Confluence Reduction for Markov Automata

Mark Timmer, Jaco van de Pol, and Mariëlle Stoelinga*

Formal Methods and Tools, Faculty of EEMCS
University of Twente, The Netherlands
{timmer, vdpol, marielle}@cs.utwente.nl

Abstract. Markov automata are a novel formalism for specifying systems exhibiting nondeterminism, probabilistic choices and Markovian rates. Recently, the process algebra MAPA was introduced to efficiently model such systems. As always, the state space explosion threatens the analysability of the models generated by such specifications. We therefore introduce confluence reduction for Markov automata, a powerful reduction technique to keep these models small. We define the notion of confluence directly on Markov automata, and discuss how to syntactically detect confluence on the MAPA language as well. That way, Markov automata generated by MAPA specifications can be reduced on-the-fly while preserving divergence-sensitive branching bisimulation. Three case studies demonstrate the significance of our approach, with reductions in analysis time up to an order of magnitude.

1 Introduction

Over the past two decades, model checking algorithms were generalised to handle more and more expressive models. This now allows us to verify probabilistic as well as hard and soft real-time systems, modelled by timed automata, Markov decision processes, probabilistic automata, continuous-time Markov chains, interactive Markov chains, and Markov automata. Except for timed automata—which incorporate real-time deadlines—all other models are subsumed by the Markov automaton (MA) [14, 13, 12]. MAs can therefore be used as a semantic model for a wide range of formalisms, such as generalised stochastic Petri nets (GSPNs) [2], dynamic fault trees [9], Arcade [8] and the domain-specific language AADL [10].

Before the introduction of MAs, the above models could not be analysed to their full extent. For instance, the semantics of a (potentially nondeterministic) GSPN were given as a fully probabilistic CTMC. To this end, weights had to be assigned to resolve the nondeterminism between immediate transitions. As argued in [20], it is often much more natural to omit most of these weights, retaining rates and probability as well as nondeterminism, and thus obtaining an MA. For example, consider the GSPN in Figure 1(a), taken from [13]. Immediate

* This research has been partially funded by NWO under grants 612.063.817 (SYRUP), 12238 (ArRangeer) and Dn 63-257 (ROCKS), and EU under 318490 (SENSATION).

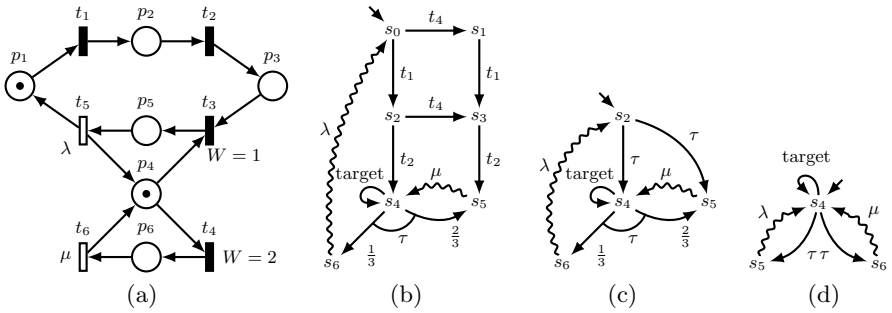


Fig. 1. A GSPN and the corresponding unreduced and reduced state spaces.

transitions are indicated in black, Markovian transitions in white, and we assume a partial weight assignment. The underlying MA is given in Figure 1(b), where s_0 corresponds to the initial situation with one token in p_1 and p_4. We added a selfloop labelled *target* to indicate a possible state of interest s_4 (having one token in p_3 and p_4), and for convenience labelled the interactive transitions of the MA by the immediate transition of the GSPN they resulted from (except for the probabilistic transition, which is the result of t_3 and t_4 together).

Recently, the data-rich process-algebraic language MAPA was introduced to efficiently specify MAs in a compositional manner [24]. As always, though, the state space explosion threatens the feasibility of model checking, especially in the presence of data and interleaving. Therefore, reduction techniques for MAs are vital to keep the state spaces of these models manageable. In this paper we introduce such a technique, generalising *confluence reduction* to MAs. It is a powerful state space reduction technique based on commutativity of transitions, removing spurious nondeterminism often arising from the parallel composition of largely independent components. To the best of our knowledge, it is the first technique of this kind for MAs. We give heuristics to apply confluence reduction directly on specifications in the MAPA language, reducing them on-the-fly while preserving divergence-sensitive branching bisimulation.

To illustrate confluence reduction, reconsider the MA in Figure 1(b) and assume that $t_1 = t_2 = t_4 = \tau$, i.e., all action-labelled transitions, except for the *target*-transition, are invisible. We are able to detect automatically that the t_1-transitions are confluent; they can thus safely be given priority over t_4, without losing any behaviour. Figure 1(c) shows the reduced state space, generated on-the-fly using confluence reduction. If all weights are omitted from the specification, an even smaller reduced state space is obtained (Figure 1(d)), while the only change in the unreduced state space is the substitution of the probabilistic choice by a nondeterministic choice.

Outline of the approach. First, we introduce the technical background of our work (Section 2). Then, we define our novel notion of confluence for MAs (Section 3). It specifies sufficient conditions for invisible transitions to not alter the

behaviour of an MA; i.e., if a transition is confluent, it could be given priority over all other transitions with the same source state.

We formally show that confluent transitions connect divergence-sensitive branching bisimilar states, and present a mapping of states to representatives to efficiently generate a reduced MA based on confluence (Section 4). We discuss how confluence can be detected symbolically on specifications in the MAPA language (Section 5) and illustrate the significance of our technique using three case studies (Section 6). We show state spaces shrinking by more than 80%, making the entire process from MAPA specification to results more than ten times as fast for some models.[1]

Related work. Confluence reduction for process algebras was first introduced for non-probabilistic systems [7], and later for probabilistic automata [25]. Also, several types of *partial order reduction* (POR) have been defined, both for non-probabilistic [26, 21, 16] and probabilistic systems [11, 4, 3]. These techniques are based on ideas similar to confluence, and have been compared to confluence recently, both in a theoretical [17] and in a practical manner [18]. The results showed that branching-time POR is strictly subsumed by confluence, and that the additional advantages of confluence can be employed nicely in the context of statistical model checking.

Compared to the earlier approaches to confluence reduction for process algebras [7, 25], our novel notion of confluence is different in three important ways:

- It can handle MAs, and hence is applicable to a larger class of systems.
- It fixes a subtle flaw in the earlier work, by introducing an underlying classification of the interactive transitions. This way we guarantee closure under unions, something that was not guaranteed before. It is key to the way we detect confluence on MAPA specifications.
- We now do preserve divergences and hence minimal reachability probabilities, incorporating a technique used earlier in [17].

Since none of the existing techniques is able to deal with MAs, we believe that our generalisation—the first reduction technique for MAs abstracting from internal transitions—is a major step forward in efficient quantitative verification.

2 Preliminaries

Definition 1 (Basics). *A* probability distribution *over a countable set S is a function $\mu\colon S \to [0,1]$ such that $\sum_{s\in S} \mu(s) = 1$. For $S' \subseteq S$, let $\mu(S') = \sum_{s\in S'} \mu(s)$. We define* $\mathrm{spt}(\mu) = \{s \in S \mid \mu(s) > 0\}$ *to be the* support *of μ, and write $\mathbb{1}_s$ for the* Dirac distribution *for s, determined by $\mathbb{1}_s(s) = 1$.*

We use $\mathrm{Distr}(S)$ *to denote the set of all probability distributions over S, and* $\mathrm{SDistr}(S)$ *for the set of all* substochastic *probability distributions over S, i.e.,*

[1] Due to space limitations, we discuss the notion of divergence-sensitive branching bisimulation only on an intuitive level, deferring the formal definitions and proofs of all our results to the appendices. These will be published in a technical report.

where $0 \leq \sum_{s \in S} \mu(s) \leq 1$. *Given a function* f, *we denote by* $f(\mu)$ *the* lifting *of* μ *over* f, *i.e.,* $f(\mu)(s) = \mu(f^{-1}(s))$, *with* $f^{-1}(s)$ *the inverse image of* s *under* f.

Given an equivalence relation $R \subseteq S \times S$, *we write* $[s]_R$ *for the equivalence class of* s *induced by* R, *i.e.,* $[s]_R = \{s' \in S \mid (s, s') \in R\}$. *We denote the set of all such equivalence classes by* S/R. *Given two probability distributions* $\mu, \mu' \in \text{Distr}(S)$ *and an equivalence relation* R, *we write* $\mu \equiv_R \mu'$ *to denote that* $\mu([s]_R) = \mu'([s]_R)$ *for every* $s \in S$.

An MA is a transition system in which the set of transitions is partitioned into probabilistic interactive transitions (equivalent to the transitions of a PA), and Markovian transitions labelled by the rate of an exponential distribution (equivalent to the transitions of a CTMC). We assume a countable universe of actions *Act*, with $\tau \in Act$ the invisible internal action.

Definition 2 (Markov automata). *A* Markov automaton (MA) *is a tuple* $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$, *where*

- S *is a countable set of* states, *of which* $s^0 \in S$ *is the* initial state;
- $A \subseteq Act$ *is a countable set of* actions;
- $\hookrightarrow \subseteq S \times A \times \text{Distr}(S)$ *is the* interactive transition relation;
- $\rightsquigarrow \subseteq S \times \mathbb{R}_{>0} \times S$ *is the* Markovian transition relation.

If $(s, a, \mu) \in \hookrightarrow$, *we write* $s \xhookrightarrow{a} \mu$ *and say that the action* a *can be* executed *from state* s, *after which the probability to go to* $s' \in S$ *is* $\mu(s')$. *If* $(s, \lambda, s') \in \rightsquigarrow$, *we write* $s \overset{\lambda}{\rightsquigarrow} s'$ *and say that* s *moves to* s' *with rate* λ.

The *rate* between two states $s, s' \in S$ is $rate(s, s') = \sum_{(s,\lambda,s') \in \rightsquigarrow} \lambda$, and the *outgoing rate of* s is $rate(s) = \sum_{s' \in S} rate(s, s')$. We require $rate(s) < \infty$ for every state $s \in S$. If $rate(s) > 0$, the *branching probability distribution* after this delay is denoted by $\mathbb{P}_s$ and defined by $\mathbb{P}_s(s') = \frac{rate(s,s')}{rate(s)}$ for every $s' \in S$.

By definition of the exponential distribution, the probability of leaving a state s within t time units is given by $1 - e^{-rate(s) \cdot t}$ (given $rate(s) > 0$), after which the next state is chosen according to $\mathbb{P}_s$.

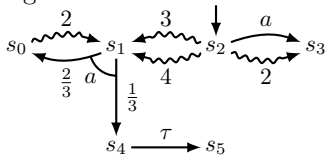
MAs adhere to the *maximal progress assumption*, prescribing τ-transitions to never be delayed. Hence, a state that has at least one outgoing τ-transition can never take a Markovian transition. This fact is captured below in the definition of extended transitions.

Definition 3 (Extended action set). *Let* $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$ *be an MA, then the* extended action set of $\mathcal{M}$ *is given by* $A^\chi = A \cup \{\chi(r) \mid r \in \mathbb{R}_{>0}\}$. *Given a state* $s \in S$ *and an action* $\alpha \in A^\chi$, *we write* $s \xrightarrow{\alpha} \mu$ *if either*

- $\alpha \in A$ *and* $s \xhookrightarrow{\alpha} \mu$, *or*
- $\alpha = \chi(rate(s))$, $rate(s) > 0$, $\mu = \mathbb{P}_s$ *and there is no* μ' *such that* $s \xhookrightarrow{\tau} \mu'$.

A transition $s \xrightarrow{\alpha} \mu$ *is called an* extended transition. *We write* $s \xrightarrow{\alpha,\mu} s'$ *if there is an extended transition* $s \xrightarrow{\alpha} \mu$ *such that* $\mu(s') > 0$. *We use* $s \xrightarrow{\alpha} t$ *to denote* $s \xrightarrow{\alpha} \mathbb{1}_t$, *and write* $s \to t$ *if there is at least one action* α *such that* $s \xrightarrow{\alpha} t$.

Example 4. Consider the MA $\mathcal{M}$ shown on the right.



For this system, $\mathit{rate}(s_2, s_1) = 3 + 4 = 7$, $\mathit{rate}(s_2) = 7 + 2 = 9$, and $\mathbb{P}_{s_2} = \mu$ such that $\mu(s_1) = \frac{7}{9}$ and $\mu(s_3) = \frac{2}{9}$. There are two extended transitions from s_2: $s_2 \xrightarrow{a} \mathbb{1}_{s_3}$ (also written as $s_2 \xrightarrow{a} s_3$) and $s_2 \xrightarrow{\chi(9)} \mathbb{P}_{s_2}$. □

We define several notions for paths and connectivity. These are based on extended transitions, and thus may contain interactive as well as Markovian steps.

Definition 5 (Paths). *Given an MA $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$,*

- *A* path *in $\mathcal{M}$ is a finite sequence $\pi^{\mathrm{fin}} = s_0 \xrightarrow{a_1,\mu_1} s_1 \xrightarrow{a_2,\mu_2} \ldots \xrightarrow{a_n,\mu_n} s_n$ from some state s_0 to a state s_n ($n \geq 0$), or an infinite sequence $\pi^{\mathrm{inf}} = s_0 \xrightarrow{a_1,\mu_1} s_1 \xrightarrow{a_2,\mu_2} s_2 \xrightarrow{a_3,\mu_3} \ldots$, with $s_i \in S$ for all $0 \leq i \leq n$ and all $0 \leq i$, respectively. We use $\mathit{prefix}(\pi, i)$ to denote $s_0 \xrightarrow{a_1,\mu_1} \ldots \xrightarrow{a_i,\mu_i} s_i$, and $\mathit{step}(\pi, i)$ for the transition $s_{i-1} \xrightarrow{a_i} \mu_i$. When π is finite we define $|\pi| = n$ and $\mathit{last}(\pi) = s_n$. We use $\mathit{finpaths}_{\mathcal{M}}$ for the set of all* finite paths *in $\mathcal{M}$, and $\mathit{finpaths}_{\mathcal{M}}(s)$ for all such paths with $s_0 = s$.*
- *We denote by $\mathit{trace}(\pi)$ the sequence of actions of π while omitting all τ-actions, and use ϵ to denote the empty sequence.*

Definition 6 (Connectivity). *Given an MA $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$ and two states $s, t \in S$, we write*

- $s \twoheadrightarrow t$ (reachability) *if there is a path from s to t;*
- $s \twoheadrightarrow \twoheadleftarrow t$ (joinability) *if there is a state u such that $s \twoheadrightarrow u$ and $t \twoheadrightarrow u$.*

We define $\twoheadleftarrow\!\!\twoheadrightarrow$ (convertibility) *as the symmetric and transitive closure of $\twoheadrightarrow$.*

Since a single state is a path as well, all three connectivity relations are reflexive. Additionally, the relation $\twoheadrightarrow$ is transitive (but not necessarily symmetric) and the relation $\twoheadrightarrow \twoheadleftarrow$ is symmetric (but not necessarily transitive). Note that, intuitively, $s \twoheadleftarrow\!\!\twoheadrightarrow t$ means that s is connected by extended transitions to t—disregarding the orientation of these transitions. Clearly, $s \twoheadrightarrow t$ implies $s \twoheadrightarrow \twoheadleftarrow t$, and $s \twoheadrightarrow \twoheadleftarrow t$ implies $s \twoheadleftarrow\!\!\twoheadrightarrow t$. These implications in general do not hold the other way.

Example 7. The system in Example 4 has infinitely many paths, for example

$$\pi = s_2 \xrightarrow{\chi(9),\mu_1} s_1 \xrightarrow{a,\mu_2} s_0 \xrightarrow{\chi(2),\mathbb{1}_{s_1}} s_1 \xrightarrow{a,\mu_2} s_4 \xrightarrow{\tau,\mathbb{1}_{s_5}} s_5$$

with $\mu_1(s_1) = \frac{7}{9}$ and $\mu_1(s_3) = \frac{2}{9}$, and $\mu_2(s_0) = \frac{2}{3}$ and $\mu_2(s_4) = \frac{1}{3}$. We have $\mathit{prefix}(\pi, 2) = s_2 \xrightarrow{\chi(9),\mu_1} s_1 \xrightarrow{a,\mu_2} s_0$, and $\mathit{step}(\pi, 2) = s_1 \xrightarrow{a} \mu_2$. Also, $\mathit{trace}(\pi) = \chi(9)\, a\, \chi(2)\, a$. It is easy to see that $s_2 \twoheadrightarrow s_5$, as well as $s_0 \twoheadrightarrow \twoheadleftarrow s_2$ and $s_5 \twoheadleftarrow\!\!\twoheadrightarrow s_3$. However, $s_0 \twoheadrightarrow s_2$ and $s_5 \twoheadrightarrow \twoheadleftarrow s_3$ do not hold. □

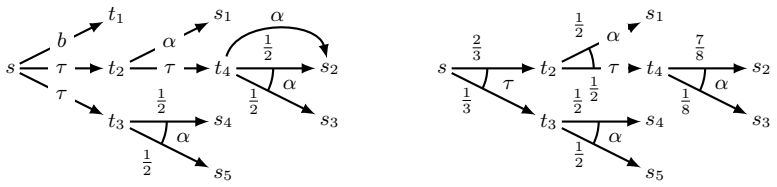


Fig. 2. An MA (left), and a tree demonstrating the branching transition $s \xRightarrow{\alpha} \mu$ (right).

2.1 Divergence-sensitive branching bisimulation

To prove our confluence reduction technique correct, we show that it preserves divergence-sensitive branching bisimulation. Basically, this means that there is an equivalence relation R linking states in the original system to states in the reduced system, in such a way that their initial states are related and all related states can mimic each other's transitions and divergences.

More precisely, for R to be a divergence-sensitive branching bisimulation, it is required that for all $(s,t) \in R$ and every extended transition $s \xrightarrow{a} \mu$, there is a *branching transition* $t \xRightarrow{a}_R \mu'$ such that $\mu \equiv_R \mu'$. The existence of such a branching transition depends on the existence of a certain *scheduler*. Schedulers resolve nondeterministic choices in an MA by selecting which transition to take given a history; they are also allowed to terminate with some probability.

Now, a state t can do a branching transition $t \xRightarrow{a}_R \mu'$ if either (1) $a = \tau$ and $\mu' = \mathbb{1}_t$, or (2) there exists a scheduler that terminates according to μ', always schedules precisely one a-transition (immediately before terminating), does not schedule any other visible transitions and does not leave the equivalence class $[t]_R$ before doing an a-transition.

Example 8. Observe the MA in Figure 2 (left). We find that $s \xRightarrow{\alpha} \mu$, with

$$\mu(s_1) = \tfrac{8}{24} \qquad \mu(s_2) = \tfrac{7}{24} \qquad \mu(s_3) = \tfrac{1}{24} \qquad \mu(s_4) = \tfrac{4}{24} \qquad \mu(s_5) = \tfrac{4}{24}$$

by the scheduling depicted in Figure 2 (right), assuming $(s, t_i) \in R$ for all t_i. □

In addition to the mimicking of transitions by branching transitions, we require R-related states to either both be able to perform an infinite invisible path with probability 1 (*diverge)*, or to both not be able to do so. We write $s \leftrightarrows_{\mathrm{b}}^{\mathrm{div}} t$ if two states s, t are divergence-sensitive branching bisimilar, and $\mathcal{M}_1 \leftrightarrows_{\mathrm{b}}^{\mathrm{div}} \mathcal{M}_2$ if two MAs are (i.e., if their initial states are so in their disjoint union).

3 Confluence for Markov automata

In [25] we defined three variants of probabilistic confluence: weak probabilistic confluence, probabilistic confluence and strong probabilistic confluence. They specify sufficient conditions for τ-transitions to not alter the behaviour of an MA. The stronger notions are easier to detect, but less powerful in their reductions.

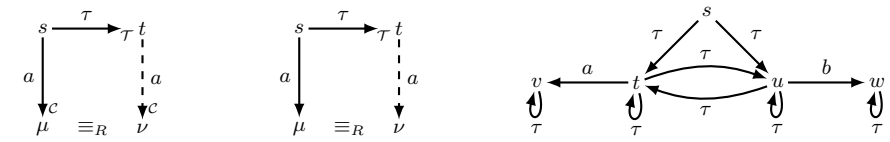
In a process-algebraic context, where confluence is detected heuristically over a syntactic description of a system, it is most practical to apply strong confluence. Therefore, in this paper we only generalise strong probabilistic confluence to the Markovian realm. Although MAs in addition to interactive transitions may also contain Markovian transitions, these are irrelevant for confluence. After all, states having a τ-transition can never execute a Markovian transition due to the maximal progress assumption. Hence, such transitions need not be mimicked. For the above reasons, the original definition of confluence for PAs might seem to still work for MAs. This is not true, however, for two reasons.

1. The old definition was not yet divergence sensitive and hence might lose divergences; for MAs it could therefore erroneously enable Markovian transitions that were disabled in the presence of divergence due to the maximal progress assumption. Hence, it would not even preserve Markovian divergence-*in*sensitive branching bisimulation. We now improve on the definition to resolve this issue, introducing τ-loops in the reduced system for states having confluent divergence in the original system (inspired by the way [17] deals with divergences). This not only makes the theory work for MAs, it even yields preservation of divergence-sensitive branching bisimulation, and hence of minimal reachability probabilities.
2. The old definition had a subtle flaw: earlier work relied on the assumption that confluent sets are closed under unions [7, 25]. In practical applications this was indeed a valid assumption, but for the theoretical notions of confluence this was not yet the case. We fix this flaw by classifying transitions into groups, defining confluence over sets of such groups and requiring transitions to be mimicked by a transition from their own group.

Additionally, we improve on the way equivalence of distributions is defined, making it slightly more powerful and, in our view, easier to understand.

Confluence classifications and confluent sets. The original lack of closure under unions was due to the requirement that confluent transitions are mimicked by confluent transitions. When taking the union of two valid sets of confluent transitions, this requirement was possibly invalidated. To solve this problem, we classify the interactive transitions of an MA into groups—allowing overlap and not requiring all interactive transitions to be in at least one group. Together, we call such a set of groups $P = \{C_1, C_2, \ldots, C_n\} \subseteq \mathscr{P}(\hookrightarrow)$ a *confluence classification*[2]. Now, instead of designating individual transitions to be confluent and requiring confluent transitions to be mimicked by confluent transitions, we designate groups in P to be confluent and require transitions from a group in P to be mimicked by transitions from the same group.

[2] We use $s \xrightarrow{a}_C \mu$ to denote that $(s \xrightarrow{a} \mu) \in C$, and abuse notation by writing $(s \xrightarrow{a} \mu) \in P$ to denote that $s \xrightarrow{a}_C \mu$ for some $C \in P$. Similarly, we subscript reachability, joinability and convertibility arrows to indicate that they only traverse transitions from a certain group or set of groups of transitions.



(a) $(s \xrightarrow{a} \mu) \in P$. (b) $(s \xrightarrow{a} \mu) \notin P$. (c) A simple state space.

Fig. 3. The confluence diagrams for $s \xrightarrow{\tau}_{\mathcal{T}} t$, and a simple state space. In (a,b): If the solid transitions are present, then so should the dashed ones be.

For a set $\mathcal{T} \subseteq P$ to be *Markovian confluent*, first of all—like in the PA setting [25, 3]—it is only allowed to contain invisible transitions with a Dirac distribution. (Still, prioritising such transitions may very well reduce probabilistic transitions as well, as we will see in Section 4.) Additionally, each transition $s \xrightarrow{a} \mu$ enabled before a transition $s \xrightarrow{\tau}_{\mathcal{T}} t$ should have a mimicking transition $t \xrightarrow{a} \nu$ such that μ and ν are connected by $\mathcal{T}$-transitions, and mimicking transitions should be from the same group. The definition is illustrated in Figure 3.

Definition 9 (Markovian confluence). *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$ be an MA and $P \subseteq \mathscr{P}(\hookrightarrow)$ a confluence classification. Then, a set $\mathcal{T} \subseteq P$ is* Markovian confluent for P *if it only contains sets of invisible transitions with Dirac distributions, and for all $s \xrightarrow{\tau}_{\mathcal{T}} t$ and all transitions $(s \xrightarrow{a} \mu) \neq (s \xrightarrow{\tau} t)$:*

$$\begin{cases} \forall C \in P \,.\, s \xrightarrow{a}_C \mu \implies \exists \nu \in \mathrm{Distr}(S) \,.\, t \xrightarrow{a}_C \nu \wedge \mu \equiv_R \nu & , \text{if } (s \xrightarrow{a} \mu) \in P \\ \qquad\qquad\quad\; \exists \nu \in \mathrm{Distr}(S) \,.\, t \xrightarrow{a} \nu \wedge \mu \equiv_R \nu & , \text{if } (s \xrightarrow{a} \mu) \notin P \end{cases}$$

with R the smallest equivalence relation such that

$$R \supseteq \{(s,t) \in \mathrm{spt}(\mu) \times \mathrm{spt}(\nu) \mid (s \xrightarrow{\tau} t) \in \mathcal{T}\}.$$

A transition $s \xrightarrow{\tau} t$ is Markovian confluent *if there exists a Markovian confluent set $\mathcal{T}$ such that $s \xrightarrow{\tau}_{\mathcal{T}} t$.*

Note that $\mu \equiv_R \nu$ requires direct transitions from the support of μ to the support of ν. Also note that, even though a (symmetric) equivalence relation R is used, transitions from the support of ν to the support of μ do not influence R.

Remark 10. Due to the confluence classification, confluent transitions are always mimicked by confluent transitions. After all, transitions from a group $C \in P$ are mimicked by transitions from C. So, if C is designed confluent by $\mathcal{T}$, then all these confluent transitions are indeed mimicked by confluent transitions.

Although the confluence classification may appear restrictive, we will see that in practice it is obtained naturally. Transitions are often instantiations of higher-level constructs, and are therefore easily grouped together. Then, it makes sense to detect the confluence of such a higher-level construct. Additionally, to show that a certain set of interactive transitions is confluent, we can just take P to consists of one group containing precisely all those transitions. Then, the requirement for P-transitions to be mimicked by the same group reduces to the old requirement that confluent transitions are mimicked by confluent transitions.

Properties of confluent sets. Since confluent transitions are always mimicked by confluent transitions, confluent paths (i.e., paths following only transitions from a confluent set) are always joinable.

Proposition 11. *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$ be an MA, $P \subseteq \mathscr{P}(\hookrightarrow)$ a confluence classification for $\mathcal{M}$ and $\mathcal{T}$ a Markovian confluent set for P. Then,*

$$s \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} t \qquad \text{if and only if} \qquad s \twoheadleftarrow\!\!\!\!\twoheadrightarrow_{\mathcal{T}} t$$

Due to the confluence classification, we now also do have a closure result. Closure under union tells us that it is safe to show confluence of multiple sets of transitions in isolation, and then just take their union as one confluent set. Also, it implies that there exists a unique maximal confluent set.

Proposition 12. *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$ be an MA, $P \subseteq \mathscr{P}(\hookrightarrow)$ a confluence classification for $\mathcal{M}$ and $\mathcal{T}_1, \mathcal{T}_2$ two Markovian confluent sets for P. Then, $\mathcal{T}_1 \cup \mathcal{T}_2$ is also a Markovian confluent set for P.*

The next example shows why Proposition 12 would not hold without the use of a confluence classification. It applies to the old notions of confluence as well.

Example 13. Consider the system in Figure 3(c). Without the requirement that transitions are mimicked by the same group, the sets

$$\mathcal{T}_1 = \{(s, \tau, u), (t, \tau, t), (u, \tau, u), (v, \tau, v), (w, \tau, w)\}$$
$$\mathcal{T}_2 = \{(s, \tau, t), (t, \tau, t), (u, \tau, u), (v, \tau, v), (w, \tau, w)\}$$

would both be perfectly valid confluent sets. Still, $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2$ is not an acceptable set. After all, whereas $t \twoheadleftarrow\!\!\!\!\twoheadrightarrow_{\mathcal{T}} u$, it fails to satisfy $t \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} u$. This property was ascertained in earlier work by requiring confluent transitions to be mimicked by confluent transitions or by explicitly requiring $\twoheadrightarrow \twoheadleftarrow_{\mathcal{T}}$ to be an equivalence relation. This is indeed not the case for $\mathcal{T}$, as the diamond starting with $s \xrightarrow{\tau} t$ and $s \xrightarrow{\tau} u$ can only be closed using the non-confluent transitions between t and u, and clearly $\twoheadrightarrow \twoheadleftarrow$ is not transitive. However, $\mathcal{T}_1$ and $\mathcal{T}_2$ do satisfy these requirements, and hence the old notions were not closed under union.

By using a confluence classification and requiring transitions to be mimicked by the same group, we ascertain that this kind of bad compositionality behaviour does not occur. After all, for $\mathcal{T}_1$ to be a valid confluent set, the confluence classification should be such that $s \xrightarrow{\tau} t$ and its mimicking transition $u \xrightarrow{\tau} t$ are in the same group. So, for $s \xrightarrow{\tau} t$ to be confluent (as prescribed by $\mathcal{T}_2$), also $u \xrightarrow{\tau} t$ would need to be confluent. The latter is impossible, since the b-transition from u cannot be mimicked from t, and hence $\mathcal{T}_2$ is disallowed. □

The final result of this section shows that confluent transitions indeed connect divergence-sensitive bisimilar states. This is a key result; it implies that confluent transitions can be given priority over other transitions without losing behaviour.

Theorem 14. *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$ be an MA, $s, s' \in S$ two of its states, $P \subseteq \mathscr{P}(\hookrightarrow)$ a confluence classification for $\mathcal{M}$ and $\mathcal{T}$ a Markovian confluent set for P. Then,*

$$s \twoheadleftarrow\!\!\!\!\twoheadrightarrow_{\mathcal{T}} s' \text{ implies } s \leftrightarrows^{\mathrm{div}}_{\mathrm{b}} s'.$$

4 State space reduction using confluence

We can reduce state spaces by prioritising confluent transitions, i.e., omitting all other transitions from a state that also enables a confluent transition. Better still, we aim at omitting all intermediate states on a confluent path altogether; after all, they are all bisimilar anyway by Theorem 14. Confluence even dictates that all visible transitions and divergences enabled from a state s can directly be mimicked from another state t if $s \twoheadrightarrow_{\mathcal{T}} t$. Hence, we can just keep following a confluent path and only retain the last state. To avoid getting stuck in an infinite confluent loop, we detect entering a bottom strongly connected component (BSCC) of confluent transitions and choose a unique *representative* from this BSCC for all states that can reach it. Since we showed that confluent joinability is transitive (Proposition 11), it follows immediately that all confluent paths emanating from a certain state s always end up in one unique BSCC.

Formally, we use the notion of a *representation map*, assigning a representative state $\varphi(s)$ to every state s, while making sure that $\varphi(s)$ indeed exhibits all behaviour of s due to being in a BSCC reachable from s.

Definition 15 (Representation map). *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow\rangle$ be an MA and $\mathcal{T}$ a Markovian confluent set for $\mathcal{M}$. Then, a function $\varphi_{\mathcal{T}} \colon S \to S$ is a* representation map *for $\mathcal{M}$ under $\mathcal{T}$ if for all $s, s' \in S$*

- $s \twoheadrightarrow_{\mathcal{T}} \varphi_{\mathcal{T}}(s)$
- $s \to_{\mathcal{T}} s' \implies \varphi_{\mathcal{T}}(s) = \varphi_{\mathcal{T}}(s')$

Note that the first requirements ensures that every representative is reachable by all states it represents, while the second takes care that all $\mathcal{T}$-related states have the same representative. Together, they imply that every representative is in a BSCC. Since all $\mathcal{T}$-related states have the same BSCC, as discussed above, it is indeed always possible to find such a representation map. We refer to [6] for the algorithm we use to construct it in our implementation.

As representatives exhibit all behaviour of the states they represent, they can be used for state space reduction. More precisely, it is possible to define the quotient of an MA modulo a representation map. This system does not have any $\mathcal{T}$-transitions anymore, except for self-loops on representatives that have outgoing $\mathcal{T}$-transitions in the original system. These ensure preservation of divergences.

Definition 16 ($\mathcal{M}/\varphi$). *Given an MA $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow\rangle$, a confluent set $\mathcal{T}$ for $\mathcal{M}$, and a representation map $\varphi \colon S \to S$ for $\mathcal{M}$ under $\mathcal{T}$, the quotient of $\mathcal{M}$ modulo φ is the smallest system $\mathcal{M}/\varphi = \langle \varphi(S), \varphi(s^0), A, \hookrightarrow_{\varphi}, \rightsquigarrow_{\varphi}\rangle$ such that*

- $\varphi(S) = \{\varphi(s) \mid s \in S\}$;
- $\varphi(s) \xhookrightarrow{a}_{\varphi} \varphi(\mu)$ *if* $\varphi(s) \xhookrightarrow{a} \mu$;
- $\varphi(s) \overset{\lambda}{\rightsquigarrow}_{\varphi} \varphi(s')$ *if* $\lambda = \sum_{\lambda' \in \Lambda(s,s')} \lambda'$ *and* $\lambda > 0$,

where $\Lambda(s, s')$ is the multiset $\{\!|\lambda' \in \mathbb{R} \mid \exists s^* \in S \,.\, \varphi(s) \overset{\lambda'}{\rightsquigarrow} s^* \wedge \varphi(s^*) = \varphi(s')|\!\}$.

Note that each interactive transition from $\varphi(s)$ in $\mathcal{M}$ is lifted to $\mathcal{M}/\varphi$ by changing all states in the support of its target distribution to their representatives. Additionally, each pair $\varphi(s), \varphi(s')$ of representative states in $\mathcal{M}/\varphi$ has a connecting Markovian transition with rate equal to the total outgoing rate of $\varphi(s)$ in $\mathcal{M}$ to states s^* that have $\varphi(s')$ as their representative (unless this sum is 0). It is easy to see that this implies $\varphi(s) \xrightarrow{\chi(\lambda)}_{\varphi} \varphi(\mu)$ if and only if $\varphi(s) \xrightarrow{\chi(\lambda)} \mu$.

Since $\mathcal{T}$-transitions connect bisimilar states, and representatives exhibit all behaviour of the states they represent, we can prove the following theorem. It shows that we indeed reached our goal of providing a reduction that is safe with respect to divergence-sensitive branching bisimulation.

Theorem 17. *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$ be an MA, $\mathcal{T}$ a Markovian confluent set for $\mathcal{M}$, and $\varphi\colon S \to S$ a representation map for $\mathcal{M}$ under $\mathcal{T}$. Then,*

$$\mathcal{M}/\varphi \leftrightarrows_{\mathrm{b}}^{\mathrm{div}} \mathcal{M}.$$

5 Symbolic detection of Markovian confluence

Although the definition of confluence in Section 3 is useful to show the correctness of our approach, it is often not feasible to check in practice. After all, we want to reduce *on-the-fly* to obtain a smaller state space without first generating the unreduced one. Therefore, we use heuristics to detect Markovian confluence in the context of the process-algebraic modelling language MAPA [24]. As these heuristics only differ slightly from the ones in [25] for probabilistic confluence, we discuss the basics and explain how the old techniques can be reused.

MAPA is data-rich and expressive, and features a restricted form: the Markovian Linear Probabilistic Process Equation (MLPPE). Every MAPA specification can be translated easily to an equivalent specification in MLPPE [24]. Hence, it suffices to define our confluence-based reduction technique on this form.

The MLPPE format. An MLPPE is a process with global variables, *interactive summands* (each yielding a set of interactive transitions) and *Markovian summands* (each yielding a set of Markovian transitions). Its semantics is given as an MA, whose states are all valuations of the global variables. Basically, in each state a nondeterministic choice is made between the summands that are enabled given these values.

Each interactive summand has a condition (the guard) that specifies for which valuations of the global variables it is enabled. If so, an action can be taken and the next state (a new valuation for the global variables) is determined probabilistically. The action and next state may also depend on the state. The Markovian summands are similar, except that they contain a rate and a unique next state instead of an action and a probabilistic next state. We assume an implicit confluence classification $P = \{C_1, \ldots, C_k\}$ that, for each interactive summand, contains a group consisting of all transitions generated by that summand.

For a precise formalisation of the language and its semantics, we refer to [24].

Confluent summands. We check for *confluent summands*: summands that are guaranteed to *only* yield confluent transitions, i.e., summands i such that the set $\mathcal{T} = \{C_i\}$ is confluent. Whenever during state space generation such a summand is enabled, all other summands can be ignored (continuing until reaching a representative in a BSCC, as explained in the previous section). By Proposition 12, the union of all confluent summands is also confluent.

Since only τ-transitions can be confluent, the only summands that might be confluent are interactive summands having action τ for all valuations of the global variables. Also, the next state of each of the transitions they generate should be unique. Finally, we verify whether all transitions that may result from these summands commute with all other transitions according to Definition 9.

We only need to check commutativity with all transitions possibly generated by the interactive summands, as the Markovian summands are never enabled at the same time as an invisible transition due to the maximal progress assumption. We over-approximate commutativity by checking whether, when both summands are enabled, they do not disable each other and do not influence each other's actions, probabilities and next states. After all, that implies that each transition can be mimicked by a transition from the same summand (and hence also that it is indeed mimicked by the same group of P). This can by formally expressed as a logical formula (see [25] for the details). Such a formula can be checked by an SMT solver, or approximated using heuristics. We implemented basic heuristics, checking mainly whether two summands are never enabled at the same time or whether the variables updated by one are not used by the other and vice versa. Additionally, some laws from the natural numbers have been implemented, taking for instance into account that $x := x + 1$ cannot disable $x > 2$. In the future, we hope to extend this to more advanced theorem proving.

6 Case studies

We implemented confluence reduction in our tool SCOOP [23]. It takes MAPA specifications as input, is able to perform several reduction techniques and can generate state spaces in multiple formats, among which the one for the IMCA tool for model checking MAs [19]. We already showed in [24] the benefits of dead variable reduction. Here, we apply only confluence reduction, to focus on the power of our novel technique. We present the size of the state spaces with and without confluence reduction, as well as the time to generate them with SCOOP and to subsequently analyse them with IMCA. That way, the impact of confluence reduction on both MA generation and analysis becomes clear[3].

We conjecture that the (quantitative) behavioural equivalence induced by branching bisimulation leaves invariant the time-bounded reachability probabilities, expected times to reachability and long-run averages computed by IMCA. This indeed turned out to be the case for all our models. A logic precisely characterising Markovian branching bisimulation would be interesting future work.

[3] The tool (for download and web-based usage), all MAPA models and a test script can be found on `http://wwwhome.cs.utwente.nl/~timmer/scoop/papers/formats`.

Specification	Original state space States	Trans.	SCOOP	IMCA	Reduced state space States	Trans.	SCOOP	IMCA	Reduction States	Time
`leader-3-7`	25,505	34,257	4.7	103.8	4,652	5,235	5.8	5.2	82%	90%
`leader-3-9`	52,465	71,034	9.7	214.3	9,058	10,149	8.8	9.9	83%	92%
`leader-3-11`	93,801	127,683	18.1	431.7	15,624	17,463	16.4	16.7	83%	93%
`leader-4-2`	8,467	11,600	2.1	74.9	2,071	2,650	2.2	5.2	76%	90%
`leader-4-3`	35,468	50,612	9.0	369.3	7,014	8,874	7.6	22.4	80%	92%
`leader-4-4`	101,261	148,024	25.9	1,325.3	17,885	22,724	20.9	62.2	82%	94%
`polling-2-2-4`	4,811	8,578	0.7	3.7	3,047	6,814	0.7	2.3	37%	32%
`polling-2-2-6`	27,651	51,098	12.6	90.9	16,557	40,004	5.4	49.1	40%	47%
`polling-2-4-2`	6,667	11,290	0.9	39.9	4,745	9,368	0.9	26.2	29%	32%
`polling-2-5-2`	27,659	47,130	4.1	1,573.8	19,721	39,192	4.1	1,053.5	29%	33%
`polling-3-2-2`	2,600	4,909	0.4	7.1	1,914	4,223	0.5	4.8	26%	29%
`polling-4-6-1`	15,439	29,506	3.2	330.0	4,802	18,869	3.2	109.3	69%	66%
`polling-5-4-1`	21,880	43,760	5.4	815.0	6,250	28,130	5.3	317.5	71%	61%
`processor-2`	2,508	4,608	0.7	2.8	1,393	2,922	0.7	1.1	44%	49%
`processor-3`	10,852	20,872	3.1	66.3	6,011	13,240	3.2	19.8	45%	67%
`processor-4`	31,832	62,356	10.7	922.5	17,565	39,558	10.0	316.5	45%	65%

Table 1. State space generation and analysis using confluence reduction (on a 2.4 GHz 4 GB Intel Core 2 Duo MacBook). Runtimes in SCOOP and IMCA are in seconds.

Leader election protocol. The first case study is a leader election protocol (Algorithm $\mathcal{B}$ from [15]), used in [25] as well to demonstrate confluence reduction for probabilistic automata. It uses asynchronous channels and allows for multiple nodes, throwing dice to break the symmetry. We added a rate 1 to a node throwing a die to get an MA model based on the original case study, making the example more relevant and interesting in the current situation. We computed the probability (with error bound 0.01) of electing a leader within 5 time units. The results are presented in Table 1, where we denote by `leader-i-j` the variant with `i` nodes and `j`-sided dice. The computed probability varies from 0.36 for `leader-4-2` to 0.95 for `leader-3-11`. Confluence saved over 90% of the total time to generate and analyse the models. The substantial reductions are due to extensive interleaving with little communication.

Queueing system. The second case study is the queueing system from [24]. It consists of multiple stations with incoming jobs, and one server that polls the stations for work. With some probability, communication fails. There can be different sizes of buffers in the stations, and multiple types of jobs with different service rates. In Table 1, we let `polling-i-j-k` denote the variant with `i` stations, all having buffers of size `j` and `k` types of jobs. Note that, although significant reductions are obtained, the reduction in states precisely corresponds to the reduction in transitions; this implies that only trivially confluent transitions could be reduced (i.e., invisible transitions without any other transitions from the same source state). We computed the expected time to the situation that all buffers are full. This turns out to be at least 1.1—for `polling-3-2-2`—and at most 124—for `polling-2-5-2`. Reductions were less substantial, due to the presence of many probabilistic and Markovian transitions.

Processor architecture. The third case study is a GSPN model of a 2×2 concurrent processor architecture, parameterised in the level k of multitasking, taken from Figure 11.7 in [1]. We constructed a corresponding MAPA model,

modelling each place as a global variable and each transition as a summand. As in [1], we computed the throughput of one of the processors, given by the long-run average of having a token in a certain place of the GSPN. Whereas [1] resolved all nondeterminism and found for instance a throughput of 0.903 for $k = 2$, we are able to retain the nondeterminism and obtain the more informative interval $[0.811, 0.995]$. (When resolving nondeterminism as before, we are able to reproduce the result 0.903.)

Our results clearly show the significant effect of confluence reduction on the state space sizes and the duration of the heavy numerical computations by IMCA. The generation times by SCOOP are not reduced as much, due to the additional overhead of computing representative states. To keep memory usage in the order of the reduced state space, the representative map is deliberately not stored and therefore potentially partly recomputed for some states.

7 Conclusions

We introduced confluence reduction for MAs: the first reduction technique for this model that abstracts from invisible transitions. We showed that it preserves divergence-sensitive branching bisimulation, and hence yields quantitatively behavioural equivalent models. In addition to working on MAs, our novel notion of confluence reduction has two additional advantages over previous notions. First, it preserves divergences, and hence does not alter minimal reachability probabilities. Second, it is closed under unions, enabling us to separately detect confluence of different sets of transitions and combine the results. We also showed that the representation map approach can still be used safely to reduce systems on-the-fly, and discussed how to detect confluence syntactically on the process-algebraic language MAPA. Case studies with our tool SCOOP on several instances of three different models show state space reductions up to 83%. We linked SCOOP to the IMCA model checker to illustrate the significant impact of these reductions on the expected time, time-bounded reachability and long-run averages computations. Due to confluence reduction, for some models the entire process from MAPA specification to results is now more than ten times as fast.

As future work we envision to search for even more powerful ways of using commutativity for state space reduction, for instance by allowing confluent transitions to be probabilistic. Preferably, this would enable even more aggressive reductions that, instead of preserving the conservative notion of bisimulation we used, preserve the more powerful weak bisimulation from [14].

Acknowledgements. We thank Stefan Blom and Joost-Pieter Katoen for their useful suggestions, and Dennis Guck for his help with the case studies.

References

[1] M. Ajmone Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis. *Modelling with Generalized Stochastic Petri Nets*. John Wiley & Sons, Inc., 1994.

[2] M. Ajmone Marsan, G. Conte, and G. Balbo. A class of generalized stochastic Petri nets for the performance evaluation of multiprocessor systems. *ACM Transactions on Computer Systems*, 2(2):93–122, 1984.

[3] C. Baier, P. R. D'Argenio, and M. Größer. Partial order reduction for probabilistic branching time. In *QAPL*, volume 153(2) of *ENTCS*, pages 97–116, 2006.
[4] C. Baier, M. Größer, and F. Ciesinski. Partial order reduction for probabilistic systems. In *QEST*, pages 230–239, 2004.
[5] C. Baier and J.-P. Katoen. *Principles of model checking*. MIT Press, 2008.
[6] S. Blom. Partial τ-confluence for efficient state space generation. Technical Report SEN-R0123, CWI, Amsterdam, 2001.
[7] S. C. C. Blom and J. C. van de Pol. State space reduction by proving confluence. In *CAV*, volume 2404 of *LNCS*, pages 596–609, 2002.
[8] H. Boudali, P. Crouzen, B. R. Haverkort, M. Kuntz, and M. I. A. Stoelinga. Architectural dependability evaluation with arcade. In *DSN*, pages 512–521, 2008.
[9] H. Boudali, P. Crouzen, and M. I. A. Stoelinga. A rigorous, compositional, and extensible framework for dynamic fault tree analysis. *IEEE Transactions on Dependable and Secure Compututing*, 7(2):128–143, 2010.
[10] M. Bozzano, A. Cimatti, J.-P. Katoen, V. Y. Nguyen, T. Noll, and M. Roveri. Safety, dependability and performance analysis of extended AADL models. *The Computer Journal*, 54(5):754–775, 2011.
[11] P. R. D'Argenio and P. Niebert. Partial order reduction on concurrent probabilistic programs. In *QEST*, pages 240–249, 2004.
[12] Y. Deng and M. Hennessy. On the semantics of Markov automata. In *ICALP*, volume 6756 of *LNCS*, pages 307–318, 2011.
[13] C. Eisentraut, H. Hermanns, and L. Zhang. Concurrency and composition in a stochastic world. In *CONCUR*, volume 6269 of *LNCS*, pages 21–39, 2010.
[14] C. Eisentraut, H. Hermanns, and L. Zhang. On probabilistic automata in continuous time. In *LICS*, pages 342–351, 2010.
[15] W. Fokkink and J. Pang. Simplifying Itai-Rodeh leader election for anonymous rings. In *Proc. of the 4th International Workshop on Automated Verification of Critical Systems (AVoCS)*, volume 128(6) of *ENTCS*, pages 53–68, 2005.
[16] P. Godefroid. *Partial-order Methods for the Verification of Concurrent Systems: an Approach to the State-explosion Problem*, volume 1032 of *LNCS*. 1996.
[17] H. Hansen and M. Timmer. A comparison of confluence and ample sets in probabilistic and non-probabilistic branching time. *To be published in TCS*, 2013.
[18] A. Hartmanns and M. Timmer. On-the-fly confluence detection for statistical model checking. In *NFM*, LNCS, 2013 (to appear).
[19] IMCA model checker. `http://www-i2.informatik.rwth-aachen.de/imca/`.
[20] J.-P. Katoen. GSPNs revisited: Simple semantics and new analysis algorithms. In *ACSD*, pages 6–11, 2012.
[21] D. Peled. All from one, one for all: on model checking using representatives. In *CAV*, volume 697 of *LNCS*, pages 409–423, 1993.
[22] M. I. A. Stoelinga. *Alea jacta est: Verification of Probabilistic, Real-time and Parametric Systems*. PhD thesis, University of Nijmegen, 2002.
[23] M. Timmer. SCOOP: A tool for symbolic optimisations of probabilistic processes. In *QEST*, pages 149–150, 2011.
[24] M. Timmer, J.-P. Katoen, J. C. van de Pol, and M. I. A. Stoelinga. Efficient modelling and generation of Markov automata. In *CONCUR*, volume 7454 of *LNCS*, pages 364–379, 2012.
[25] M. Timmer, M. I. A. Stoelinga, and J. C. van de Pol. Confluence reduction for probabilistic systems. In *TACAS*, volume 6605 of *LNCS*, pages 311–325, 2011.
[26] A. Valmari. Stubborn sets for reduced state space generation. In *APN*, volume 483 of *LNCS*, pages 491–515, 1989.

A Divergence-sensitive branching bisimulation

MAs may contain states in which nondeterministic choices arise. Schedulers can be used to specify how these choices are resolved. Our schedulers can select from interactive transitions as well as Markovian transitions, as both might be enabled at the same time. This is due to the fact that we consider *open* MAs, in which the timing of visible actions is still to be determined by their context.

Definition 18 (Schedulers). *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$ be an MA, and $\to$ its set of extended transitions. Then, a* scheduler *for $\mathcal{M}$ is a function*

$$\mathcal{S} \colon \mathit{finpaths}_{\mathcal{M}} \to \mathrm{Distr}(\{\bot\} \cup \to),$$

such that, for every $\pi \in \mathit{finpaths}_{\mathcal{M}}$, the transitions $s \xrightarrow{\alpha} \mu$ that are scheduled by $\mathcal{S}$ after π are indeed possible, i.e., $\mathcal{S}(\pi)(s, \alpha, \mu) > 0$ implies $s = \mathit{last}(\pi)$. The decision of not choosing any transition is represented by $\bot$.

We define the sets of finite and maximal paths enabled by a given scheduler, and define how each scheduler induces a probability distribution over paths (as in [25]).

Definition 19 (Finite and maximal paths). *Let $\mathcal{M}$ be an MA and $\mathcal{S}$ a scheduler for $\mathcal{M}$. Then, the set of* finite paths of $\mathcal{M}$ under $\mathcal{S}$ *is given by*

$$\mathit{finpaths}^{\mathcal{S}}_{\mathcal{M}} = \{\pi \in \mathit{finpaths}_{\mathcal{M}} \mid \forall 0 \leq i < |\pi| \,.\, \mathcal{S}(\mathit{prefix}(\pi, i))(\mathit{step}(\pi, i+1)) > 0\}.$$

We define $\mathit{finpaths}^{\mathcal{S}}_{\mathcal{M}}(s) \subseteq \mathit{finpaths}^{\mathcal{S}}_{\mathcal{M}}$ as the set of all such paths starting in s. The set of maximal paths of $\mathcal{M}$ under $\mathcal{S}$ *is given by*

$$\mathit{maxpaths}^{\mathcal{S}}_{\mathcal{M}} = \{\pi \in \mathit{finpaths}^{\mathcal{S}}_{\mathcal{M}} \mid \mathcal{S}(\pi)(\bot) > 0\}.$$

Similarly, $\mathit{maxpaths}^{\mathcal{S}}_{\mathcal{M}}(s)$ is the set of maximal paths of $\mathcal{M}$ under $\mathcal{S}$ starting in s.

Definition 20 (Path probabilities). *Let $\mathcal{M}$ be an MA with a state s, and $\mathcal{S}$ a scheduler for $\mathcal{M}$. Then, we define the function $P^{\mathcal{S}}_{\mathcal{M},s} \colon \mathit{finpaths}_{\mathcal{M}}(s) \to [0,1]$ by*

$$P^{\mathcal{S}}_{\mathcal{M},s}(s) = 1; \qquad P^{\mathcal{S}}_{\mathcal{M},s}(\pi \xrightarrow{a,\mu} t) = P^{\mathcal{S}}_{\mathcal{M},s}(\pi) \cdot \mathcal{S}(\pi)(\mathit{last}(\pi), a, \mu) \cdot \mu(t).$$

A scheduler also induces a probability to terminate in some state s' when starting in state s. Following [25], we define this by $F^{\mathcal{S}}_{\mathcal{M}}(s)(s')$. Note that the distribution $F^{\mathcal{S}}_{\mathcal{M}}(s)$ may be substochastic, as $\mathcal{S}$ does not necessarily terminate.

Definition 21 (Final state probabilities). *Let $\mathcal{M}$ be an MA and $\mathcal{S}$ a scheduler for $\mathcal{M}$. Then, we define the function $F^{\mathcal{S}}_{\mathcal{M}} \colon S \to \mathrm{SDistr}(S)$ by*

$$F^{\mathcal{S}}_{\mathcal{M}}(s) = \Big\{ s' \mapsto \sum_{\substack{\pi \in \mathit{maxpaths}^{\mathcal{S}}_{\mathcal{M}}(s) \\ \mathit{last}(\pi) = s'}} P^{\mathcal{S}}_{\mathcal{M},s}(\pi) \cdot \mathcal{S}(\pi)(\bot) \mid s' \in S \Big\} \qquad \forall s \in S.$$

Example 22. For the system in Example 4 we can define a scheduler $\mathcal{S}$ by

$$\mathcal{S}(\epsilon)(s_2 \xrightarrow{\chi(9)} \mu_1) = 1 \quad \mathcal{S}(s_2 \xrightarrow{\chi(9),\mu_1} s_3)(\bot) = 1 \quad \mathcal{S}(\pi_1)(s_1 \xrightarrow{a} \mu_2) = 1$$
$$\mathcal{S}(\pi_0)(s_0 \xrightarrow{\chi(2)} \mathbb{1}_{s_1}) = \tfrac{1}{2} \quad \mathcal{S}(\pi_0)(\bot) = \tfrac{1}{2} \quad \mathcal{S}(\pi_4)(s_4 \xrightarrow{\tau} \mathbb{1}_{s_5}) = 1 \quad \mathcal{S}(\pi_5)(\bot) = 1$$

with μ_1 and μ_2 as in Example 7, and each π_i any path ending in s_i. For the path π given in Example 7, we find $P^{\mathcal{S}}_{\mathcal{M},s_2}(\pi) = (1\cdot\frac{7}{9})\cdot(1\cdot\frac{2}{3})\cdot(\frac{1}{2}\cdot 1)\cdot(1\cdot\frac{1}{3})\cdot(1\cdot 1) = \frac{7}{81}$, for each step multiplying the probability of taking the transition by the probability of selecting the given next state. Using the formula for infinite geometric series, we find that $F^{\mathcal{S}}_{\mathcal{M}}(s_2)$ assigns probability $\frac{4}{18}$ to s_3, $\frac{7}{18}$ to s_0 and $\frac{7}{18}$ to s_5. □

We now define branching steps for MAs. Intuitively, a state s can do a branching step $s \overset{a}{\Longrightarrow}_R \mu$ if there exists a scheduler that terminates according to μ, always schedules precisely one a-transition (immediately before terminating), does not schedule any other visible transitions and does not leave the equivalence class $[s]_R$ before doing an a-transition. Additionally, every state can do a branching τ-step to itself. Due to the use of extended transitions as a uniform manner of dealing with both interactive and Markovian transitions, this definition precisely coincides with the definition of branching steps for PAs [25].

Definition 23 (Branching steps). *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow\rangle$ be an MA, $s \in S$, and R an equivalence relation over S. Then, $s \overset{a}{\Longrightarrow}_R \mu$ if either (1) $a = \tau$ and $\mu = \mathbb{1}_s$, or (2) there exists a scheduler $\mathcal{S}$ such that $F^{\mathcal{S}}_{\mathcal{M}}(s) = \mu$ and for every maximal path $s \xrightarrow{a_1,\mu_1} s_1 \xrightarrow{a_2,\mu_2} s_2 \xrightarrow{a_3,\mu_3} \ldots \xrightarrow{a_n,\mu_n} s_n \in maxpaths^{\mathcal{S}}_{\mathcal{M}}(s)$ it holds that $a_n = a$, as well as $a_i = \tau$ and $(s, s_i) \in R$ for all $1 \leq i < n$.*

Based on these branching steps, we define branching bisimulation for MAs as a natural extension of the notion of naive weak bisimulation from [14]. It can easily be seen that naive weak bisimulation is immediately implied by our notion of branching bisimulation.

Definition 24 (Branching bisimulation). *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow\rangle$ be an MA, then an equivalence relation $R \subseteq S \times S$ is a* branching bisimulation for $\mathcal{M}$ *if for all $(s,t) \in R$ and every extended transition $s \xrightarrow{a} \mu$, there is a transition $t \overset{a}{\Longrightarrow}_R \mu'$ such that $\mu \equiv_R \mu'$. We say that $p, q \in S$ are* branching bisimilar*, denoted by $p \leftrightarroweq_{\mathrm{b}} q$, if there is a branching bisimulation R for $\mathcal{M}$ with $(p,q) \in R$.*

Two MAs are branching bisimilar if their initial states are, in the disjoint union of the two systems (see Remark 5.3.4 of [22] for the details). For a more elaborate discussion on branching bisimulation, we refer to [25].

Minimal probabilities (e.g., of eventually seeing an a-action) are not invariant under branching bisimulation. Consider for instance a system consisting of two states, connected by an a-transition and both having a τ-selfloop. Due to these divergences, the a-transition never has to happen. Still, this system is branching bisimilar to the same system without the τ-selfloops. However, in that case the minimal probability of traversing the a-transition is 1.

Hence, divergence-sensitive notions of bisimulation have been introduced that take into account that diverging states are always mapped to diverging states [5].

Definition 25 (Divergence-sensitive relations). *An equivalence relation R is* divergence sensitive *if for all $(s, s') \in R$ it holds that*

$$\exists \mathcal{S} \,.\, \forall \pi \in \mathit{finpaths}^{\mathcal{S}}_{\mathcal{M}}(s) \,.\, \mathit{trace}(\pi) = \epsilon \wedge \mathcal{S}(\pi)(\bot) = 0$$
$$\iff$$
$$\exists \mathcal{S}' \,.\, \forall \pi \in \mathit{finpaths}^{\mathcal{S}'}_{\mathcal{M}}(s') \,.\, \mathit{trace}(\pi) = \epsilon \wedge \mathcal{S}'(\pi)(\bot) = 0$$

Two MAs $\mathcal{M}_1, \mathcal{M}_2$ are divergence-sensitive branching bisimilar*, in which case we write $\mathcal{M}_1 \leftrightarrows^{\mathrm{div}}_{\mathrm{b}} \mathcal{M}_2$, if they are branching bisimilar and the equivalence relation to show this is divergence sensitive.*

Hence, if $(s, s') \in R$ and R is divergence sensitive, then s can diverge (perform an endless series of τ-transitions with probability 1) if and only if s' can.

B Proofs

In all proofs, whenever a confluent set $\mathcal{T}$ is given, we abuse notation by writing *confluent transition* to denote a transition in this set $\mathcal{T}$. Note that, in general, there might also be confluent transitions that are not in $\mathcal{T}$.

Proposition 11. *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$ be an MA, $P \subseteq \mathscr{P}(\hookrightarrow)$ a confluence classification for $\mathcal{M}$ and $\mathcal{T}$ a Markovian confluent set for P. Then,*

$$s \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} t \qquad \textit{if and only if} \qquad s \twoheadleftarrow\!\!\!\!\twoheadrightarrow_{\mathcal{T}} t$$

Proof. We separately prove both directions of the equivalence.

($\Longrightarrow$) Let $s \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} t$. Then, by definition there is a state u such that $s \twoheadrightarrow_{\mathcal{T}} u$ and $t \twoheadrightarrow_{\mathcal{T}} u$. This immediately implies that $s \twoheadleftarrow\!\!\!\!\twoheadrightarrow_{\mathcal{T}} t$.

($\Longleftarrow$) Let $s \twoheadleftarrow\!\!\!\!\twoheadrightarrow_{\mathcal{T}} t$. This means that there is a path from s to t such as

$$s_0 \leftarrow s_1 \rightarrow s_2 \rightarrow s_3 \leftarrow s_4 \leftarrow s_5 \rightarrow s_6,$$

where $s_0 = s$, $s_6 = t$ and each of the transitions is in $\mathcal{T}$. Note that $s_i \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} s_{i+1}$ for all s_i, s_{i+1} on this path. After all, if $s_i \rightarrow s_{i+i}$ then they can join at s_{i+1}, otherwise they can join at s_i. Hence, to show that $s \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} t$, it suffices to show that $\twoheadrightarrow \twoheadleftarrow_{\mathcal{T}}$ is transitive.

Let $s' \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} s$ and $s \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} s''$. We show that $s' \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} s''$. Let t' be a state such that $s \twoheadrightarrow_{\mathcal{T}} t'$ and $s' \twoheadrightarrow_{\mathcal{T}} t'$, and likewise, let t'' be a similar state for s and s''. If we can show that there is some state t such that $t' \twoheadrightarrow_{\mathcal{T}} t$ and $t'' \twoheadrightarrow_{\mathcal{T}} t$, we have the result. Let a minimal confluent path from s to t' be given by $s_0 \rightarrow_{\mathcal{T}} s_1 \rightarrow_{\mathcal{T}} \cdots \rightarrow_{\mathcal{T}} s_n$, with $s_0 = s$ and $s_n = t'$. By induction on the length of this path, we show that for each state s_i on it, there is some state t such that $s_i \twoheadrightarrow_{\mathcal{T}} t$ and $t'' \twoheadrightarrow_{\mathcal{T}} t$. Since t' is also on the path, this completes the argument.

Base case. There clearly is a state t such that $s_0 \twoheadrightarrow_{\mathcal{T}} t$ and $t'' \twoheadrightarrow_{\mathcal{T}} t$, namely t'' itself. After all, $s_0 = s$ and $s \twoheadrightarrow_{\mathcal{T}} t''$, and $\twoheadrightarrow_{\mathcal{T}}$ is reflexive.

Inductive case. Let there be a state t_k such that $s_k \twoheadrightarrow_{\mathcal{T}} t_k$ and $t'' \twoheadrightarrow_{\mathcal{T}} t_k$. We show that there exists a state t_{k+1} such that $s_{k+1} \twoheadrightarrow_{\mathcal{T}} t_{k+1}$ and $t'' \twoheadrightarrow_{\mathcal{T}} t_{k+1}$. Let $s_k \xrightarrow{\tau} u$ be the first transition on the $\mathcal{T}$-path from s_k to t_k. Let $s_k \xrightarrow{\tau} s_{k+1}$ be the $\mathcal{T}$-transition between s_k and s_{k+1}. Since it is in $\mathcal{T}$, there must be at least one group $C \in P \cap \mathcal{T}$ such that $s_k \xrightarrow{\tau}_C s_{k+1}$.

By definition of confluence, since $(s_k \xrightarrow{\tau} u) \in \mathcal{T}$ and $s_k \xrightarrow{\tau}_C s_{k+1}$ for some $C \in P$, either (1) $s_{k+1} = u$ (the transitions coincide), or (2) there is a transition $u \xrightarrow{\tau}_C u'$ such that $\mathbb{1}_{s_{k+1}} \equiv_R \mathbb{1}_{u'}$, with R the equivalence relation given in Definition 9.

In case (1), we directly find $s_{k+1} \twoheadrightarrow_{\mathcal{T}} t_k$. Hence, we can just take $t_{k+1} = t_k$. In case (2), either $s_{k+1} = u'$ or $s_{k+1} \xrightarrow{\tau}_{\mathcal{T}} u'$. In both cases, if $u = t_k$, we can take $t_{k+1} = u'$ and indeed $s_{k+1} \twoheadrightarrow_{\mathcal{T}} t_{k+1}$ and $t'' \twoheadrightarrow_{\mathcal{T}} t_{k+1}$. Otherwise, we can use the same reasoning to show that there is a state t_{k+1} such that $u' \twoheadrightarrow_{\mathcal{T}} t_{k+1}$ and $t'' \twoheadrightarrow_{\mathcal{T}} t_{k+1}$, based on $u \twoheadrightarrow_{\mathcal{T}} t_k$, $t'' \twoheadrightarrow_{\mathcal{T}} t_k$ and $u \xrightarrow{\tau}_{\mathcal{T}} u'$. Since the path from u to t_k is one transition shorter than the path from s_k to t_k, this argument terminates. □

Proposition 12. *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$ be an MA, $P \subseteq \mathscr{P}(\hookrightarrow)$ a confluence classification for $\mathcal{M}$ and $\mathcal{T}_1, \mathcal{T}_2$ two Markovian confluent sets for P. Then, $\mathcal{T}_1 \cup \mathcal{T}_2$ is also a Markovian confluent set for P.*

Proof. Let $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2$. Clearly, $\mathcal{T}$ still only contain invisible transitions with Dirac distributions, since $\mathcal{T}_1$ and $\mathcal{T}_2$ do. Consider a transition $(s \xrightarrow{\tau}_{\mathcal{T}} t)$, and another transition $s \xrightarrow{a} \mu$. We need to show that

$$\begin{cases} \forall C \in P \,.\, s \xrightarrow{a}_C \mu \implies \exists \nu \in \mathrm{Distr}(S) \,.\, t \xrightarrow{a}_C \nu \ \wedge \mu \equiv_R \nu & \text{, if } (s \xrightarrow{a} \mu) \in P \\ \qquad\qquad\qquad\quad \exists \nu \in \mathrm{Distr}(S) \,.\, t \xrightarrow{a} \nu \quad \wedge \mu \equiv_R \nu & \text{, if } (s \xrightarrow{a} \mu) \notin P \end{cases}$$

where R is the smallest equivalence relation such that

$$R \supseteq \{(s,t) \in \mathrm{spt}(\mu) \times \mathrm{spt}(\nu) \mid (s \xrightarrow{\tau} t) \in \mathcal{T}\}.$$

Without loss of generality, assume that $s \xrightarrow{\tau}_{\mathcal{T}_1} t$. Hence, by definition of Markovian confluence, we find that

$$\begin{cases} \forall C \in P \,.\, s \xrightarrow{a}_C \mu \implies \exists \nu \in \mathrm{Distr}(S) \,.\, t \xrightarrow{a}_C \nu \ \wedge \mu \equiv_{R_1} \nu & \text{, if } (s \xrightarrow{a} \mu) \in P \\ \qquad\qquad\qquad\quad \exists \nu \in \mathrm{Distr}(S) \,.\, t \xrightarrow{a} \nu \quad \wedge \mu \equiv_{R_1} \nu & \text{, if } (s \xrightarrow{a} \mu) \notin P \end{cases}$$

where R_1 is the smallest equivalence relation such that

$$R_1 \supseteq \{(s,t) \in \mathrm{spt}(\mu) \times \mathrm{spt}(\nu) \mid (s \xrightarrow{\tau} t) \in \mathcal{T}_1\}$$

Note that $R \supseteq R_1$ since $\mathcal{T} \supseteq \mathcal{T}_1$. Therefore, $\mu \equiv_{R_1} \nu$ implies $\mu \equiv_R \nu$ (using Proposition 5.2.1.5 from [22]). The result now immediately follows. □

Lemma 26. *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$ be an MA, $s, s' \in S$ two of its states, $a \in A$, $\mu \in \mathrm{Distr}(S)$, $P \subseteq \mathscr{P}(\hookrightarrow)$ a confluence classification for $\mathcal{M}$ and $\mathcal{T}$ a Markovian confluent set for P. Then,*

$$s \twoheadrightarrow_{\mathcal{T}} s' \wedge s \xrightarrow{a} \mu \Longrightarrow (a = \tau \wedge \mu \equiv_R \mathbb{1}_{s'}) \vee \left(\exists \nu \in \mathrm{Distr}(S) \,.\, s' \xrightarrow{a} \nu \wedge \mu \equiv_R \nu\right)$$

where $R = \{(u, v) \mid u \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} v\}$.

Proof. Let $s, s' \in S$ be such that $s \twoheadrightarrow_{\mathcal{T}} s'$, and assume a transition $s \xrightarrow{a} \mu$. Let $R = \{(u, v) \mid u \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} v\}$. We show that either $a = \tau \wedge \mu \equiv_R \mathbb{1}_{s'}$ or that there exists a transition $s' \xrightarrow{a} \nu$ such that $\mu \equiv_R \nu$, by induction on the length of the confluent path from s to s'. Let $s_0 \xrightarrow{\tau}_{\mathcal{T}} s_1 \xrightarrow{\tau}_{\mathcal{T}} \ldots \xrightarrow{\tau}_{\mathcal{T}} s_{n-1} \xrightarrow{\tau}_{\mathcal{T}} s_n$, with $s_0 = s$ and $s_n = s'$, denote this path. Then, we show that

$$(a = \tau \wedge \mu \equiv_R \mathbb{1}_{s'}) \vee \left(\exists \nu \in \mathrm{Distr}(S) \,.\, s_i \xrightarrow{a} \nu \wedge \mu \equiv_R \nu\right)$$

holds for every state s_i on this path. For the base case s this is immediate, since $s \xrightarrow{a} \mu$ and the relation $\equiv_R$ is reflexive.

As induction hypothesis, assume that the formula holds for some state s_i $(0 \leq i < n)$. We show that it still holds for state s_{i+1}. If the above formula was true for s_i due to the clause $a = \tau \wedge \mu \equiv_R \mathbb{1}_{s'}$, then this still holds for s_{i+1}. So, assume that $s_i \xrightarrow{a} \nu$ such that $\mu \equiv_R \nu$.

Since $s_i \xrightarrow{\tau}_{\mathcal{T}} s_{i+1}$ and $s_i \xrightarrow{a} \nu$, by definition of confluence either (1) $a = \tau$ and $\nu = \mathbb{1}_{s_{i+1}}$, or (2) there is a transition $s_{i+1} \xrightarrow{a} \nu'$ such that $\nu \equiv_{R'} \nu'$, where R' is the smallest equivalence relation such that

$$R' \supseteq \{(s, t) \in \mathrm{spt}(\nu) \times \mathrm{spt}(\nu') \mid (s \xrightarrow{\tau} t) \in \mathcal{T}\}.$$

(1) In the first case, $\nu = \mathbb{1}_{s_{i+1}}$ implies that $\nu \equiv_R \mathbb{1}_{s'}$ as there is a $\mathcal{T}$-path from s_{i+1} to s' and hence $(s_{i+1}, s') \in R$. Since we assumed that $\mu \equiv_R \nu$, and the relation $\equiv_R$ is transitive, this yields $\mu \equiv_R \mathbb{1}_{s'}$. Together with $a = \tau$, this completes the proof.

(2) In the second case, note that $R \supseteq R'$. After all, $R = \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} = \twoheadleftarrow\!\!\!\!\rightarrow_{\mathcal{T}}$ (by Proposition 11), and obviously $(s, t) \in R'$ implies that $s \twoheadleftarrow\!\!\!\!\rightarrow_{\mathcal{T}} t$. Hence, $\nu \equiv_{R'} \nu'$ implies $\nu \equiv_R \nu'$ (using Proposition 5.2.1.5 from [22]). Since we assumed that $\mu \equiv_R \nu$, by transitivity of $\equiv_R$ we obtain $\mu \equiv_R \nu'$. Hence, there is a transition $s_{i+1} \xrightarrow{a} \nu'$ such that $\mu \equiv_R \nu'$, which completes the proof. □

Theorem 14. *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$ be an MA, $s, s' \in S$ two of its states, $P \subseteq \mathscr{P}(\hookrightarrow)$ a confluence classification for $\mathcal{M}$ and $\mathcal{T}$ a Markovian confluent set for P. Then,*

$$s \twoheadleftarrow\!\!\!\!\rightarrow_{\mathcal{T}} s' \text{ implies } s \leftrightarrows_{\mathrm{b}}^{\mathrm{div}} s'.$$

Proof. We show that $s \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} s'$ implies $s \leftrightarrows_{\mathrm{b}}^{\mathrm{div}} s'$. By Proposition 11, this is equivalent to the theorem. So, assume that $s \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} s'$. To show that $s \leftrightarrows_{\mathrm{b}}^{\mathrm{div}} s'$, consider the relation

$$R = \{(u, v) \mid u \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} v\}$$

Clearly $(s, s') \in R$, and from Proposition 11 and the obvious fact that $\twoheadleftarrow\!\!\!\!\rightarrow_{\mathcal{T}}$ is an equivalence relation, it follows that R is an equivalence relation as well. It remains to show that R is a divergence-sensitive branching bisimulation. Hence, let $(p, q) \in R$, i.e., $p \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} q$. We need to show that for every extended transition $p \xrightarrow{a} \mu$ there is a transition $q \stackrel{a}{\Longrightarrow}_R \mu'$ such that $\mu \equiv_R \mu'$.

So, assume such a transition $p \xrightarrow{a} \mu$. Let r be a state such that $p \twoheadrightarrow_{\mathcal{T}} r$ and $q \twoheadrightarrow_{\mathcal{T}} r$. By Lemma 26, either (1) $a = \tau \wedge \mu \equiv_R \mathbb{1}_r$ or (2) there is a distribution $\nu \in \mathrm{Distr}(S)$ such that $r \xrightarrow{a} \nu \wedge \mu \equiv_R \nu$.

(1) In the first case, note that $q \twoheadrightarrow_{\mathcal{T}} r$ immediately implies that $q \stackrel{\tau}{\Longrightarrow}_R \mathbb{1}_r$. After all, we can schedule the (invisible) confluent transitions from q to r and then terminate. Indeed, all intermediate states are clearly related by R. Together with the assumption that $\mu \equiv_R \mathbb{1}_r$, this completes the argument.
(2) In the second case, note that $q \twoheadrightarrow_{\mathcal{T}} r$ and $r \stackrel{a}{\to} \nu$ together immediately imply that $q \stackrel{a}{\Longrightarrow}_R \nu$. After all, we can schedule the (invisible) confluent transitions from q to r, perform the transition $r \stackrel{a}{\to} \nu$ and then terminate. Indeed, all intermediate states before the a-transition are clearly related by R. Together with the assumption that $\mu \equiv_R \nu$, this completes the argument.

It remains to show that R is divergence sensitive. So, let $(s, s') \in R$ (and hence $s \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} s'$) and assume that there is a scheduler $\mathcal{S}$ such that

$$\forall \pi \in \mathit{finpaths}^{\mathcal{S}}_{\mathcal{M}}(s) \,.\, \mathit{trace}(\pi) = \epsilon \wedge \mathcal{S}(\pi)(\bot) = 0$$

It is well known that we can assume that such diverging schedulers are memoryless and deterministic.

We show that there also is a diverging scheduler from s'. First, note that since $s \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} s'$, there is a state t such that $s \twoheadrightarrow_{\mathcal{T}} t$ and $s' \twoheadrightarrow_{\mathcal{T}} t$. We show that there is a diverging scheduler from t; then, the result follows as from s' we can schedule to first follow the confluent (and hence invisible) transitions to t and then continue with the diverging scheduler from t.

Let $s_0 \stackrel{\tau}{\to}_{\mathcal{T}} s_1 \stackrel{\tau}{\to}_{\mathcal{T}} s_2 \stackrel{\tau}{\to}_{\mathcal{T}} \ldots \stackrel{\tau}{\to}_{\mathcal{T}} s_n$ be the confluent path from s to t; hence, $s_0 = s$ and $s_n = t$. It might be the case that some states on this path also occur on the tree associated with $\mathcal{S}$; hence, for those states a diverging scheduler already exists. Let s_i be the last state on the path from s_0 to s_n that occurs on the tree of $\mathcal{S}$. We show that s_n also has a diverging scheduler by induction on the length of the path from s_i to s_n; note that the base case is immediate.

Assume that s_j (with $i \leq j < n$) has a diverging scheduler $\mathcal{S}'$. We show that s_{j+1} has one too. If s_{j+1} occurs on the tree associated with $\mathcal{S}'$ this is immediate, so from now on assume that it does not. From s_j there now is a confluent transition $s_j \stackrel{\tau}{\to}_{\mathcal{T}} s_{j+1}$ and an invisible (not necessarily confluent) transition $s_j \stackrel{\tau}{\to} \mu$ (chosen by $\mathcal{S}'$ as first step of the diverging path form s_j). By definition of confluence, either these transitions coincide or there is a transition $s_{j+1} \stackrel{\tau}{\to} \nu$ such that $\mu \equiv_{R'} \nu$, with R' the smallest equivalence relation such that $R \supseteq \{(s,t) \in \mathrm{spt}(\mu) \times \mathrm{spt}(\nu) \mid (s \stackrel{\tau}{\to} t) \in \mathcal{T}\}$. The first option is impossible, since we assumed that s_{j+1} is not on the tree associated with $\mathcal{S}'$. Therefore, there is a transition $s_{j+1} \stackrel{\tau}{\to} \nu$ such that $\mu \equiv_{R'} \nu$. We schedule this transition from s_{j+1} in order to diverge. Hence, we still need to show that it is possible to diverge from all states $q \in \mathrm{spt}(\nu)$.

By definition of R', $\mu \equiv_{R'} \nu$ implies that each state $q \in \mathrm{spt}(\nu)$ is either (1) in $\mathrm{spt}(\mu)$ as well or (2) has an incoming confluent transition $p \stackrel{\tau}{\to}_{\mathcal{T}} q$ with $p \in \mathrm{spt}(\mu)$. In the first case, we can diverge from q using $\mathcal{S}'$. In the second case, we have reached the situation of a state q with an incoming confluent transition from a state p that has a diverging scheduler. Now, the above reasoning can be applied again, taking $s_j = p$ and $s_{j+1} = q$. Either at some point overlap with the

scheduler of p occurs, or this argument is repeated indefinitely; in both cases, divergence is obtained. □

Theorem 17. *Let $\mathcal{M} = \langle S, s^0, A, \hookrightarrow, \rightsquigarrow \rangle$ be an MA, $\mathcal{T}$ a Markovian confluent set for $\mathcal{M}$, and $\varphi\colon S \to S$ a representation map for $\mathcal{M}$ under $\mathcal{T}$. Then,*

$$\mathcal{M}/\varphi \leftrightarrows_{\mathrm{b}}^{\mathrm{div}} \mathcal{M}.$$

Proof. We denote the extended transition relation of $\mathcal{M}$ by $\rightarrow$, and the one of $\mathcal{M}/\varphi$ by $\rightarrow_\varphi$. We take the disjoint union $\mathcal{M}'$ of $\mathcal{M}$ and $\mathcal{M}/\varphi$, to provide a bisimulation relation over this state space that contains their initial states. We denote the transition relation of $\mathcal{M}'$ by $\rightarrow'$. Note that, based on whether $s \in \mathcal{M}$ or $s \in \mathcal{M}/\varphi$, a transition $s \xrightarrow{a}' \mu$ corresponds to either $s \xrightarrow{a} \mu$ or $s \xrightarrow{a}_\varphi \mu$.

To distinguish between for instance a state $\varphi(s) \in \mathcal{M}$ and the corresponding state $\varphi(s) \in \mathcal{M}/\varphi$, we denote all states $s, \varphi(s)$ from $\mathcal{M}$ just by $s, \varphi(s)$, and all states $s, \varphi(s)$ from $\mathcal{M}/\varphi$ by $\hat{s}, \hat{\varphi}(s)$.

Let R be the smallest equivalence relation containing the set

$$\{(s, \hat{\varphi}(s)) \mid s \in S\},$$

i.e., R relates all states from $\mathcal{M}$ that have the same representative to each other and to this mutual representative from $\mathcal{M}/\varphi$. Clearly, $(s^0, \hat{\varphi}(s^0)) \in R$.

Note that given this equivalence relation R, for every probability distribution μ we have $\mu \equiv_R \varphi(\mu)$ (no matter whether $\varphi(\mu)$ is in $\mathcal{M}$ or in $\mathcal{M}/\varphi$). After all, the lifting over φ just changes the states in the support of μ to their representatives; as R relates precisely such states, clearly $\mu \equiv_R \varphi(\mu)$. This observation is used several times in the proof below.

Now, let $(s, s') \in R$ and assume that there is an extended transition $s \xrightarrow{a}' \mu$. We show that also $s' \xRightarrow{a}'_R \mu'$ such that $\mu \equiv_R \mu'$. Note that there are four possible cases to consider with respect to the origin of s and s', indicated by the presence or absence of hats:

- Case 1: $(\hat{s}, \hat{s}')$. Since every equivalence class of R contains precisely one representative from $\mathcal{M}/\varphi$, we find that $\hat{s} = \hat{s}'$. Hence, the result follows directly by the scheduler that takes the transition $s \xrightarrow{a}' \mu$ and then terminates.
- Case 2: (s, s'). If both states are in $\mathcal{M}$, then the quotient is not involved and $\varphi(s) = \varphi(s')$. By definition of the representation map, we find $s \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} s'$. Using Theorem 14, this immediately implies that $s' \xRightarrow{a}_{R'} \mu'$ such that $\mu \equiv_{R'} \mu'$ for $R' = \{(u, v) \mid u \twoheadrightarrow \twoheadleftarrow_{\mathcal{T}} v\}$. Since all states connected by $\mathcal{T}$-transitions are required to have the same representative, we have $R \supseteq R'$. Hence, also $s' \xRightarrow{a}_R \mu'$, as this is then less restrictive. Moreover, $\mu \equiv_R \mu'$ by Proposition 5.2.1.5 from [22]. Finally, note that $s' \xRightarrow{a}_R \mu'$ implies $s' \xRightarrow{a}'_R \mu'$.
- Case 3: $(\hat{s}, s')$. Since $\hat{s}$ is in $\mathcal{M}/\varphi$ and s' is not, by definition of R we find that $\hat{s} = \hat{\varphi}(s')$. Hence, by assumption $\hat{\varphi}(s') \xrightarrow{a}_\varphi \mu$, and thus by definition of the extended arrow either (1) $a \in A$ and $\hat{\varphi}(s') \xhookrightarrow{a}_\varphi \mu$, or (2) $a = \chi(\lambda)$ for $\lambda = \mathit{rate}(\hat{\varphi}(s'))$, $\lambda > 0$, $\mu = \mathbb{P}_{\hat{\varphi}(s')}$ and there is no μ' such that $\hat{\varphi}(s') \xhookrightarrow{\tau}_\varphi \mu'$. We make a case distinction based on this.

(1) Let $a \in A$ and $\hat{\varphi}(s') \xhookrightarrow{a}_{\varphi} \mu$. By definition of the quotient, this implies that there is a transition $\varphi(s') \xhookrightarrow{a} \mu'$ in $\mathcal{M}$ such that $\mu = \varphi(\mu')$. By definition of the representation map, there is a $\mathcal{T}$-path (which is invisible and deterministic) from s' to $\varphi(s')$ in $\mathcal{M}$. Hence, $s' \stackrel{a}{\Longrightarrow}_R \mu'$ (and therefore also $s' \stackrel{a}{\Longrightarrow}'_R \mu'$) by the scheduler from s' that first goes to $\varphi(s')$ and then executes the $\varphi(s') \xhookrightarrow{a} \mu'$ transition. Note that the transition is indeed branching, as all steps in between have the same representative and thus are related by R.

It remains to show that $\mu \equiv_R \mu'$. We already saw that $\mu = \varphi(\mu')$; hence, the result follows from the observation that $\mu \equiv_R \varphi(\mu)$ for every μ.

(2) Let $a = \chi(\lambda)$ for $\lambda = \mathit{rate}(\hat{\varphi}(s'))$, $\lambda > 0$, $\mu = \mathbb{P}_{\hat{\varphi}(s')}$ and there is no μ' such that $\hat{\varphi}(s') \xhookrightarrow{\tau}_{\varphi} \mu'$. Note that this means that from $\hat{\varphi}(s')$ there is a total outgoing rate of λ, spreading out according to μ. Hence, given an arbitrary state $\hat{u}$ in $\mathcal{M}/\varphi$, we have

$$\mu(\hat{u}) = \frac{\mathit{rate}(\hat{\varphi}(s'), \hat{u})}{\lambda}$$

By definition of the quotient there is at most one Markovian transition between any pair of states in $\mathcal{M}/\varphi$, so for every $\hat{u} \in \mathrm{spt}(\mu)$, there is precisely one Markovian transition $\hat{\varphi}(s') \stackrel{\lambda'}{\rightsquigarrow}_{\varphi} \hat{u}$ with $\lambda' = \mu(\hat{u}) \cdot \lambda$. By definition of the quotient we then also find that λ' is the sum of all outgoing Markovian transitions in $\mathcal{M}$ from $\varphi(s')$ to states t such that $\varphi(t) = u$. Since each state in $\mathcal{M}$ has precisely one representative and $\hat{\varphi}(s')$ has a Markovian transition to all representatives of states reached from $\varphi(s')$ by Markovian transitions, it follows that the total outgoing rate of $\varphi(s')$ is also λ.

Additionally, there is no outgoing τ-transition from $\varphi(s')$, since by definition of the quotient this would have resulted in a τ-transition from $\hat{\varphi}(s')$, which we assumed is not present. Hence, there is an extended transition $\varphi(s') \xrightarrow{\chi(\lambda)} \mu'$ in $\mathcal{M}$. As the total outgoing rates of $\varphi(s')$ and $\hat{\varphi}(s')$ are equal, and the sum of all outgoing Markovian transitions from $\varphi(s')$ to states t such that $\varphi(t) = u$ equals the rate from $\hat{\varphi}(s')$ to $\hat{u}$, we find that $\mu \equiv_R \mu'$ since R equates states to their representative and to other states with the same representative.

By definition of the representation map, there is a $\mathcal{T}$-path (which is invisible and deterministic) from s' to $\varphi(s')$ in $\mathcal{M}$. Hence, $\varphi(s') \xrightarrow{\chi(\lambda)} \mu'$ implies that $s' \stackrel{\chi(\lambda)}{\Longrightarrow}_R \mu'$ and therefore also $s' \stackrel{\chi(\lambda)}{\Longrightarrow}'_R \mu'$. As $\chi(\lambda) = a$ and we already saw that $\mu \equiv_R \mu'$, this completes this part of the proof.

– Case 4: $(s, \hat{s}')$. Since $\hat{s}'$ is in $\mathcal{M}/\varphi$ and s is not, by definition of R we find that $\hat{s}' = \hat{\varphi}(s)$. By definition of the representation map, there is a $\mathcal{T}$-path from s to $\varphi(s)$ in $\mathcal{M}$. Hence, since $s \xrightarrow{a} \mu$, by Lemma 26 we have either (1) $a = \tau \wedge \mu \equiv_{R'} \mathbb{1}_{\varphi(s)}$, or (2) there exists a transition $\varphi(s) \xrightarrow{a} \nu$ such that $\mu \equiv_{R'} \nu$, for $R' = \{(u, v) \mid u \twoheadrightarrow\!\!\!\!\leftarrow_{\mathcal{T}} v\}$. Again, as in case 2 we can safely substitute R' by R.

(1) We need to show that $\hat{\varphi}(s) \stackrel{\tau}{\Longrightarrow}'_R \mu'$ such that $\mathbb{1}_{\varphi(s)} \equiv_R \mu'$. By definition of branching steps, we trivially have $\hat{\varphi}(s) \stackrel{\tau}{\Longrightarrow}'_R \mathbb{1}_{\hat{\varphi}(s)}$. Note that indeed $\mathbb{1}_{\varphi(s)} \equiv_R \mathbb{1}_{\hat{\varphi}(s)}$, since $(\varphi(s), \hat{\varphi}(s)) \in R$.

(2) If $\varphi(s) \xrightarrow{a} \nu$, by definition of the extended arrow either (2.a) $a \in A$ and $\varphi(s) \overset{a}{\hookrightarrow} \mu$, or (2.b) $a = \chi(\lambda)$ for $\lambda = \mathit{rate}(\varphi(s))$, $\lambda > 0$, $\mu = \mathbb{P}_{\varphi(s)}$ and there is no μ' such that $\varphi(s) \overset{\tau}{\hookrightarrow} \mu'$.

In case of (2.a), by definition of the quotient we find that $\hat{\varphi}(s) \xrightarrow{a}_{\varphi} \varphi(\nu)$. Hence, also $\hat{\varphi}(s) \overset{a}{\Longrightarrow}'_R \varphi(\nu)$. As observed above, $\varphi(\nu) \equiv_R \nu$. Also, since $\mu \equiv_R \nu$ by assumption, transitivity of $\equiv_R$ yields $\mu \equiv_R \varphi(\nu)$.

In case of (2.b), $\varphi(s)$ has a total outgoing rate of λ and this is spread out according to μ. That is, for each state t, there is a rate of $\lambda \cdot \mu(t)$ from $\varphi(s)$ to t. Let $C = [t]_R$ for some state t, and let λ' be the total rate from $\varphi(s)$ to C. By definition of the quotient, this implies that there is a rate of λ' from $\hat{\varphi}(s)$ to $\hat{\varphi}(t)$ in $\mathcal{M}/\varphi$ as well. Since the only element of C reachable from $\hat{\varphi}(s)$ is $\hat{\varphi}(t)$, this implies that there is a rate from $\hat{\varphi}(s)$ to C of λ'. Hence, for an arbitrary equivalence class C we find identical rates from $\varphi(s)$ to C and from $\hat{\varphi}(s)$ to C. This immediately implies that the outgoing rates of $\varphi(s)$ and $\hat{\varphi}(s)$ coincide, and that $\mathbb{P}_{\varphi(s)} \equiv_R \mathbb{P}_{\hat{\varphi}(s)}$. By definition of extended transitions now $\hat{\varphi}(s) \xrightarrow{\chi(\lambda)}_{\varphi} \mathbb{P}_{\hat{\varphi}(s)}$, and hence $\hat{\varphi}(s) \overset{a}{\Longrightarrow}'_R \mathbb{P}_{\hat{\varphi}(s)}$. Since $\mu = \mathbb{P}_{\varphi(s)}$ and $\mathbb{P}_{\varphi(s)} \equiv_R \mathbb{P}_{\hat{\varphi}(s)}$, this completes this part of the proof.

It remains to show that R is divergence sensitive. So, let $(s, s') \in R$. Again, we make a case distinction based on the origin of s and s'. Like before, if both states are in $\mathcal{M}/\varphi$ then they coincide, and hence the result immediately follows. Also, if both states are in $\mathcal{M}$, then divergence of s' is implied by divergence of s. After all, having the same representative they must be connected by confluent transitions, and hence Theorem 14 and the fact that the quotient is not involved give the result.

So, we only need to show (1) whether divergence in a state s in $\mathcal{M}$ implies divergence in its representative $\hat{\varphi}(s)$ in $\mathcal{M}/\varphi$, and (2) whether divergence in a state $\hat{t} \in \mathcal{M}/\varphi$ implies divergence in all states s in $\mathcal{M}$ such that $\varphi(s) = t$.

(1) Assume that there is a diverging scheduler for some state s in $\mathcal{M}$. We need to show that there also is a diverging scheduler for $\hat{\varphi}(s)$ in $\mathcal{M}/\varphi$. First of all note that, by Theorem 14, divergence of s implies divergence of $\varphi(s)$ in $\mathcal{M}$. Hence, we can assume that there is a scheduler $\mathcal{S}$ such that

$$\forall \pi \in \mathit{finpaths}^{\mathcal{S}}_{\mathcal{M}}(\varphi(s)) \,.\, \mathit{trace}(\pi) = \epsilon \wedge \mathcal{S}(\pi)(\bot) = 0$$

It is well known that we can assume that this diverging scheduler is memoryless and deterministic.

By the existence of $\mathcal{S}$, there must be some transition $\varphi(s) \overset{\tau}{\hookrightarrow} \mu$ such that every state $t \in \mathrm{spt}(\mu)$ is also diverging. By the definition of the quotient, then there also is a transition $\hat{\varphi}(s) \xrightarrow{\tau}_{\varphi} \varphi(\mu)$ in $\mathcal{M}/\varphi$. We can now construct a diverging scheduler for $\hat{\varphi}(s)$ that starts with this transition. Then, it invisibly ends up in either one of a set of states that are all representatives of diverging states. From all those states, the above argument can be repeated to take the next invisible transition. As this process can be extended indefinitely, indeed $\hat{\varphi}(s)$ is diverging too.

(2) Assume that there is a scheduler $\mathcal{S}$ such that

$$\forall \pi \in \mathit{finpaths}^{\mathcal{S}}_{\mathcal{M}/\varphi}(\hat{s}) \,.\, \mathit{trace}(\pi) = \epsilon \wedge \mathcal{S}(\pi)(\bot) = 0$$

for some state $\hat{s}$ in $\mathcal{M}/\varphi$. It is well known that we can assume that this diverging scheduler is memoryless and deterministic.

We need to show that there also is a diverging scheduler for every state s' in $\mathcal{M}$ such that $\varphi(s') = s$. First of all note that, by Theorem 14, divergence of $\varphi(s')$ implies divergence of s' in $\mathcal{M}$. Hence, it suffices to show divergence of $\varphi(s')$ based on divergence of $\hat{\varphi}(s')$ $(= \hat{s})$.

By the existence of $\mathcal{S}$, there must be some transition $\hat{\varphi}(s') \xhookrightarrow{\tau}_{\varphi} \mu$ such that every state $\hat{t} \in \mathrm{spt}(\mu)$ is also diverging. By the definition of the quotient, then there also is a transition $\varphi(s') \xhookrightarrow{\tau} \nu$ in $\mathcal{M}$ such that $\varphi(\nu) = \mu$. Hence, we can now construct a diverging scheduler for $\varphi(s')$ that starts with this transition. Then, it invisibly ends up in either one of a set of states that all have a diverging representative. From all those states, the above argument can be repeated to take the next invisible transition. As this process can be extended indefinitely, indeed $\varphi(s')$ (and hence s') is diverging too. □