Deciding Bisimilarities on Distributions

Christian Eisentraut[1], Holger Hermanns[1], Julia Krämer[1],
Andrea Turrini[1], and Lijun Zhang[2,3,1]

[1] Saarland University – Computer Science, Saarbrücken, Germany
[2] State Key Laboratory of Computer Science, Institute of Software,
Chinese Academy of Sciences, Beijing, China
[3] DTU Informatics, Technical University of Denmark, Denmark

Abstract. Probabilistic automata (*PA*) are a prominent compositional concurrency model. As a way to justify property-preserving abstractions, in the last years, bisimulation relations over *probability distributions* have been proposed both in the strong and the weak setting. Different to the usual bisimulation relations, which are defined over *states*, an algorithmic treatment of these relations is inherently hard, as their carrier set is uncountable, even for finite *PA*s. The coarsest of these relation, weak distribution bisimulation, stands out from the others in that no equivalent state-based characterisation is known so far. This paper presents an equivalent state-based reformulation for weak distribution bisimulation, rendering it amenable for algorithmic treatment. Then, decision procedures for the probability distribution-based bisimulation relations are presented.

1 Introduction

Weak probabilistic bisimilarity is a well-established behavioural equivalence on probabilistic automata (*PA*) [20]. However, it is arguably too fine [6, 9]. As an example, consider the two automata in Fig. 1, where a single visible step, embedding a probabilistic decision is depicted on the left, while on the right this is split into a visible step followed by an internal, thus invisible probabilistic decision of the very same kind (indicated by τ). Intuitively, an observer should not be able to distinguish between the two automata. However, they are not weak probabilistic bisimilar.

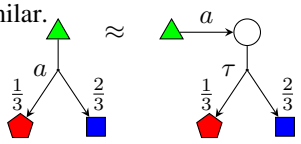


Fig. 1. Distribution bisimilarity

Markov Automata are a compositional behavioural model for continuous time stochastic and nondeterministic systems [5, 8, 9] subsuming Interactive Markov Chains (*IMC*) [12] and Probabilistic Automata. Markov automata weak bisimilarity has been introduced as an elegant and powerful way of abstracting from internal computation cascades, yielding the coarsest reasonable bisimilarity [5]. It is a conservative extension of *IMC* weak bisimilarity, and also extends weak probabilistic bisimilarity on *PA*. But different from standard bisimulation notions, Markov automata weak bisimulations are defined as relations on subprobability distributions instead of states. Translated back to the *PA* setting, this *weak distribution bisimilarity* enables to equate automata such as the ones in Fig. 1. The equivalence of these two systems rests on the ability to relate distributions. If we are only allowed to relate states, we must fail to prove bisimilarity since

we would need to require the presence of a state bisimilar to state $\circ$ on the left. This indicates that weak distribution bisimilarity is coarser than weak probabilistic bisimilarity on *PA*. It can be regarded as the symmetric version [8] of weak probabilistic forward similarity [20], the coarsest precongruence preserving trace distributions [16, 17]. The idea of distribution bisimilarity can also be instantiated in the strong setting [11], where internal computations are not abstracted from.

In this paper, we present decision algorithms for distribution bisimilarities in the strong and weak sense. Strong distribution bisimilarity requires only a minor adaptation of the polynomial time decision algorithm for strong probabilistic bisimilarity [1]. However, a decision algorithm for weak distribution bisimilarity cannot follow the traditional partition refinement approach directly. This is caused by the uncountability of the underlying carrier set, which here is the set of all distributions over the automaton's state space. The key contribution of this paper is an equivalent reformulation of weak distribution bisimulation in a state-based manner. This makes it eventually amenable to an algorithmic treatment. To arrive there, we have to tweak the usual approach to state-based characterisations of bisimulations: instead of all, *only specific* transitions of one state must be matched by its bisimilar counterpart. To identify those transitions, we introduce the concept of behaviour *preserving* transitions.

Based on this state-based characterisation, we then adapt the standard partition refinement algorithm [2, 14, 18] to decide weak bisimilarity. The algorithm successively refines the current equivalence relation by checking the conditions of the state-based characterisations. identifying the set of preserving transitions, the overall complexity of the algorithm is exponential.

The main contribution of this paper is a state-based characterisation of weak distribution bisimilarity, and a decision algorithm based on it. We develop our findings in the setting of probabilistic automata, they however carry over to Markov automata weak bisimilarity, where only the notion of maximal progress, inherited from *IMC*, requires technical care.

Organisation of the paper. After the preliminaries in Sec. 2, we introduce in Sec. 3 the state-based characterisation of the weak bisimilarity in the context of probabilistic automata. We devote Sec. 4 to prove the equivalence between state-based and distribution-based weak bisimilarities. We describe in Sec. 5 the decision procedure and we conclude the paper by Sec. 6 with a discussion on related and future work and by Sec. 7 with some general remarks.

2 Preliminaries

For a set X, we denote by $\mathrm{SubDisc}(X)$ the set of discrete sub-probability distributions over X. Given $\rho \in \mathrm{SubDisc}(X)$, we denote by $|\rho|$ the probability mass $\rho(X)$ of a subdistribution, by $\mathrm{Supp}(\rho)$ the set $\{\, x \in X \mid \rho(x) > 0 \,\}$, by $\rho(\bot)$ the value $1 - \rho(X)$ where $\bot \notin X$, and by δ_x, where $x \in X \cup \{\bot\}$, the *Dirac* distribution such that $\rho(y) = 1$ for $y = x$, 0 otherwise; $\delta_\bot$ represents the empty distribution such that $\rho(X) = 0$. We call a distribution ρ *full*, or simply a *probability* distribution, if $|\rho| = 1$. The set of all discrete probability distributions over X is denoted by $\mathrm{Disc}(X)$.

The lifting $\mathcal{L}(\mathcal{B}) \subseteq \mathrm{SubDisc}(X) \times \mathrm{SubDisc}(X)$ [15] of an equivalence relation $\mathcal{B}$ on X is defined as: for $\rho_1, \rho_2 \in \mathrm{SubDisc}(X)$, $\rho_1 \ \mathcal{L}(\mathcal{B}) \ \rho_2$ if and only if for each $\mathcal{C} \in X/\mathcal{B}$, $\rho_1(\mathcal{C}) = \rho_2(\mathcal{C})$. We define the distribution $\rho := \rho_1 \oplus \rho_2$ by $\rho(s) = \rho_1(s) + \rho_2(s)$ provided $|\rho| \leq 1$, and conversely we say ρ can be split into ρ_1 and ρ_2. Since $\oplus$ is associative and commutative, we may use the notation $\bigoplus$ for arbitrary finite sums. Similarly, we define $\rho := \rho_1 \ominus \rho_2$ by $\rho(s) = \max\{\rho_1(s) - \rho_2(s), 0\}$. For notation convenience, for a state s, we denote by $\rho \ominus s$ the distribution $\rho \ominus \delta_s$.

It is often convenient to consider distributions as relations over $X \times \mathbb{R}^{\geq 0}$ and thus explicitly denote the distribution μ by the relation $\{ (s : p_s) \mid s \in X, p_s = \mu(s) \}$.

A Probabilistic Automaton (*PA*) [20] $\mathcal{A}$ is a quadruple $(S, \bar{s}, \Sigma, D)$, where S is a finite set of *states*, $\bar{s} \in S$ is the *start* state, Σ is the set of *actions*, and $D \subseteq S \times \Sigma \times \mathrm{Disc}(S)$ is a *probabilistic transition relation*. The set Σ is partitioned into two sets $H = \{\tau\}$ and E of internal (hidden) and external actions, respectively; we refer to $\bar{s}$ also as the *initial* state and we let s,t,u,v, and their variants with indexes range over S and a, b over actions. In this work we consider only finite *PA*s, i.e., automata such that S and D are finite.

A transition $tr = (s, a, \mu) \in D$, also denoted by $s \xrightarrow{a} \mu$, is said to *leave* from state s, to be *labelled* by a, and to *lead* to μ, also denoted by μ_{tr}. We denote by $src(tr)$ the *source* state s, by $act(tr)$ the *action* a, and by $trg(tr)$ the *target* distribution μ. We also say that s enables action a, that action a is enabled from s, and that (s, a, μ) is enabled from s. Finally, we denote by $D(s)$ the set of transitions enabled from s, i.e., $D(s) = \{ tr \in D \mid src(tr) = s \}$, and similarly by $D(a)$ the set of transitions with action a, i.e., $D(a) = \{ tr \in D \mid act(tr) = a \}$.

Weak Transitions. An *execution fragment* of a *PA* $\mathcal{A}$ is a finite or infinite sequence of alternating states and actions $\alpha = s_0 a_1 s_1 a_2 s_2 \ldots$ starting from a state s_0, also denoted by $first(\alpha)$, and, if the sequence is finite, ending with a state denoted by $last(\alpha)$, such that for each $i > 0$ there exists a transition $(s_{i-1}, a_i, \mu_i) \in D$ such that $\mu_i(s_i) > 0$. The *length* of α, denoted by $|\alpha|$, is the number of occurrences of actions in α. If α is infinite, then $|\alpha| = \infty$. Denote by $frags(\mathcal{A})$ the set of execution fragments of $\mathcal{A}$ and by $frags^*(\mathcal{A})$ the set of finite execution fragments of $\mathcal{A}$. An execution fragment α is a *prefix* of an execution fragment α', denoted by $\alpha \leqslant \alpha'$, if the sequence α is a prefix of the sequence α'. The *trace* $trace(\alpha)$ of α is the sub-sequence of external actions of α; we denote by ε the empty trace. Similarly, we define $trace(a) = a$ for $a \in E$ and $trace(\tau) = \varepsilon$.

A *scheduler* for a *PA* $\mathcal{A}$ is a function $\sigma \colon frags^*(\mathcal{A}) \to \mathrm{SubDisc}(D)$ such that for each finite execution fragment α, $\sigma(\alpha) \in \mathrm{SubDisc}(D(last(\alpha)))$. Note that by using sub-probability distributions, it is possible that with some non-zero probability no transition is chosen after α, that is, the computation stops after α. A scheduler is *determinate* [2] if for each pair of execution fragments α, α', if $trace(\alpha) = trace(\alpha')$ and $last(\alpha) = last(\alpha')$, then $\sigma(\alpha) = \sigma(\alpha')$. A scheduler is *Dirac* if for each α, $\sigma(\alpha)$ is a Dirac distribution. Given a scheduler σ and a finite execution fragment α, the distribution $\sigma(\alpha)$ describes how transitions are chosen to move on from $last(\alpha)$. A scheduler σ and a state s induce a probability distribution $\mu_{\sigma,s}$ over execution fragments as follows. The basic measurable events are the cones of finite execution fragments, where the cone of α, denoted by C_α, is the set $\{ \alpha' \in frags(\mathcal{A}) \mid \alpha \leqslant \alpha' \}$. The probability $\mu_{\sigma,s}$ of a

cone C_α is recursively defined as:

$$\mu_{\sigma,s}(C_\alpha) = \begin{cases} 0 & \text{if } \alpha = t \text{ for a state } t \neq s, \\ 1 & \text{if } \alpha = s, \\ \mu_{\sigma,s}(C_{\alpha'}) \cdot \sum_{tr \in D(a)} \sigma(\alpha')(tr) \cdot \mu_{tr}(t) & \text{if } \alpha = \alpha' a t. \end{cases}$$

Standard measure theoretical arguments ensure that $\mu_{\sigma,s}$ extends uniquely to the σ-field generated by cones. We call the resulting measure $\mu_{\sigma,s}$ a *probabilistic execution fragment* of $\mathcal{A}$ and we say that it is generated by σ from s. Given a finite execution fragment α, we define $\mu_{\sigma,s}(\alpha)$ as $\mu_{\sigma,s}(\alpha) = \mu_{\sigma,s}(C_\alpha) \cdot \sigma(\alpha)(\bot)$, where $\sigma(\alpha)(\bot)$ is the probability of terminating the computation after α has occurred.

We say that there is a *weak combined transition* from $s \in S$ to $\mu \in \mathrm{Disc}(S)$ labelled by $a \in \Sigma$, denoted by $s \stackrel{a}{\Longrightarrow}_{\mathrm{c}} \mu$, if there exists a scheduler σ such that the following holds for the induced probabilistic execution fragment $\mu_{\sigma,s}$: (1) $\mu_{\sigma,s}(\mathit{frags}^*(\mathcal{A})) = 1$; (2) for each $\alpha \in \mathit{frags}^*(\mathcal{A})$, if $\mu_{\sigma,s}(\alpha) > 0$ then $\mathit{trace}(\alpha) = \mathit{trace}(a)$ (3) for each state t, $\mu_{\sigma,s}(\{\, \alpha \in \mathit{frags}^*(\mathcal{A}) \mid \mathit{last}(\alpha) = t \,\}) = \mu(t)$. In this case, we say that the weak combined transition $s \stackrel{a}{\Longrightarrow}_{\mathrm{c}} \mu$ is induced by σ.

We remark that $\mathit{trace}(\alpha) = \mathit{trace}(a)$ is equivalent to $\mathit{trace}(\alpha) = \varepsilon$ for $a = \tau$ and $\mathit{trace}(\alpha) = a$ for $a \in E$. Moreover, the first two conditions can be equivalently replaced by $\mu_{\sigma,s}(\{\, \alpha \in \mathit{frags}^*(\mathcal{A}) \mid \mathit{trace}(\alpha) = \mathit{trace}(a) \,\}) = 1$.

Given a set of *allowed* transitions $\breve{A} \subseteq D$, we say that there is an *allowed weak combined transition* [13] from s to μ with label a respecting $\breve{A}$, denoted by $s \stackrel{a \downarrow \breve{A}}{\Longrightarrow}_{\mathrm{c}} \mu$, if there exists a scheduler σ inducing $s \stackrel{a}{\Longrightarrow}_{\mathrm{c}} \mu$ such that for each $\alpha \in \mathit{frags}^*(\mathcal{A})$, $\mathrm{Supp}(\sigma(\alpha)) \subseteq \breve{A}$.

Albeit the definition of weak combined transitions is somewhat intricate, this definition is just the obvious extension of weak transitions on labelled transition systems to the setting with probabilities. See [21] for more details on weak combined transitions.

Example 1. As an example of weak combined transition, consider the probabilistic automaton depicted in Fig. 2 and the probability distribution $\mu = \{(\pentagon : \frac{3}{4}), (⑤ : \frac{1}{4})\}$. It is immediate to verify that the weak combined transition ① $\stackrel{\tau}{\Longrightarrow}_{\mathrm{c}} \mu$ is induced by the Dirac determinate scheduler σ defined as follows: $\sigma(①) = \delta_{tr_1}$, $\sigma(① \tau ②) = \delta_{tr_2}$, $\sigma(① \tau ③) = \delta_{tr_3}$, $\sigma(① \tau ② \tau ④) = \sigma(① \tau ③ \tau ④) = \delta_{tr_4}$, and $\sigma(\alpha) = \delta_\bot$ for each other finite execution fragment α. If we consider all transitions but tr_2 as allowed transitions $\breve{A}$, then there is no scheduler inducing ① $\stackrel{\tau \downarrow \breve{A}}{\Longrightarrow}_{\mathrm{c}} \mu$. In fact, using this set of allowed transitions, the maximal probability of reaching $\pentagon$ from ① is $\frac{1}{4}$ by the execution fragment ① τ ③ τ ④ τ $\pentagon$. □

We say that there is a *weak (allowed) hyper transition* from $\rho \in \mathrm{SubDisc}(S)$ to $\mu \in \mathrm{SubDisc}(S)$ labelled by $a \in \Sigma$, denoted by $\rho \stackrel{a}{\Longrightarrow}_{\mathrm{c}} \mu$ ($\rho \stackrel{a \downarrow \breve{A}}{\Longrightarrow}_{\mathrm{c}} \mu$), if there exists a family of (allowed) weak combined transitions $\{s \stackrel{a}{\Longrightarrow}_{\mathrm{c}} \mu_s\}_{s \in \mathrm{Supp}(\rho)}$ ($\{s \stackrel{a \downarrow \breve{A}}{\Longrightarrow}_{\mathrm{c}} \mu_s\}_{s \in \mathrm{Supp}(\rho)}$) such that $\mu = \bigoplus_{s \in \mathrm{Supp}(\rho)} \rho(s) \cdot \mu_s$.

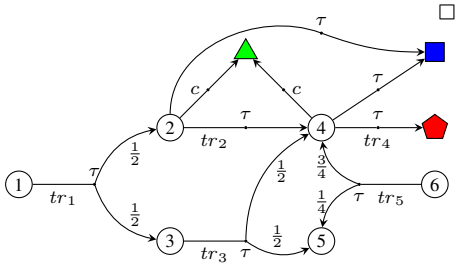


Fig. 2. A probabilistic automaton

3 Probabilistic Bisimulations

For non-stochastic systems, the idea of bisimulation can be formalised as a binary symmetric relation $\mathcal{B}$ over states where each pair of states $(s,t) \in \mathcal{B}$ satisfies that whenever $s \xrightarrow{a} s'$ for some state s', then there exists a state t' such that $t \xrightarrow{a} t'$ and $s \ \mathcal{B} \ t$. Strong bisimilarity is the union of all such strong bisimulations. Bisimulation can be seen as a game [7, 22, 23], and therefore one often calls s the *challenger* proposing a transition and t the *defender*. Phrased differently, in a bisimulation, *every* transition of a challenger must be matched by *some* transition of its corresponding defender. Weak bisimulation and bisimilarity is defined analogously, but with the strong transition arrow $\xrightarrow{a}$ replaced by its weak variant $\stackrel{a}{\Longrightarrow}$ that in addition allows to perform arbitrary sequences of τ actions before and after the action a is performed.

When translating the idea of bisimulation to probabilistic systems, it is generalised in order to account for the probabilistic setting: Transitions $\longrightarrow$ and $\Longrightarrow$ are replaced by their combined variants $\longrightarrow_c$ and $\Longrightarrow_c$, and target states s' and t' become target distributions μ and γ over states, respectively. Finally, target distributions must match up to the lifting of $\mathcal{B}$ to distributions ($\mathcal{L}(\mathcal{B})$). For a detailed motivation of these adaptions we refer the interested reader to [20]. Strong and weak probabilistic bisimulation can then be defined as follows.

Definition 1 (Strong and Weak Probabilistic Bisimulations). *For a probabilistic automaton $\mathcal{A} = (S, \bar{s}, \Sigma, D)$, a symmetric relation $\mathcal{B}$ over S is a probabilistic bisimulation, if each pair of states $(s,t) \in \mathcal{B}$ satisfies for every $a \in \Sigma$: $s \overset{a}{\rightsquigarrow} \mu$ implies $t \overset{a}{\rightsquigarrow} \gamma$ for some $\gamma \in \mathrm{Disc}(S)$ and $\mu \ \mathcal{L}(\mathcal{B}) \ \gamma$.*

We call $\mathcal{B}$ strong, if $\rightsquigarrow = \longrightarrow_c$ and weak if $\rightsquigarrow = \Longrightarrow_c$. The union of all strong (weak) bisimulation relations is called strong (weak) bisimilarity. For a uniform presentation, our definitions differ from the standard in the challenger's transition, which usually chooses a strong and not combined transition. The resulting bisimilarities can, however, be shown to be identical.

It is worthwhile to observe that weak probabilistic bisimulation is often considered too fine when it comes to intuitively unobservable behavioural differences [6, 9]. This has been already illustrated in Fig. 1, where weak probabilistic bisimulation fails to equate the automata on the left and the right hand side. We are going to shed some more light on this.

Example 2. (Weak Probabilistic Bisimulation is Too Fine) Consider again the *PA* depicted in Fig. 2, where non-circular shaped states are supposed to have pairwise distinct behaviour. Intuitively, the observable behaviour of state ① cannot be distinguished from that of state ⑥: whenever the action c happens, or likewise, any of the non-round states is reached, this happens with the same probability for both ① and ⑥. In [20], this intuition of what the coarsest reasonable notion of observability is, has been formalised as *trace distribution precongruence*, that has been proven equivalent [16] to the notion of *weak probabilistic forward similarity*. The latter relates states to probability distributions over states. However, weak probabilistic bisimilarity distinguishes between the two states, as already the first transition of ① to the distribution $\gamma = \left(\frac{1}{2}\delta_{②}\right) \oplus \left(\frac{1}{2}\delta_{③}\right)$ cannot be matched by ⑥. The reason is that the only distribution reachable from ⑥ is

$\mu = \left(\frac{3}{4}\delta_{④}\right) \oplus \left(\frac{1}{4}\delta_{⑤}\right)$. Clearly, for μ $\mathcal{L}(\mathcal{B})$ γ to hold, all states ②, ③, ④ and ⑤ must be equivalent. However, this cannot be the case, as for example ⑤ cannot perform any transition, while state ④ can perform a transition labelled with c. This means that although, ① and ⑥ show the same *observable* behaviour with the same probability, they are distinguished by weak probabilistic bisimilarity.

Notably, all the distributions $\delta_{①}$, $\left(\frac{3}{4}\delta_{④}\right) \oplus \left(\frac{1}{4}\delta_{⑤}\right)$, and γ are pairwise trace distribution precongruent (and weak probabilistic forward similar). □

With these motivations in mind, several probabilistic bisimulation variants defined on *probability (sub)distributions* over states have been introduced for the strong setting [11] and for the weak setting [5, 8, 9]. For the weak setting, there currently exist three different variations, however, two of them essentially coincide [6]. We will recall these notions in the following. Again, our definitions differ from the original definitions for the sake of a uniform presentation, which allows to highlight differences and similarities clearly.

Definition 2 (Strong and Weak Probabilistic Distribution Bisimulations). *For a PA $\mathcal{A} = (S, \bar{s}, \Sigma, D)$, a symmetric relation $\mathcal{B}$ over* $\mathrm{SubDisc}(S)$ *is a probabilistic distribution bisimulation, if each pair of subdistributions $(\mu, \gamma) \in \mathcal{B}$ satisfies $|\mu| = |\gamma|$ and for every $a \in \Sigma$*

(a) $\mu \overset{a}{\rightsquigarrow} \mu'$ *implies* $\gamma \overset{a}{\rightsquigarrow} \gamma'$ *for some* $\gamma \in \mathrm{SubDisc}(S)$ *and* $\mu' \; \mathcal{B} \; \gamma'$.
(b) $\mu = \mu_1 \oplus \mu_2$ *implies* $\gamma = \gamma_1 \oplus \gamma_2$ *for some* $\gamma_1, \gamma_2 \in \mathrm{SubDisc}(S)$ *such that* $\mu_i \; \mathcal{B} \; \gamma_i$ *for* $i \in \{1, 2\}$.

As before, we obtain the strong and weak variants by replacing $\rightsquigarrow$ by $\longrightarrow_c$ and $\Longrightarrow_{\mathrm{c}}$ respectively; the corresponding bisimilarities are defined as the union of all respective bisimulations.

As shown in [11] (for strong) and [8] (for weak), these distribution-based bisimilarities are indeed reformulations of their state-based counterparts, in so far that for two states s and t, the distributions δ_s and δ_t are bisimilar in the distribution-based bisimulations, if and only if s and t are bisimilar in the respective state-based counterparts.

The weak bisimilarities defined in [9] and [5] (for Markov automata) coincide [6], if restricted to probabilistic automata, but do not correspond to any known state-based bisimilarity. We can define them as follows.

Definition 3 (Weak Distribution Bisimulation). *For a PA $\mathcal{A} = (S, \bar{s}, \Sigma, D)$, a symmetric relation $\mathcal{B}$ over* $\mathrm{SubDisc}(S)$ *is a weak distribution bisimulation, if each pair of subdistributions $(\mu, \gamma) \in \mathcal{B}$ satisfies $|\mu| = |\gamma|$ and for every $a \in \Sigma$*

(a) $\mu \overset{a}{\Longrightarrow}_{\mathrm{c}} \mu'$ *implies* $\gamma \overset{a}{\Longrightarrow}_{\mathrm{c}} \gamma'$ *for some* $\gamma' \in \mathrm{SubDisc}(S)$ *and* $\mu' \; \mathcal{B} \; \gamma'$.
(b) $\mu = \mu_1 \oplus \mu_2$ *implies* $\gamma \overset{\tau}{\Longrightarrow}_{\mathrm{c}} \gamma_1 \oplus \gamma_2$ *for some* $\gamma_1, \gamma_2 \in \mathrm{SubDisc}(S)$ *such that* $\mu_i \; \mathcal{B} \; \gamma_i$ *for* $i \in \{1, 2\}$.

The union of all weak distribution bisimulation relations is called weak distribution bisimilarity, denoted by $\approx$. It is an equivalence relation and the coarsest weak distribution bisimulation relation. Two *PA*s are weak distribution bisimilar if the Dirac distributions of their initial states are weak distribution bisimilar in the direct sum of the two *PA*s, i.e., in the automaton whose components are the disjoint union of the components of the two automata. We project the relation $\approx$ to states (denoted $\approx_\delta$) as follows. We say that two states s, t are related by $\approx_\delta \subseteq S \times S$, if and only if $\delta_s \approx \delta_t$.

The strength of this definition is the introduction of a weak transition in Condition (b). As already noted in [8], this is in fact the only difference to weak *probabilistic* distribution bisimulation (Def. 2).

While in Ex. 2 we have argued that the distributions $\gamma = \left(\frac{1}{2}\delta_{②}\right) \oplus \left(\frac{1}{2}\delta_{③}\right)$ and $\mu = \left(\frac{3}{4}\delta_{④}\right) \oplus \left(\frac{1}{4}\delta_{⑤}\right)$ are not weak probabilistic bisimilar in the *PA* of Fig. 2, they satisfy $\mu \approx \gamma$, because $\gamma \stackrel{\tau}{\Longrightarrow}_{\mathrm{c}} \mu$, which is effectively the only transition of γ, and it thus directly satisfies Condition (a) and (b) of Def. 3.

So while distribution-based bisimulations give rise to coarser and more natural notions of equality, they also have severe drawbacks. A distribution-based bisimulation relation that is to be constructed in order to prove two systems bisimilar is much harder to define than for a state-based relation: For state-based bisimulations only the set of reachable states must be considered and suitably related pairwise. In contrast for distribution-based systems the potentially uncountable set of all reachable distributions needs to be considered. This gets problematic when it comes to algorithmic checks for bisimilarity, for example, in the context of verification of systems and state-space minimisation by bisimulation quotienting. Standard partition refinement approaches usually applied in this context seem infeasible here, as even for finite state space, the problem space (i.e., the reachable distributions) is uncountable.

For the strong and weak distribution-based bisimilarities according to Def. 2 the above issue is not a problem, since they can be reduced to the state-based setting. For weak distribution bisimilarity according to Def. 3, the situation is more complicated as no state-based characterisation is known, and it is by far not obvious how to arrive at such a characterisation. To approach this, we will now give an intuitive explanation why the fact that weak probabilistic bisimilarity is too distinctive seems rooted in the fact that it is a naturally state-based relation, and then explain how to overcome the problem while maintaining the state-based bisimulation approach as far as possible.

For the discussion that follows, we assume a generic underlying notion of observation equivalence such as a trace distribution-based equivalence. We call a state s *behaviourally pivotal*, if $s \xrightarrow{\tau} \mu$ implies that s and μ are not observation equivalent, i.e., μ is not able to perform $\mu \stackrel{\tau}{\Longrightarrow}_{\mathrm{c}} \rho$ such that s and ρ are observation equivalent.

Ex. 2. (cont'd) (Behaviourally Pivotal States) Assume again that all non-round states of the *PA* in Fig. 2 induce pairwise distinct behaviour (for example each state can only perform a different external action). Then state ④ is behaviourally pivotal, since none of its internal successor distributions $\delta_{■}$ and $\delta_{▲}$ can behaviourally match the other, and thus cannot preserve the behaviour of s. Trivially, also ⑤ is behaviourally pivotal, since it has no successors. In contrast, all other states are *not* behaviourally pivotal, as for each of them the behaviour is fully preserved by one of its respective τ-successor distributions. In particular, state ② is not behaviourally pivotal since its behaviour is fully preserved by $\delta_{④}$ via transition ② $\xrightarrow{\tau}$ $\delta_{④}$. □

Consider the probability distribution $\mu = \left(\frac{3}{4}\delta_{④}\right) \oplus \left(\frac{1}{4}\delta_{⑤}\right)$ over behaviourally pivotal states. From the perspective of the individual behaviour of the single states in its support, this distribution is different from the distribution $\gamma = \left(\frac{1}{2}\delta_{②}\right) \oplus \left(\frac{1}{2}\delta_{③}\right)$ over non-pivotal states. For example, from the perspective of an observer, ③ $\in \mathrm{Supp}(\gamma)$ can perform the transition to ⬟ with *at most* probability $\frac{1}{2}$. In comparison, state ④ $\in \mathrm{Supp}(\mu)$ can perform this transition with probability 1, while ⑤ $\in \mathrm{Supp}(\mu)$ cannot perform this transition at all.

However, as we have discussed in Ex. 2, both distributions as such can be regarded as observation equivalent. Weak probabilistic bisimilarity, however, focusing on state-

wise behaviour, needs to distinguish between μ and γ regardless of the fact that distribution γ, consisting only of non-pivotal states, can by no means be noticed by an observer, as it is merely skipped over on the way from ① to μ.

From the discussion so far, we will now derive necessary steps to recast Def. 3 in a state-based setting. As we have seen, the fact that weak probabilistic bisimilarity is arguably too fine is mainly due to the fact that it is too much focused on single state behaviour. More precisely, the problem is that it treats behaviourally non-pivotal states (e.g., ② and ③) in the same way as pivotal states (e.g., ④ and ⑤). To overcome this, a state-based characterisation of weak distribution bisimilarity will first of all identify pivotal states, and then, speaking from the game perspective on bisimulation, allow the bisimulation challenger only to propose a challenging transition to a distribution over pivotal states.

Example 3. When we want to show that ① and ⑥ are weak distribution bisimilar, then the challenger should not be allowed to propose the transition to $\gamma = \left(\frac{1}{2}\delta_{②}\right) \oplus \left(\frac{1}{2}\delta_{③}\right)$, which has non-pivotal states in its support (actually both states are non-pivotal). Instead, it may only propose $\left(\frac{3}{4}\delta_{④}\right) \oplus \left(\frac{1}{4}\delta_{⑤}\right)$. □

In fact, our approach will not characterise pivotal states explicitly, but rather use a set of distinguished internal transitions $(s, \tau, \mu) \in D(\tau)$ with the property that δ_s and μ are behaviourally equivalent. We call such transitions *preserving*. As a state is pivotal if it has no internal successor distribution that can fully mimic its behaviour, the set of pivotal state then is exactly the set of all states that *do not* enable a preserving transition.

The technically crucial idea of our approach is to define the bisimulation relation $\mathcal{B}$ over states and the set P of distinguished transitions simultaneously. The definitions of $\mathcal{B}$ and P will be mutually dependent. This allows us to use the information from set P to identify pivotal states when defining the bisimulation $\mathcal{B}$. Vice versa, the information provided from the bisimulation $\mathcal{B}$ allows us to determine when a state has a τ-successor distribution, that is behaviourally equivalent. As it is technically more convenient, we will not formally define the notion of pivotal states in the sequel, but directly work with the notion of preserving transitions instead.

Definition 4 (Preserving Transitions). *Let $\mathcal{B}$ be an equivalence relation on S. We call an internal transition $(s, \tau, \gamma) \in D(\tau)$* preserving with respect to *$\mathcal{B}$ if whenever $s \stackrel{a}{\Longrightarrow}_{\mathrm{c}} \mu$ then there exist μ', γ' such that $\mu \stackrel{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}} \mu'$, $\gamma \stackrel{a}{\Longrightarrow}_{\mathrm{c}} \gamma'$, and $\mu' \; \mathcal{L}(\mathcal{B}) \; \gamma'$.*
We call a set $P \subseteq D(\tau)$ preserving with respect to *$\mathcal{B}$ if it only consists of preserving transitions.*

Example 4. In Fig 2, transitions tr_1, tr_2, tr_3 and tr_5 are preserving, while all other transitions are not. It is especially interesting to note that tr_2 is preserving while the other internal transition leaving ② is not, as ■ is not behaviourally equivalent to ②. □

Given a set P of preserving transitions, we from now on call weak (hyper) transitions of the form $\stackrel{\tau \downarrow P}{\Longrightarrow}$ *preserving* weak (hyper) transitions, and $\stackrel{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}}$ *preserving* weak combined (hyper) transitions.

Definition 5 (State-Based Characterisation of Weak Distribution Bisimulation). *An* equivalence *relation $\mathcal{B}$ on S is called a* state-based weak distribution bisimulation*, if there is a set $P \subseteq D(\tau)$ that is preserving with respect to $\mathcal{B}$ and whenever $s \; \mathcal{B} \; t$,*

1. *if* $s \stackrel{a}{\Longrightarrow}_{\mathrm{c}} \mu$ *for some* μ*, then* $t \stackrel{a}{\Longrightarrow}_{\mathrm{c}} \gamma$ *for some* γ*, such that there exists* μ' *such that* $\mu \stackrel{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}} \mu'$ *and* $\mu' \; \mathcal{L}(\mathcal{B}) \; \gamma$*;*
2. *if* $s \stackrel{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}} \mu$ *for some* μ*, then* $t \stackrel{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}} \gamma$ *for some* γ*, such that there exists* μ' *such that* $\mu \stackrel{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}} \mu'$ *and* $\mu' \; \mathcal{L}(\mathcal{B}) \; \gamma$*.*

We write $s \approx_{\mathrm{s}} t$ if there exists a state-based weak distribution bisimulation relating s and t.

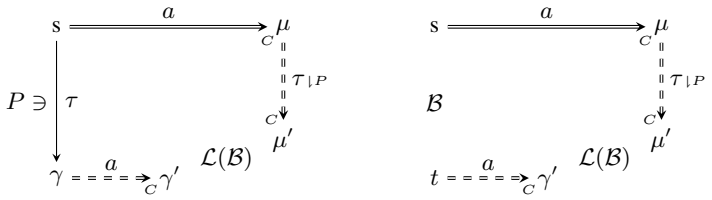


Fig. 3. Preserving transitions (left) and Condition 1 of state-based weak distribution bisimulation

In Fig. 3, preserving transitions and state-based weak distribution bisimulation are explained graphically. Solid lines denote challenger transitions, dashed lines defender transitions. Different to weak probabilistic bisimulation, the role of the defender in the bisimulation game is no longer linked exclusively to transitions of t. Although the weak allowed hyper transition from μ to μ' originates from a successor distribution of s rather than t, the defender can choose this transition. As a consequence, the defender does not need to match the challenging distribution μ directly, but it is allowed to choose an arbitrary distribution μ', which it is able to match, as long as $\mu \stackrel{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}} \mu'$. In intuitive terms, a transition (t, τ, ξ) is in P if t is non-pivotal; the existence of a transition $\mu \stackrel{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}} \mu'$ (with $\mu \neq \mu'$) means that μ must contain non-pivotal states, and thus, we liberate the defender from its obligation to match μ by allowing it to match μ' instead. At the same time, if μ was a distribution exclusively over pivotal states, then no $\mu' \neq \mu$ would exist such that $\mu \stackrel{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}} \mu'$. Thus, the defender is forced to match exactly distributions over pivotal states. Intuitively, we want a transition $s \xrightarrow{\tau} \gamma$ to be contained in P, exactly if s and γ allow the same observations, which in turn means that s is non-pivotal. Formally, this is achieved by defining P completely analogous to the state-based characterisation of weak distribution bisimulation. The only difference is that the role of the defender is played by a distribution γ, instead of a state.

So far, we have left Condition 2 of Def. 5 unmentioned, which expresses that if one of two related states can perform transitions within P, then also the other state must be able to match these transitions within P. This condition might come unexpected, and we claim, that the condition can be dropped without affecting the resulting notion of bisimilarity. Yet, currently it is needed for the proof of Thm. 1, which establishes that the distribution-based and the state-based characterisation of weak distribution bisimilarity are indeed equivalent.

Example 5. (State-Based Weak Distribution Bisimulation) Consider again the *PA* depicted in Fig. 2 and suppose that states ⬟, ▲, and ■ are not weak bisimilar. The equiva-

lence relation $\mathcal{B}$ whose non-singleton classes are $\{①, ⑥\}$ and $\{②, ④\}$ is a state-based weak distribution bisimulation, given $P = \{tr_1, tr_2, tr_3, tr_5\}$.

Checking the step condition for the pair $(②, ④)$ is trivial, so let us focus on the pair $(①, ⑥)$. Each weak combined transition $⑥ \stackrel{\tau}{\Longrightarrow}_{\mathrm{c}} \mu$ enabled by ⑥, can be matched by ① by reaching μ_{tr_5} (via preserving transitions tr_1, tr_2, and tr_3 chosen with probability 1) and then behaving as in $⑥ \stackrel{\tau}{\Longrightarrow}_{\mathrm{c}} \mu$. If we stay in ⑥ with non-zero probability, then we remain in ① with the same probability and the lifting condition is satisfied.

Now, consider the weak transition $① \stackrel{\tau}{\Longrightarrow}_{\mathrm{c}} \mu$ enabled by ①, where $\mu = \{(② : \frac{1}{2}), (③ : \frac{1}{2})\}$ (this is actually the ordinary transition tr_1). ⑥ has no way to reach μ so it needs help of ① to match such a transition: ⑥ performs the transition $⑥ \stackrel{\tau}{\Longrightarrow}_{\mathrm{c}} \gamma$ where $\gamma = \{(④ : \frac{3}{4}), (⑤ : \frac{1}{4})\}$, i.e., it performs tr_5, while μ reaches γ by the preserving weak hyper transition $\mu \stackrel{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}} \gamma$ by choosing with probability 1 preserving transitions tr_2 from ② and tr_3 from ③ and then stopping.

The transition $① \stackrel{\tau}{\Longrightarrow}_{\mathrm{c}} \mu$ is not the only weak combined transition enabled by ①. It enables, for instance, the weak combined transition $① \stackrel{\tau}{\Longrightarrow}_{\mathrm{c}} \rho$ where $\rho = \{(■ : \frac{1}{2}), (③ : \frac{1}{2})\}$. ⑥ matches this transition by enabling $⑥ \stackrel{\tau}{\Longrightarrow}_{\mathrm{c}} \phi$ where $\phi = \{(■ : \frac{1}{2}), (④ : \frac{1}{4}), (⑤ : \frac{1}{4})\}$ that can be reached from ρ by the preserving weak hyper transition $\rho \stackrel{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}} \phi$ obtained by performing no transitions from ■ and choosing tr_3 (that is preserving) with probability 1 and then stopping. There are several other transitions enabled by ① that can be matched in a similar way. □

Finally, we want to remark that weak probabilistic distribution bisimulation given in Def. 2 is obtained from Def. 5 by requiring $P = \emptyset$, since, when $P = \emptyset$, we have that $s \stackrel{\tau \downarrow \emptyset}{\Longrightarrow}_{\mathrm{c}} \mu$ implies $\mu = \delta_s$ as well as $\mu \stackrel{\tau \downarrow \emptyset}{\Longrightarrow}_{\mathrm{c}} \mu'$ implies $\mu' = \mu$.

4 Correctness of the Characterisation

The correctness of the state-based characterisation of weak distribution bisimilarity will be formalised by Thm. 1. We obtain this equality in a slightly restricted setting where we collapse probability 1 cycles, or *maximal end components* (*mec*s) [4], i.e., τ-cycles where it is possible to return to each state of the cycle with probability 1. This restriction, that is due to technical reasons, does not affect the general applicability of Thm. 1 since collapsing *mec*s preserves $\approx_\delta$, as stated by Lemma 1.

We will now define the restricted setting in which we will then establish the correctness proof. Along the way, we will present insightful examples where the unrestricted setting has caused unexpected difficulties. In the restricted setting, we will only consider *PA*s, where no cyclic structure in the following sense exists.

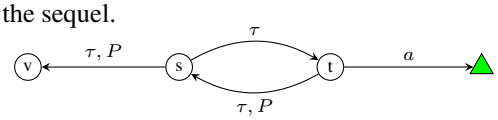
Definition 6 (Maximal End Components). *Given a PA $\mathcal{A}$ with set of states S, a* maximal end component *(mec) is a maximal set $C \subseteq S$ such that for each $s, t \in C$: $s \stackrel{\tau}{\Longrightarrow}_{\mathrm{c}} \delta_t$ and $t \stackrel{\tau}{\Longrightarrow}_{\mathrm{c}} \delta_s$.*

The definition stems from [4]. The set of all *mec*s is a disjoint partitioning of S. Thus, the relation $=_{mec}$, where $s =_{mec} t$ if and only if s and t lie in the same *mec*, is an equivalence relation on states. All states that lie in the same *mec* can mutually reach

each other with τ transitions with probability 1[4]. It is thus straightforward to show that such states are weak distribution bisimilar.

Lemma 1. *$s =_{mec} t$ implies $s \approx_\delta t$.*

Surprisingly, the presence of *mec*s in a *PA* leads to unexpected results. In general, it is folklore knowledge that replacing the weak challenger transition $s \stackrel{a}{\Longrightarrow}_c \mu$ of weak bisimulation by a strong challenger transition $s \stackrel{a}{\longrightarrow} \mu$ leads to equipotent characterisations of the induced bisimilarities. For state-based weak distribution bisimilarity, this is not the case. We will refer to this variation of Definition 5 as the *strong challenger characterisation* in the sequel.



Example 6. (Strong Challenger Characterisation is Broken in the Presence of Mecs.) Consider the automaton above. All transitions in this example are Dirac transitions. We label two transitions with τ, P in order to express that they are elements of P, the set of supposedly preserving transitions considering the strong challenger characterisation. Note that, however, the transition from (s) to (v) is not a preserving transition in the sense of the original definition with respect to any bisimulation relation $\mathcal{B}$, since (s) can reach ▲ with a weak a transition, whereas (v) cannot perform an a transition at all. However, all conditions of the strong challenger characterisation are satisfied. The only non-preserving strong transition (s) can perform is the one to (t). Now it is enough that (t) can reach (v) via preserving transitions, by using $t \stackrel{\tau}{\longrightarrow} \delta_s$ and $s \stackrel{\tau}{\longrightarrow} \delta_v$. For completeness, it is easy to check that the transition from (t) to (s) satisfies the conditions to be a preserving transition. With this result, it is straightforward to construct two bisimulations $\mathcal{B}_1$ and $\mathcal{B}_2$ (satisfying the strong challenger characterisation), where $\mathcal{B}_1$ is the reflexive, transitive and symmetric closure of the relation containing only the pair $(\text{(v)}, \text{(s)})$ and $\mathcal{B}_2$ accordingly containing $(\text{(s)}, \text{(t)})$. It is easy to check that for both $\mathcal{B}_1$ and $\mathcal{B}_2$ our choice of preserving transitions satisfies the strong challenger characterisation. If this characterisation now indeed was equivalent to $\approx_\delta$, the restriction of $\approx$ to states, then also (v) $\approx_\delta$ (s) and (s) $\approx_\delta$ (t) would hold and thus, by transitivity, also (v) $\approx_\delta$ (t). But clearly, this cannot hold, as (t) can perform an a-transition while (v) cannot. □

These considerations led us to only consider *mec*-contracted *PA*s in the following.

Definition 7 (*Mec*-Contracted *PA*). *A PA $\mathcal{A}$ is called* mec-contracted*, if for each pair of states $s, t \in S$, $s \stackrel{\tau}{\Longrightarrow}_c \delta_t$ and $t \stackrel{\tau}{\Longrightarrow}_c \delta_s$ implies $s = t$.*

Obviously, the quotient under $=_{mec}$ is a *mec*-contracted automaton, where the quotient under $=_{mec}$ of a *PA* $\mathcal{A}$ is defined as follows:

Definition 8 (Quotient under $=_{mec}$). *Given a PA $\mathcal{A} = (S, \bar{s}, \Sigma, D)$ and the equivalence relation $=_{mec}$ on S, the* quotient under $=_{mec}$ *of $\mathcal{A}$ is the automaton $\mathcal{A}' = (S/=_{mec}, [\bar{s}]_{=_{mec}}, \Sigma, D/=_{mec})$ where $D/=_{mec} = \{ ([s]_{=_{mec}}, a, [\mu]_{=_{mec}}) \mid (s, a, \mu) \in D \}$ and $[\mu]_{=_{mec}} \in \mathrm{Disc}(S/=_{mec})$ is the probability distribution defined for each $\mathcal{C} \in S/=_{mec}$ as $[\mu]_{=_{mec}}(\mathcal{C}) = \mu(\mathcal{C})$.*

[4] Note that *mec*s are not necessarily bottom strongly connected components, as a *mec* may well be escaped by τ transitions.

In the restricted setting we have introduced, the following theorem states that the state-based characterisation (Def. 5) of weak distribution bisimulation is indeed equivalent to the original distribution-based definition (Def. 3).

Theorem 1 (Equivalence of Characterisations). *If $\mathcal{A}$ is a mec-contracted PA, then for every t and t' in S, $t \approx_\delta t'$ if and only if $t \approx_s t'$.*

5 Decision Procedure

In this section we investigate algorithms for distribution bisimilarity that decide whether two states of an automaton are equivalent. For strong and weak probabilistic distribution bisimilarity, we can rely on existing decision algorithms for the corresponding state-based characterisations [1, 2]. If we want to relate a pair of distributions (μ, ν), we can introduce two fresh states from which a transition with a fresh label goes to distribution μ, respectively ν, and then check the bisimilarity of these two states with the above mentioned algorithms tailored to the state-based setting. For weak distribution bisimilarity, we can proceed accordingly, provided we have a decision algorithm for state-based weak distribution bisimilarity. In the rest of this section, we will devise such an algorithm.

More precisely, the algorithm constructs $S/\approx_{\mathrm{s}}$, the set of equivalence classes of states under $\approx_{\mathrm{s}}$. In contrast to all known bisimulation variants, we cannot blindly apply the standard partition refinement approach [2, 14, 18], since we potentially split equivalence classes that should not be split as the result of a negative interference between the set of preserving transitions and the current partitioning, as Ex. 7 will show. We shortly repeat the general idea of partition refinement to illustrate the problems we face. Partition refinement starts with an initial partition $\mathbf{W}$, which only consists of a single set (called a block) containing all states. Thus, all states are assumed to be pairwise state-based weak distribution bisimilar. This assumption is then checked, and usually there is a reason to split the block. Refining the partition then means we successively split a block in two (or more) blocks, whenever it contains states still assumed state-based weak distribution bisimilar in the previous iteration of the refinement loop, while in the current iteration they violate any of the state-based weak distribution bisimulation conditions. When no more splitting is possible, the algorithm has found the largest state-based weak distribution bisimulation and returns that.

In our setting, we have to manipulate also the set of preserving transitions P, since it depends on the equivalence relation induced by the partition. The obvious way is to start initially with the set $D(\tau)$ of all internal transitions. Transitions are then eliminated from this set, as soon as they violate Def. 4. However, as it turns out, the two procedures, partition refinement and transition elimination, interfere negatively. Focusing on Condition 1 of Def. 5, the challenging transition $s \stackrel{a}{\Longrightarrow}_{\mathrm{c}} \mu$ is only dependent on the transition relation underlying the given PA, but not on the current partition or P. In contrast, Condition 2 demands $s \stackrel{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}} \mu$. However, the existence of such transition depends on P, which itself varies over the refinement process. As a consequence, we can obtain false negatives, if P still contains a transition starting from s that will not be contained in the final P, while the corresponding transition from t has already been eliminated from P during an earlier refinement step.

Example 7. Let $s \xrightarrow{\tau}$ ▲ and $s \xrightarrow{\tau}$ ⬟ and also $t \xrightarrow{\tau}$ ▲ and $t \xrightarrow{\tau}$ ⬟. Assume that states ▲ and ⬟ are not weak distribution bisimilar. Then, clearly, none of the transitions is preserving. However, s and t are obviously weak distribution bisimilar. Assume the transition $t \xrightarrow{\tau}$ ▲ has been eliminated from the candidate set P, but $s \xrightarrow{\tau}$ ▲ has not. Then, when we check whether s and t satisfy the second condition of Def. 5, $s \overset{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}} \delta_{\blacktriangle}$ holds, but $t \overset{\tau \downarrow P}{\Longrightarrow}_{\mathrm{c}} \delta_{\blacktriangle}$ does not. Thus, s and t will be erroneously split. □

DECIDE($\mathcal{A}$)

0.1: $\mathcal{A}' = $ QUOTIENT-UNDER-MEC($\mathcal{A}$)
0.2: $\mathbf{W} = \emptyset$
0.3: **for all** $P \subseteq D(\tau)$ **do**
0.4: $\quad \mathbf{W}' = $ QUOTIENT-WRT-PRES($\mathcal{A}', P$)
0.5: $\quad \mathbf{W} = $ JOIN($\mathbf{W}, \mathbf{W}'$)
0.6: **return** $\mathbf{W}$

QUOTIENT-WRT-PRES($\mathcal{A}, P$)

0.1: $\mathbf{W} = \{S\}$;
0.2: **repeat**
0.3: $\quad \mathbf{W}' = \mathbf{W}$;
0.4: $\quad$ **if** CONSISTENCY($P, \mathbf{W}$) **then**
0.5: $\quad\quad$ **return** $\emptyset$
0.6: $\quad (\mathcal{C}, a, \rho) = $ FINDSPLIT($\mathbf{W}, P$);
0.7: $\quad \mathbf{W} = $ REFINE($\mathbf{W}, (\mathcal{C}, a, \rho)$);
0.8: **until** $\mathbf{W} = \mathbf{W}'$
0.9: **return** $\mathbf{W}$

If we remove Condition 2 from Def. 5, then we can show that the set P can be correctly refined with respect to $\mathbf{W}$. Since currently we have to maintain such condition, we adopt a brute force approach, where we first fix P, and refine $\mathbf{W}$ according to the standard partition refinement approach with respect to the set P.[5]

We repeat the refinement described for every possible set of preserving transitions. This is done inside the **for** loop of the main procedure, DECIDE, of the algorithm. This means we consider all subsets of $D(\tau)$, which, unfortunately, is of size in $\mathcal{O}(2^{|D|})$.

The partition refinement happens in procedure QUOTIENT-WRT-PRES, which is parameterised by P. This procedure is entirely unsurprising except for a consistency check performed in procedure CONSISTENCY: During each refinement iteration of $\mathbf{W}$ in QUOTIENT-WRT-PRES, we check whether the currently assumed set P actually still satisfies Def. 4. If it does not, we stop refining and immediately return $\mathbf{W} = \emptyset$.

After each call of QUOTIENT-WRT-PRES, in procedure DECIDE the returned partitioning $\mathbf{W}$ is joined with the previously computed partitioning $\mathbf{W}'$. Procedure JOIN computes the partitioning that results from the union of the two partitionings. Treating $\mathbf{W}$ and $\mathbf{W}'$ as equivalence relations over S, it computes the reflexive, transitive and symmetric closure of $\mathbf{W} \cup \mathbf{W}'$. Thus, when QUOTIENT-WRT-PRES returns $\emptyset$ in order to indicate that no weak distribution bisimulation exists for the current candidate P, this result will not change $\mathbf{W}'$.

As the algorithm is based on the state-based characterisation of weak distribution bisimulation, we cannot apply the algorithm on arbitrary *PA*s directly, but only on *mec*-contracted. Therefore, we have to transform every input *PA* into a *mec*-contracted *PA* before further processing. This is done in Line 1 of procedure DECIDE, where procedure QUOTIENT-UNDER-MEC is applied. This procedure computes the quotient *PA* with respect to $=_{mec}$. Clearly, this quotient is *mec*-contracted by definition. Deciding $=_{mec}$ is very efficient [3]. Lemma 1 guarantees the soundness of this approach with respect to deciding $\approx_{\delta}$.

[5] In the following, we will treat $\mathbf{W}$ both as a set of partitions and as an equivalence relation, wherever convenient, without further mentioning.

5.1 Matching Weak Transitions, Consistency Checking, and Splitting

Before we provide explanations of the procedures FINDSPLIT and REFINE, we first discuss how to construct matching weak transitions. The following enables us to effectively compute the existence of two matching weak transitions.

Proposition 1 (cf. [13, Prop. 3]). *Given a PA $\mathcal{A}$, two sub-probability distributions $\rho_1, \rho_2 \in \mathrm{SubDisc}(S)$ such that $|\rho_1| = |\rho_2| > 0$, two actions $a_1, a_2 \in \Sigma$, two sets $\check{A}_1, \check{A}_2 \subseteq D$ of transitions, and an equivalence relation $\boldsymbol{W}$ on S, the existence of $\mu_1, \mu_2 \in \mathrm{SubDisc}(S)$ such that*

$$\rho_1 \overset{a_1 \downarrow \check{A}_1}{\Longrightarrow}_{\mathrm{c}} \mu_1,\ \rho_2 \overset{a_2 \downarrow \check{A}_2}{\Longrightarrow}_{\mathrm{c}} \mu_2, \text{ and } \mu_1\ \mathcal{L}(\boldsymbol{W})\ \mu_2$$

can be checked in polynomial time.

The proof that this check, that we denote by $P(\mathbf{W}, \rho_1, a_1, \check{A}_1, \rho_2, a_2, \check{A}_2)$, can be performed in polynomial time relies on the construction of a generalised flow problem, that in turn can be encoded into an LP-problem of polynomial size spanned by the parameters ρ_1, ρ_2, a_1, a_2, $\check{A}_1$, $\check{A}_2$, and $\mathbf{W}$. Details are given in [13] whose Prop. 3 considers $\rho_1', \rho_2' \in \mathrm{Disc}(S)$; the above proposition follows by choosing the normalised distributions $\rho_i' = \rho_i/|\rho_i|$ for $i = 1, 2$. An exponential algorithm solving this task has been given in [2].

CONSISTENCY$(P, \mathbf{W})$

```
0.1: for all (s, τ, ρ) ∈ P do
0.2:     for all (s, a, μ) ∈ 𝒯 do
0.3:         if
             P(W, μ, τ, P, ρ, a, D)
             has no solution then
0.4:             return false
0.5: return true
```

FINDSPLIT$(\mathbf{W}, P)$

```
0.1: for all (s, a, ρ) ∈ 𝒯 do
0.2:     for all t ∈ [s]_W do
0.3:         if (s, a, ρ) ∈ 𝒯_P then
0.4:             if P(W, ρ, τ, P, δ_t, a, P) has
                 no solution then
0.5:                 return ([s]_W, a, ρ)
0.6:         else
0.7:             if P(W, ρ, τ, P, δ_t, a, D) has
                 no solution then
0.8:                 return ([s]_W, a, ρ)
0.9: return (∅, τ, δ_⊥)
```

Now we are ready to explain the remaining procedures. Following the same line as for instance [2], QUOTIENT-WRT-PRES makes use of a sub-procedure REFINE, which actually creates a finer partitioning, as long as there is a partition containing two states that violate the bisimulation condition, which is checked for in procedure FINDSPLIT. More precisely, as in [2], procedure REFINE divides partition $\mathcal{C}$ into two new partitions according to the discriminating behaviour $\xrightarrow{a} \mu$, which has been identified by FINDSPLIT before. We do not provide REFINE explicitly.

In FINDSPLIT, the sets $\mathcal{T}$ and $\mathcal{T}_X \subseteq \mathcal{T}$ contain all transitions and all candidate preserving transitions, respectively, we have to match: $\mathcal{T}$ is the set of combined weak transitions and $\mathcal{T}_X$ is the set of combined candidate preserving weak transitions (defined by a scheduler using only candidate transitions in X) for state-based weak distribution bisimulation. Note that it is sufficient to use for $\mathcal{T}$ ($\mathcal{T}_X$) the set of (preserving) weak transitions defined by Dirac determinate schedulers (on preserving transitions X),

which is a finite set (cf. [2, Prop. 3, 4]). Unfortunately, this set may be exponential, which also gives rise to an overall exponential run-time complexity of the algorithm.

Both procedures FINDSPLIT and CONSISTENCY rely on Prop. 1. By verifying $P(\mathbf{W}, \mu, \tau, P, \delta_t, a, D)$ in their conditional statement, they check the corresponding conditions from Def. 4 (preserving transitions) and Def. 5 (state-based characterisation of weak distribution bisimulation), respectively.

6 Related Work

Recently, the problem of a decision algorithm for *MA* weak bisimilarity has been addressed by Schuster and Siegle [19]. The treatment uses the concept of tangible states, which seems dual to our preserving transitions in the sense that a state is tangible if and only if it has no outgoing preserving transitions. The algorithm presented is a nested fixed-point computation with exponential time complexity. It iteratively refines a candidate state partition while iteratively enlarging the set of candidate tangible states. No correctness proof is provided. A particular obstacle we see is that some of the crucial correctness arguments need to be applied to candidate partitions which by construction do not represent weak bisimulation relations (except for the last one, provided the algorithm is correct). But these arguments are established to hold only in case the partitions do indeed represent weak bisimulation relations.

7 Concluding Remarks

This paper has developed a decision algorithm for weak distribution bisimulation on probabilistic automata. It can be extended straightforwardly to Markov automata. This algorithm can be considered as the nucleus for extending the compositional specification and reasoning means in use for *IMC* to the more expressive *MA* setting. Albeit being a distribution-based relation, we managed to circumvent uncountability in the carrier set by a state-based characterisation. The main obstacle has not been the issue of finding an alternative characterisation of $\approx_\delta$ and deriving a decision algorithm from there. Rather, the formal proof that the characterisation is indeed equivalent to the one of [9] has been very challenging. As Ex. 6 and Ex. 7 show, the pitfalls are hidden in seemingly obvious places. The presented algorithm uses worst-case exponential time and polynomial space, and we are investigating its theoretical and practical runtime characteristics further.

Acknowledgements. This work is supported by the DFG/NWO bilateral research programme ROCKS, by the DFG as part of the SFB/TR 14 AVACS, by the EU FP7 Programme under grant agreement no. 295261 (MEALS), 318490 (SENSATION), and 318003 (TREsPASS), by IDEA4CPS and MT-LAB, a VKR Centre of Excellence. Andrea Turrini is supported by the Cluster of Excellence "Multimodal Computing and Interaction" (MMCI), part of the German Excellence Initiative.

References

1. C. Baier, B. Engelen, and M. E. Majster-Cederbaum. Deciding bisimilarity and similarity for probabilistic processes. *J. Comput. Syst. Sci.*, 60(1):187–231, 2000.
2. S. Cattani and R. Segala. Decision algorithms for probabilistic bisimulation. In *CONCUR*, volume 2421 of *LNCS*, pages 371–385, 2002.
3. K. Chatterjee and M. R. Henzinger. Faster and dynamic algorithms for maximal end-component decomposition and related graph problems in probabilistic verification. In *SODA*, pages 1318–1336, 2011.
4. L. de Alfaro. *Formal Verification of Probabilistic Systems*. PhD thesis, Stanford University, 1997.
5. Y. Deng and M. Hennessy. On the semantics of Markov automata. In *ICALP*, pages 307–318, 2011.
6. Y. Deng and M. Hennessy. On the semantics of Markov automata. *I&C*, 222:139–168, 2012.
7. A. Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundamenta Mathematicae*, 49:129–144, 1961.
8. C. Eisentraut, H. Hermanns, and L. Zhang. Concurrency and composition in a stochastic world. In *CONCUR*, volume 6269 of *LNCS*, pages 21–39, 2010.
9. C. Eisentraut, H. Hermanns, and L. Zhang. On probabilistic automata in continuous time. In *LICS*, pages 342–351, 2010.
10. C. Eisentraut, H. Hermanns, and L. Zhang. On probabilistic automata in continuous time. Reports of SFB/TR 14 AVACS 62, SFB/TR 14 AVACS, 2010.
11. M. Hennessy. Exploring probabilistic bisimulations, part I. *Formal Aspects of Computing*, 24(4-6):749–768, 2012.
12. H. Hermanns. *Interactive Markov Chains: The Quest for Quantified Quality*, volume 2428 of *LNCS*. Springer, 2002.
13. H. Hermanns and A. Turrini. Deciding probabilistic automata weak bisimulation in polynomial time. In *FSTTCS*, pages 435–447, 2012.
14. P. C. Kanellakis and S. A. Smolka. CCS expressions, finite state processes, and three problems of equivalence. *I&C*, 86(1):43–68, 1990.
15. K. G. Larsen and A. Skou. Bisimulation through probabilistic testing (preliminary report). In *POPL*, pages 344–352, 1989.
16. N. A. Lynch, R. Segala, and F. W. Vaandrager. Compositionality for probabilistic automata. In *CONCUR*, volume 2761 of *LNCS*, pages 208–221, 2003.
17. N. A. Lynch, R. Segala, and F. W. Vaandrager. Observing branching structure through probabilistic contexts. *SIAM J. on Computing*, 37(4):977–1013, 2007.
18. A. Philippou, I. Lee, and O. Sokolsky. Weak bisimulation for probabilistic systems. In *CONCUR*, volume 1877 of *LNCS*, pages 334–349, 2000.
19. J. Schuster and M. Siegle. Markov automata: Deciding weak bisimulation by means of "non-naïvely" vanishing states. Available at `http://arxiv.org/abs/1205.6192`.
20. R. Segala. *Modeling and Verification of Randomized Distributed Real-Time Systems*. PhD thesis, MIT, 1995.
21. R. Segala. Probability and nondeterminism in operational models of concurrency. In *CONCUR*, volume 4137 of *LNCS*, pages 64–78, 2006.
22. C. Stirling. Local model checking games (extended abstract). In *CONCUR*, volume 962 of *LNCS*, pages 1–11.
23. W. Thomas. On the Ehrenfeucht-Fraïssé game in theoretical computer science. In *TAPSOFT*, volume 668 of *LNCS*, pages 559–568, 1993.