On the Semantics of Markov Automata

Yuxin Deng[1,3,4,⋆] and Matthew Hennessy[2,⋆⋆]

[1] Shanghai Jiao Tong University
[2] Trinity College Dublin
[3] Carnegie Mellon University
[4] Chinese Academy of Sciences

Abstract. Markov automata describe systems in terms of events which may be nondeterministic, may occur probabilistically, or may be subject to time delays. We define a novel notion of weak bisimulation for such systems and prove that this provides both a sound and complete proof methodology for a natural extensional behavioural equivalence between such systems, a generalisation of *reduction barbed congruence*, the well-known touchstone equivalence for a large variety of process description languages.

1 Introduction

Markov Automata (MA) [8] describe system behaviour in terms of non-deterministic, probabilistic and timed events. The first two kinds of events are well-known from Probabilistic Automata (PA) [22,23] and Probabilistic Labelled Transition Systems (pLTSs) [5], while the third are taken to be random delays, governed by negative exponential distributions parametrised by some delay $\lambda \in \mathbb{R}^+$. As explained in [10], timed events can be given a straightforward operational semantics in terms of their parametric delays.

For example, consider the MAs in Figure 1, taken from [8]. From the initial state of the first automaton, s, there is a race between two possible timed events, denoted by double-headed arrows, each governed by the same rate, 4λ, for some arbitrary $\lambda \in \mathbb{R}^+$. If one of these events wins, the state of the automaton changes to s_a, from which some external action a can happen. If the other timed event wins, the change of state is to s_1, from which an internal unobservable action, denoted by τ, can occur. Moreover, the effect of this internal action is probabilistic; fifty percent of the time the state change will be to s_b, where action b can occur, while with the same probability the change will be to s_c, where c can occur. Formally this probabilistic behaviour is represented as an action from a state, such as s_1, to a distribution over states, represented as a darkened circle connected to states in the support of the distribution, labelled with their probabilities.

On the other hand the second automaton is much more straightforward. From its initial state there is a race between three timed events, two running at the same rate and one at double the rate. Then one of the (external) actions a, b, c occurs depending on which event wins the race.

⋆ Partially supported by the Natural Science Foundation of China under Grant No. 61033002, the Qatar National Research Fund under grant NPRP 09-1107-1-168, and the Opening Fund of Top Key Discipline of Computer Software and Theory in Zhejiang Provincial Colleges at Zhejiang Normal University.

⋆⋆ Supported financially by SFI project no. SFI 06 IN.1 1898.

L. Aceto, M. Henzinger, and J. Sgall (Eds.): ICALP 2011, Part II, LNCS 6756, pp. 307–318, 2011.

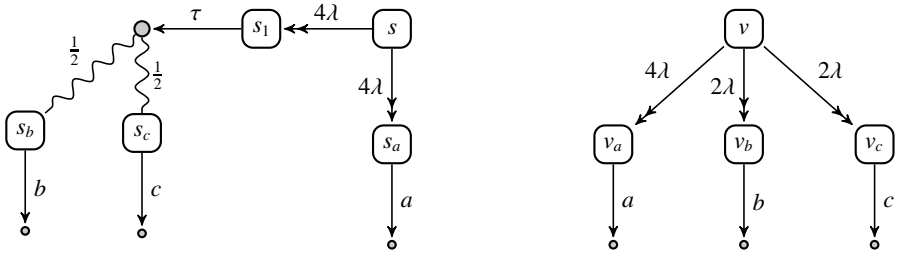

Fig. 1. Timed transitions and distributions

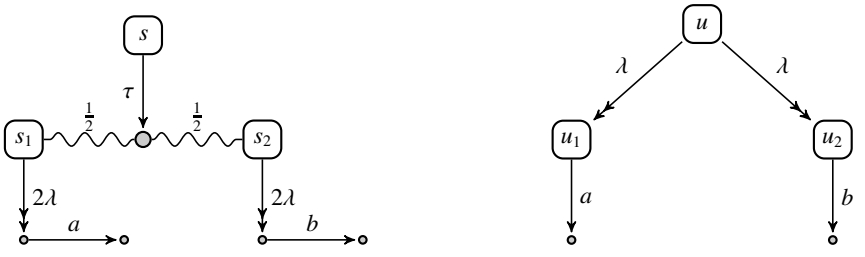

Fig. 2. Timed transitions and distributions, again

Providing a satisfactory behavioural model of MAs is necessarily a complicated undertaking. But as pointed out in [8], because of the nature of their underlying distributions, the timed events can be satisfactorily explained in terms of simple probabilistic distributions determined by their rates. They propose a translation of MAs into PAs (cf. Section 2). Since behavioural theories have already been developed for PAs [24,6,18,4], we therefore automatically obtain such theories for MAs, via their induced PAs.

However, if one uses a standard behavioural theory for PAs, such as weak bisimulation equivalence as defined in [15,24,18,6] then the two MAs in Figure 1 are distinguished. Instead the authors of [8] propose a new bisimulation equivalence between PAs, which enjoys standard properties such as compositionality, and which identifies these two MAs. But as the authors point out their equivalence still distinguishes between the MAs in Figure 2. The question naturally arises: which MAs should be distinguished, and which be deemed equivalent. This is the topic of the current paper.

We approach the question indirectly, by giving criteria for reasonable behavioural equivalences between MAs; this induces a *touchstone* extensional equivalence between systems, namely the largest equivalence, $\approx_{\text{behav}}$, which satisfies these criteria. Thus two MAs should only be distinguished on the basis of the chosen criteria.

Having an independent notion of which systems should, and which should not, be distinguished, one can then justify a particular notion of bisimulation by showing that it captures precisely the touchstone equivalence, $\approx_{\text{behav}}$. In other words, a particular definition of bisimulation is appropriate because $\approx_{bis}$, the associated bisimulation equivalence,

(i) is *sound* w.r.t. the touchstone equivalence, that is $s_1 \approx_{bis} s_2$ implies $s_1 \approx_{\text{behav}} s_2$
(ii) provides a *complete* proof methodology for the touchstone equivalence, that is $s_1 \approx_{\text{behav}} s_2$ implies $s_1 \approx_{bis} s_2$.

This approach originated in [13] but has now been widely used for different process description languages; for example see [14,20] for its application to higher-order process languages, [19] for mobile ambients and [9] for asynchronous languages. Moreover in each case the distinguishing criteria are more or less the same. The touchstone equivalence should be

(i) *compositional*; that is preserved by natural operators for constructing systems
(ii) *barb-preserving*; barbs are simple experiments observers perform on systems
(iii) *reduction-closed*; this is a natural condition on the reduction semantics of systems which ensures that nondeterministic choices are in some sense preserved.

We adapt this approach to MAs. Using natural versions of these criteria for MAs we obtain an appropriate touchstone equivalence, *reduction barbed congruence* ($\approx_{rbc}$). We then develop a new theory of bisimulations which is both sound and complete for $\approx_{rbc}$.

The remainder of the paper is organised as follows. In the next section we give our definition of Markov automata MA, a slight generalisation of that in [8]; in addition to the timed events parametrised on specific delays, we have special timed events which have indefinite, or imprecise delay times associated with them. In order to model the delay operators probabilistically, we then show how to translate an MA into a PA, as suggested in [8]. For this purpose we use a slight variation, called MLTSs, in which there are distinguished actions labelled by weights. We then develop our new definition of bisimulation equivalence for MLTSs, thereby inducing bisimulation equivalence between MAs; this construction is illustrated via examples. In Section 3 we show how MAs can be composed, using a parallel operator based on CCS [17]. In fact this is extended to an interpretation of an Markovian extension of CCS, mCCS, as an MA. We then show that bisimulation equivalence is preserved by this form of composition.

Section 4 contains the main theoretical results of the paper. We give a formal definition of the touchstone equivalence $\approx_{rbc}$, and outline the proof that this is captured precisely by our new notion of bisimulation. The paper ends with a brief comparison with related work in Section 5.

2 Markov Automata

A (discrete) probability subdistribution over a set S is a function $\Delta : S \to [0,1]$ with $\sum_{s \in S} \Delta(s) \leq 1$; the support of such an Δ is the set $\lceil \Delta \rceil = \{ s \in S \mid \Delta(s) > 0 \}$. A subdistribution is a (total, or full) distribution if $\sum_{s \in S} \Delta(s) = 1$. The point distribution $\overline{s}$ assigns probability 1 to s and 0 to all other elements of S, so that $\lceil \overline{s} \rceil = s$. We use $\mathcal{D}_{sub}(S)$ to denote the set of subdistributions over S, and $\mathcal{D}(S)$ its subset of full distributions.

We write $\mathbb{R}^+$ for the set of all positive real numbers. Let $\{\Delta_k \mid k \in K\}$ be a set of subdistributions, possibly infinite. Then $\sum_{k \in K} \Delta_k$ is the real-valued function in $S \to \mathbb{R}^+$ defined by $(\sum_{k \in K} \Delta_k)(s) := \sum_{k \in K} \Delta_k(s)$. This is a partial operation on subdistributions because for some state s the sum of $\Delta_k(s)$ might exceed 1. If the index set is finite, say $\{1..n\}$, we often write $\Delta_1 + \ldots + \Delta_n$. For p a real number from $[0,1]$ we use $p \cdot \Delta$ to denote the subdistribution given by $(p \cdot \Delta)(s) := p \cdot \Delta(s)$. If $\sum_{k \in K} p_k = 1$ for some collection of $p_k \geq 0$, and the Δ_k are distributions, then so is $\sum_{k \in K} p_k \cdot \Delta_k$.

Definition 1. *A* Markov automaton *(MA), is a quadruple* $\langle S, \mathsf{Act}_\tau, \rightarrow, \mapsto \rangle$*, where*

(i) S *is a set of states*
(ii) Act_τ *is a set of transition labels, with distinguished element* τ
(iii) the relation $\rightarrow$ *is a subset of* $S \times \mathsf{Act}_\tau \times \mathcal{D}(S)$
(iv) the relation $\mapsto$ *is a subset of* $S \times (\mathbb{R}^+ \cup \{\delta\}) \times \mathcal{D}(S)$

satisfying (a) $s \overset{\boldsymbol{d}}{\mapsto} \Delta$ *implies* $s \not\xrightarrow{\tau}$, $\boldsymbol{d} = \delta$ *or* λ*, (b)* $s \overset{\delta}{\mapsto} \Delta_1$ *and* $s \overset{\delta}{\mapsto} \Delta_2$ *implies* $\Delta_1 = \Delta_2$*. The MA is* finitary *if* S *is finite and each state has only finitely many outgoing transitions.*

This is a mild generalisation of the MAs in [8]; for example we allow the residual of a timed action to be a distribution, and *maximal progress*, assumption (a), is built in to the definition. But the major extension is the introduction of the indefinite delay actions denoted by the special action δ, $s \overset{\delta}{\mapsto} \Delta$; this can be viewed as a timed action whose underlying rate is unknown. Such indefinite actions, often called *passive* when they are external, are widely used in the literature [3,12], although their precise properties vary between publications; see [10], page 66 for a discussion.

Following [8], we study MAs indirectly, by considering their derived structures.

Definition 2. *A* Markov labelled transition system *(MLTS) is a triple* $\langle S, \mathsf{Act}_\tau, \rightarrow \rangle$*, where* S *and* Act_τ *are as in Definition 1, and* $\rightarrow$ *is a subset of* $S \times (\mathsf{Act}_{\tau,\delta} \cup \mathbb{R}^+) \times \mathcal{D}(S)$ *satisfying* $s \xrightarrow{\lambda_1} \Delta_1$ *and* $s \xrightarrow{\Delta_2} \Delta_2$ *implies* $\lambda_1 = \lambda_2$ *and* $\Delta_1 = \Delta_2$*, in addition to the constraints corresponding to (a) and (b) in Definition 1. Here* $\mathsf{Act}_{\tau,\delta}$ *means* $\mathsf{Act}_\tau \cup \{\delta\}$. □

A (non-probabilistic) labelled transition system (LTS) may be viewed as a degenerate MLTS — one in which only point distributions are used, and the special actions labelled by δ and $\lambda \in \mathbb{R}^+$ are vacuous. An MLTS is *finitary* if the state set S is finite and for each $s \in S$ the set $\{(\mu, \Delta) \mid s \xrightarrow{\mu} \Delta,\ \mu \in \mathsf{Act}_{\tau,\delta} \cup \mathbb{R}^+,\ \Delta \in \mathcal{D}(S)\}$ is finite.

Admittedly MAs and MLTSs are very similar; the difference lies in the intent. We interpret the former in the latter, thereby modelling the passage of time probabilistically. The essential ingredient in the interpretation is the function on the states of an MA, defined by $\mathsf{Rate}(s) = \sum\{\lambda_i \mid s \overset{\lambda_i}{\mapsto} \Delta_i\}$. Given an MA M as in Definition 1 the MLTS $\mathsf{mlts}(M)$ is given by $\langle S, \mathsf{Act}_\tau, \rightarrow \rangle$ where:

(a) for $\mu \in \mathsf{Act}_\tau$ the actions $s \xrightarrow{\mu} \Delta$ are inherited from M
(b) $s \xrightarrow{\delta} \Delta$ whenever $s \overset{\delta}{\mapsto} \Delta$ in M
(c) for $\lambda \in \mathbb{R}^+$, $s \xrightarrow{\lambda} \Delta$ if $\mathsf{Rate}(s) = \lambda > 0$ and $\Delta = \sum\{p_i \cdot \Delta_i \mid s \overset{\lambda_i}{\mapsto} \Delta_i\}$ where $p_i = \frac{\lambda_i}{\mathsf{Rate}(s)}$

Example 1. The derived MLTSs of the two MAs in Figure 1 are given in Figure 3. Note that the time dependent race between the evolution of s to s_a or s_1 in Figure 1 is represented in Figure 3 by a single arrow labelled by the total rate of s to a distribution representing the chances of s_1 and s_2 winning the race. Similarly in the second MA the race from v to v_a, v_b, v_c is now represented by a single weighted arrow to a similar distribution. The weights on these arrows will be used for compositional reasoning. □

In an MLTS actions are only performed by states, but in general we allow distributions over states to perform an action. For this purpose, we *lift* these relations so that they also apply to subdistributions [5].

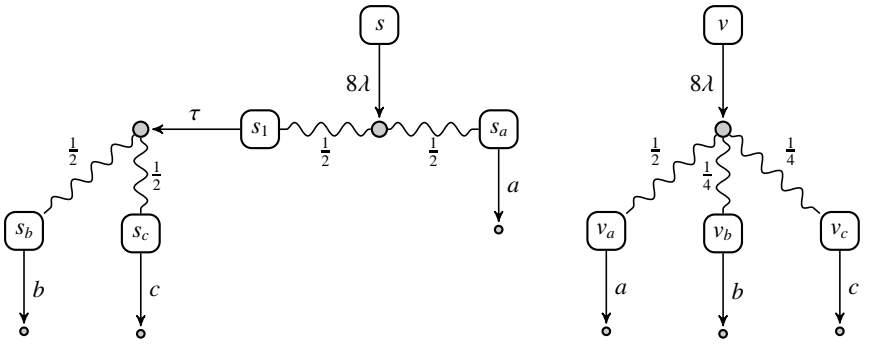

Fig. 3. Derived MLTSs of MAs in Figure 1

Definition 3. *Let $\mathcal{R} \subseteq S \times \mathcal{D}_{sub}(S)$ be a relation from states to subdistributions in an MLTS. Then $\overline{\mathcal{R}} \subseteq \mathcal{D}_{sub}(S) \times \mathcal{D}_{sub}(S)$ is the smallest relation that satisfies*

(i) $s\ \mathcal{R}\ \Theta$ implies $\overline{s}\ \overline{\mathcal{R}}\ \Theta$, and
(ii) $\Delta_i\ \overline{\mathcal{R}}\ \Theta_i$ for $i \in I$ implies $(\sum_{i\in I} p_i \cdot \Delta_i)\ \overline{\mathcal{R}}\ (\sum_{i\in I} p_i \cdot \Theta_i)$ for any $p_i \in [0,1]$ with $\sum_{i\in I} p_i \leq 1$. □

We apply this operation to the relations $\xrightarrow{\mu}$ in the MLTS for $\mu \in \mathsf{Act}_{\tau,\delta}$, where we also write $\xrightarrow{\mu}$ for $\overline{\xrightarrow{\mu}}$. Thus as source of a relation $\xrightarrow{\mu}$ we now also allow distributions, and even subdistributions.

Definition 4 (Hyper-derivations). *In an MLTS a hyper-derivation consists of a collection of subdistributions $\Delta, \Delta_k^{\rightarrow}, \Delta_k^{\times}$, for $k \geq 0$, with the following properties:*

$$\begin{aligned} \Delta &= \Delta_0^{\rightarrow} + \Delta_0^{\times} \\ \Delta_0^{\rightarrow} &\xrightarrow{\tau} \Delta_1^{\rightarrow} + \Delta_1^{\times} \\ &\vdots \\ \Delta_k^{\rightarrow} &\xrightarrow{\tau} \Delta_{k+1}^{\rightarrow} + \Delta_{k+1}^{\times} \\ &\vdots \\ \Delta' &= \sum_{k=0}^{\infty} \Delta_k^{\times} \end{aligned} \tag{1}$$

We call Δ' a hyper-derivative *of Δ, and write $\Delta \Longrightarrow \Delta'$.* □

We refer to [5] for more discussion about hyper-derivations.
With these concepts we can now define the appropriate notion of weak moves in a MLTS, with which we may then use to define our concept of bisimulations. We write $\Delta \overset{\tau}{\Longrightarrow} \Delta'$ to mean $\Delta \Longrightarrow \Delta'$ and $\Delta \overset{\alpha}{\Longrightarrow} \Delta'$, for $\alpha \in \mathsf{Act}_\delta \cup \mathbb{R}^+$, to mean $\Delta \Longrightarrow \xrightarrow{\lambda} \Longrightarrow \Delta'$.

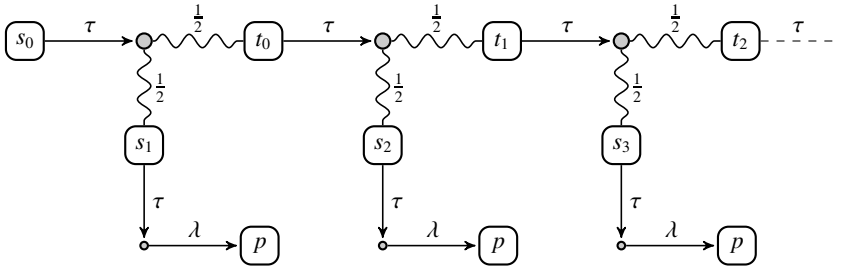

Fig. 4. Limiting internal moves

Definition 5 (Bisimulations). *For $\mathcal{R} \subseteq \mathcal{D}(S) \times \mathcal{D}(S)$, where S is the set of states in an MLTS, let $\mathcal{B}(\mathcal{R})$ be the relation over $\mathcal{D}(S) \times \mathcal{D}(S)$ determined by letting $\Delta\ \mathcal{B}(\mathcal{R})\ \Theta$ if, for each $\mu \in \mathsf{Act}_{\tau,\delta} \cup \mathbb{R}^+$ and all finite sets of probabilities $\{ p_i \mid i \in I \}$ satisfying $\sum_{i\in I} p_i = 1$,*

(i) whenever $\Delta \stackrel{\mu}{\Longrightarrow} \sum_{i\in I} p_i \cdot \Delta_i$, for any distributions Δ_i, there are distributions Θ_i with $\Theta \stackrel{\mu}{\Longrightarrow} \sum_{i\in I} p_i \cdot \Theta_i$, such that $\Delta_i\ \mathcal{R}\ \Theta_i$ for each $i \in I$

(ii) symmetrically, whenever $\Theta \stackrel{\mu}{\Longrightarrow} \sum_{i\in I} p_i \cdot \Theta_i$, for any distributions Θ_i, there are distributions Δ_i with $\Delta \stackrel{\mu}{\Longrightarrow} \sum_{i\in I} p_i \cdot \Delta_i$, such that $\Delta_i\ \mathcal{R}\ \Theta_i$ for each $i \in I$.

The largest solution to $\mathcal{R} = \mathcal{B}(\mathcal{R})$ is denoted by $\approx_{bis}$. □

Because of the form of the functional $\mathcal{B}$ it is easy to establish that $\approx_{bis}$ is an equivalence relation. However, due to the use of weak arrows, and the quantification over sets of probabilities, it is not easy to exhibit witness bisimulations. So we give an alternative characterisation of $\approx_{bis}$ in terms of a relation between *states* and distributions.

Definition 6 (Simple bisimulations). *For $\mathcal{R} \subseteq S \times \mathcal{D}(S)$, where again S is the set of states in an MLTS, let $\mathcal{SB}(\mathcal{R})$ be the relation over $S \times \mathcal{D}(S)$ defined by letting $s\ \mathcal{SB}(\mathcal{R})\ \Theta$ if, for each $\mu \in \mathsf{Act}_{\tau,\delta} \cup \mathbb{R}^+$,*

(i) whenever $s \stackrel{\mu}{\longrightarrow} \Delta'$, there is some $\Theta \stackrel{\mu}{\Longrightarrow} \Theta'$, such that $\Delta'\ \overline{\mathcal{R}}\ \Theta'$

(ii) there exists some $\Delta \in \mathcal{D}(S)$ such that $\overline{s} \stackrel{\tau}{\Longrightarrow} \Delta$ and $\Theta\ \overline{\mathcal{R}}\ \Delta$.

We use $\approx_{sbis}$ to denote the largest solution to $\mathcal{R} = \mathcal{SB}(\mathcal{R})$. □

Example 2. Consider again Figure 3. We have $s \approx_{sbis} \overline{v}$ because the following relation $\{\langle s, \overline{v}\rangle,\ \langle s_1, \frac{1}{2}\cdot\overline{v_b} + \frac{1}{2}\cdot\overline{v_c}\rangle,\ \langle s_a, \overline{v_a}\rangle,\ \langle s_b, \overline{v_b}\rangle,\ \langle s_c, \overline{v_c}\rangle,\ \langle v, \overline{s}\rangle,\ \langle v_a, \overline{s_a}\rangle,\ \langle v_b, \overline{s_b}\rangle,\ \langle v_c, \overline{s_c}\rangle\}$ is a simple bisimulation and therefore $s \approx_{sbis} \overline{v}$.

Consider the MA in Figure 4; starting from the initial state s_0 an ever increasing number of internal τ moves are performed before the eventual timed λ action, but with ever decreasing probability. The relation

$$\{\, \langle \lambda.p, \overline{s_i}\rangle,\ \langle \lambda.p, \overline{t_i}\rangle,\ \langle s_i, \overline{\lambda.p}\rangle,\ \langle t_i, \overline{\lambda.p}\rangle \mid i \geq 0 \,\}$$

is a simple bisimulation, and therefore $s_0 \approx_{sbis} \overline{\lambda.p}$, where $\lambda.p$ describes in an obvious manner the MA which does a timed action at rate λ and evolves to the state p.

Now consider the MA in Figure 2. We have $s \not\approx_{sbis} \overline{u}$ because the transition $s \xrightarrow{\tau} \frac{1}{2} \cdot \overline{s_1} + \frac{1}{2} \cdot \overline{s_2}$ cannot be matched by any transition from u. The state u cannot enable internal actions, so the only weak internal transition from $\overline{u}$ is $\overline{u} \stackrel{\tau}{\Longrightarrow} \overline{u}$. However, the derivative $\overline{u}$ is not able to simulate $\frac{1}{2} \cdot \overline{s_1} + \frac{1}{2} \cdot \overline{s_2}$ according to the lifted relation $\overline{\approx_{sbis}}$. Suppose for a contradiction that $(\frac{1}{2} \cdot \overline{s_1} + \frac{1}{2} \cdot \overline{s_2}) \; \overline{\approx_{sbis}} \; \overline{u}$. Then we must have $s_1 \approx_{sbis} \overline{u}$ and $s_2 \approx_{sbis} \overline{u}$; obviously neither of these hold. □

The two relations $\approx_{bis}$ and $\approx_{sbis}$ are closely related, as stated by the theorem below.

Theorem 1. *Let Δ and Θ be two distributions in a finitary MLTS.*

(i) If $\Delta \approx_{bis} \Theta$ then there is some Θ' with $\Theta \stackrel{\tau}{\Longrightarrow} \Theta'$ and $\Delta \; \overline{\approx_{sbis}} \; \Theta'$
(ii) If $\Delta \; \overline{\approx_{sbis}} \; \Theta$ then $\Delta \approx_{bis} \Theta$. □

For the remainder of the paper we will apply these relations, developed for MLTSs, to the states and distributions of MAs. For example we write $s \approx_{sbis} \Delta$, where s is a state in an MA M and Δ a distribution, to mean $s \approx_{sbis} \Delta$ in the derived $\mathsf{mlts}(M)$.

3 Composing Markov Automata

Here we assume that the set of actions Act is equipped with a complementation function $\bar{\cdot} : \mathsf{Act} \to \mathsf{Act}$ satisfying $\overline{\overline{a}} = a$; we say $\overline{a}$ is the complement of a. Then given two MAs, $M_i = \langle S_1, \mathsf{Act}_\tau, \to, \mapsto, \rangle$ for $i = 1, 2$, their composition $(M_1 \mid M_2)$ is given by $\langle S_1 \mid S_2, \mathsf{Act}_\tau, \to, \mapsto, \rangle$ where the set of states $S_1 \mid S_2 = \{ s_1 \mid s_2 \mid s_i \in S_i, i = 1, 2 \}$ and the relations are determined by the rules in Figure 5. The rules use the obvious extension of the function $\mid$ on pairs of states to pairs of distributions. To be precise $\Delta|\Theta$ is the distribution defined by: $(\Delta|\Theta)(s) = \Delta(s_1) \times \Theta(s_2)$ if $s = s_1|s_2$, and 0 otherwise. It can be checked that if M_1 and M_2 are Markov automata, then so is $(M_1 \mid M_2)$. We can internalise this composition relation by saying an Markov automaton M is *par-closed* if $(M \mid M)$ is already a sub-MA of M.

The simplest way of constructing a par-closed MA is by interpreting a process algebra as a universal Markov automaton. To this end we introduce the language mCCS whose terms are given by:

$$P, Q ::= \mathbf{0} \mid \delta.P \mid \lambda.D, \; \lambda \in \mathbb{R}^+ \mid \mu{:}D, \; \mu \in \mathsf{Act}_\tau \mid P + Q \mid P \mid Q \mid A$$
$$D ::= (\oplus_{i \in I} p_i \cdot P_i)$$

where A ranges over a set of process constants, with each of which is associated a definition, $A \Leftarrow \mathrm{Def}(A)$. mCCS is interpreted as an Markov automaton whose states are all the terms in the language, and whose arrows are determined by the rules in Figure 6, together with those in Figure 5; we have omitted the obvious symmetric counterparts to the rules (EXT.L), (EXT.L.L) and (EXT.D.L). Other operations, such as the standard hiding $Q\backslash a, a \in Act$, can also be easily given an interpretation. We say a process P from mCCS is *finitary* if the sub-MA consisting of all states reachable from P is finitary, and we use *finitary* mCCS to refer to the MA consisting of all such finitary P.

The rules (ACTION) and (DELAY) use the notation $[\![D]\!]$, where D has the form $(\oplus_{i \in I} p_i \cdot P_i)$, to denote the obvious distribution over process terms, whose support consists of

$$\frac{s \xrightarrow{\mu} \Delta}{s \mid t \xrightarrow{\mu} \Delta \mid \overline{t}}\ (\textsc{par.l}) \qquad \frac{t \xrightarrow{\mu} \Theta}{s \mid t \xrightarrow{\mu} \overline{s} \mid \Theta}\ (\textsc{par.r})$$

$$\frac{s \xrightarrow{a} \Delta,\ t \xrightarrow{\overline{a}} \Theta}{s \mid t \xrightarrow{\tau} \Delta \mid \Theta}\ (\textsc{par.i})$$

$$\frac{s \overset{\mathbf{d}}{\mapsto} \Delta,\ t \overset{\delta}{\mapsto} \Theta,\ s \mid t \not\xrightarrow{\tau}}{s \mid t \overset{\mathbf{d}}{\mapsto} \Delta \mid \Theta}\ (\textsc{par.l.t}) \qquad \frac{s \overset{\delta}{\mapsto} \Delta, t \overset{\mathbf{d}}{\mapsto} \Theta,\ s \mid t \not\xrightarrow{\tau}}{s \mid t \overset{\mathbf{d}}{\mapsto} \Delta \mid \Theta}\ (\textsc{par.r.t})\ \mathbf{d} = \delta, \lambda$$

Fig. 5. Composing Markov automata

$P_1, \ldots P_n$, each with weight p_i respectively. Most of the other rules should be self-explanatory, although the justification for the rules for λ transitions depends on non-trivial properties of exponential distributions, as explained in detail in [10].

Nevertheless, this interpretation of mCCS is quite different than that of other Markovian process calculi, such as those in [10,3]. First the actions $\mu : D$ are *insistent* rather than *lazy*; they do not allow time to pass. For example the process $(\lambda.Q \mid a{:}P)$ is stuck with respect to time; it can not perform any timed action. This is because the parallel operator requires each component to perform a timed action, which $a : P$ can not do, before time can pass. To obtain lazy actions one can define $a.P$ by the declaration $A \Leftarrow a{:}P + \delta.A$. Then we have the transition $\lambda.Q \mid a.P \overset{\lambda}{\mapsto} Q \mid a.P$ by an application of the rule (PAR.L.T) to the transitions $\lambda.Q \overset{\lambda}{\mapsto} Q$ and $a.P \overset{\delta}{\mapsto} a.P$.

The parallel operator is even more constraining in that at most one of its components can perform a definite delay. Again this is reminiscent of many existing Markovian process algebras [2,3], although these tend to have delays associated with external actions. But in the setting of mCCS the net effect is an operational semantics very similar to that in [8]. For example consider the process $Q = (\lambda_1.P_1 \mid \lambda_2.P_2)$. This has three timed actions:

(i) $Q \overset{\lambda_1}{\mapsto} (P_1 \mid \lambda_2.P_2)$ via an application of the rule (PAR.L.T) to the actions $\lambda_1.P_1 \overset{\lambda_1}{\mapsto} P_1$ and $\lambda_2.P_2 \overset{\delta}{\mapsto} \lambda_2.P_2$

(ii) $Q \overset{\lambda_2}{\mapsto} (\lambda_1 P_2 \mid P_2)$ via an application of (PAR.R.T) to the actions $\lambda_1.P_1 \overset{\delta}{\mapsto} \lambda_1.P$ and $\lambda_2.P_2 \overset{\lambda_2}{\mapsto} P_2$

(iii) $Q \overset{\delta}{\mapsto} Q$ via an application of either of (PAR.L.T) or (PAR.R.T) to the transitions $\lambda_1.P_1 \overset{\delta}{\mapsto} \lambda_1.P_1$ and $\lambda_1.P_1 \overset{\delta}{\mapsto} \lambda_1.P_1$.

Theorem 2 (Compositionality). *Let Δ, Θ and Γ be any distributions in a finitary par-closed MA. If $\Delta \approx_{bis} \Theta$ then $\Delta \mid \Gamma \approx_{bis} \Theta \mid \Gamma$.* □

$$\text{(ACTION)}\quad \mu{:}D \xrightarrow{\mu} [\![D]\!] \qquad \text{(RECURSION)}\quad \frac{\mathrm{Def}(A) \xrightarrow{\alpha} \Delta}{A \xrightarrow{\alpha} \Delta}\ \alpha = \mu, \lambda, \delta$$

$$\text{(EXT.L)}\quad \frac{P \xrightarrow{\mu} \Delta,}{P + Q \xrightarrow{\mu} \Delta} \qquad \text{(EXT.L.L)}\quad \frac{P \overset{\lambda}{\mapsto} \Delta,\ Q \not\xrightarrow{\tau}}{P + Q \overset{\lambda}{\mapsto} \Delta}$$

$$\text{(DELAY)}\quad \lambda.D \overset{\lambda}{\mapsto} [\![D]\!], \qquad \text{(D.}\delta\text{)}\quad \lambda.D \overset{\delta}{\mapsto} \overline{\lambda.D}$$

$$\text{(}\delta\text{.E)}\quad \frac{P \xrightarrow{\mu} \Delta}{\delta.P \xrightarrow{\mu} \Delta} \qquad \text{(}\delta\text{.D)}\quad \frac{P \not\xrightarrow{\tau}}{\delta.P \overset{\delta}{\mapsto} \overline{P}}$$

$$\text{(EXT)}\quad \frac{P \overset{\delta}{\mapsto} \Delta_1,\ Q \overset{\delta}{\mapsto} \Delta_2}{P + Q \overset{\delta}{\mapsto} \Delta_1 + \Delta_2} \qquad \text{(EXT.D.L)}\quad \frac{P \overset{\delta}{\mapsto} \Delta,\ Q \not\overset{\delta}{\mapsto},\ Q \not\xrightarrow{\tau}}{P + Q \overset{\delta}{\mapsto} \Delta}$$

Fig. 6. Operational semantics of mCCS

4 Soundness and Completeness

Consider an arbitrary par-closed MA $M = \langle S, \mathsf{Act}_\tau, \rightarrow, \mapsto \rangle$. Experimenting on processes in M consists in observing what communications a process can perform, as it evolves by both internal moves and the passage of time. To make this *evolution* precise let $\Delta \Longrightarrow \Delta'$ be the least reflexive relation satisfying:
(a) $\Delta \Longrightarrow \Delta_1$ and $\Delta_1 \xrightarrow{\tau} \Delta'$ implies $\Delta \Longrightarrow \Delta'$
(b) $\Delta \Longrightarrow \Delta_1$ and $\Delta_1 \xrightarrow{\lambda} \Delta'$ implies $\Delta \Longrightarrow \Delta'$, where $\lambda \in \mathbb{R}^+$
(c) $\Delta_i \Longrightarrow \Delta_i'$ for each $i \in I$ implies $(\sum_{i\in I} p_i \cdot \Delta_i) \Longrightarrow (\sum_{i\in I} p_i \cdot \Delta_i')$ where $\sum_{i\in I} p_i = 1$.

Thus $\Delta \Longrightarrow \Delta'$ is a relation between distributions in the $\mathsf{mlts}(M)$ which allows reduction either by internal actions τ or definite delay actions λ; with the latter the reductions are to distributions determined by the rates of the states in the support of Δ.

Definition 7 (Barbs). *For $\Delta \in \mathcal{D}(S)$ and $a \in \mathsf{Act}$, let $\mathcal{V}_a(\Delta) = \sum\{\Delta(s) \mid s \xrightarrow{a}\}$. We write $\Delta \Downarrow_a^{\geq p}$ whenever $\Delta \Longrightarrow \Delta'$, where $\mathcal{V}_a(\Delta') \geq p$.* □

Then we say a relation $\mathcal{R}$ is *barb-preserving* if $\Delta\ \mathcal{R}\ \Theta$ then $\Delta \Downarrow_a^{\geq p}$ iff $\Theta \Downarrow_a^{\geq p}$. It is *reduction-closed* if $\Delta\ \mathcal{R}\ \Theta$ implies
(i) whenever $\Delta \Longrightarrow \Delta'$, there is a $\Theta \Longrightarrow \Theta'$ such that $\Delta'\ \mathcal{R}\ \Theta'$
(ii) whenever $\Theta \Longrightarrow \Theta'$, there is a $\Delta \Longrightarrow \Delta'$ such that $\Delta'\ \mathcal{R}\ \Theta'$.

Finally, we say a relation $\mathcal{R}$ is *compositional* if $\Delta_1\ \mathcal{R}\ \Delta_2$ implies $(\Delta_1 \mid \Theta)\ \mathcal{R}\ (\Delta_2 \mid \Theta)$.

Definition 8. *In a par-closed MA, let $\approx_{rbc}$ be the largest relation over the states which is barb-preserving, reduction-closed and compositional.* □

Example 3. Consider the two processes $P_1 = \lambda_1.Q_1$ and $P_2 = \lambda_2.Q_2$ where $\lambda_1 < \lambda_2$ and Q_i are two arbitrary processes. We can show that $P_1 \not\approx_{rbc} P_2$ by exhibiting a *testing*

process T such that the barbs of $(P_1 \mid T)$ and $(P_2 \mid T)$ are different. For example let $T = \delta.\tau.\mathbf{0} + \lambda_1.succ$. In $(P_i \mid T)$ there is a race between two timed events; in $(P_2 \mid T)$ their rates are λ_1 versus λ_2 while in $(P_1 \mid T)$ both events have the same rate. If the timed event in the test wins out, the action *succ* will occur. Consequently $(P_1 \mid T) \Downarrow_{succ}^{\geq \frac{1}{2}}$. However $(P_2 \mid T)$ does not have this barb; instead $(P_2 \mid T) \Downarrow_{succ}^{\geq q}$, where $q = \frac{\lambda_1}{\lambda_1 + \lambda_2}$; q is strictly smaller than $\frac{1}{2}$ since $\lambda_1 < \lambda_2$.

It follows that $\lambda_1.\lambda_2.P \not\approx_{rbc} \lambda_2.\lambda_1.P$ when λ_1 and λ_2 are different. □

Example 4. Consider the processes $P_1 = a:Q$, $P_2 = a.Q$, and $P_3 = \lambda.P_2$, where Q is an arbitrary process, and we have seen that $a.Q$ is shorthand for a recursively defined process $A = a:Q + \delta.A$.

Note that according to our semantics P_1 does not let time pass. Let T be the testing process $\lambda.(\bar{a}.succ + \tau.\mathbf{0})$. The process $P_1 \mid T$ cannot evolve, thus $(P_1 \mid T) \not\Downarrow_{succ}^{>0}$[1]. However, we have $P_2 \mid T \overset{\lambda}{\mapsto} P_2 \mid (\bar{a}.succ + \tau.\mathbf{0}) \xrightarrow{\tau} Q \mid succ$, thus $(P_2 \mid T) \Downarrow_{succ}^{\geq 1}$. The only comparable barb for P_3 is $(P_3 \mid T) \Downarrow_{succ}^{\geq \frac{1}{2}}$, because if the timed event in the test takes place, then by maximal progress the τ action must happen before the timed event in the process. It follows that the three processes P_1, P_2 and P_3 can be distinguished. □

Although by definition $\approx_{rbc}$ is closed w.r.t. the evolution relation $\Longrightarrow$, in fact it is also closed w.r.t. the individual components, and indeed the definite delay operator.

Proposition 1. *Suppose $\Delta \approx_{rbc} \Theta$.*

(i) If $\Delta \overset{\mu}{\Longrightarrow} \Delta'$ with $\mu \in \mathsf{Act}_\tau$ then $\Theta \overset{\mu}{\Longrightarrow} \Theta'$ such that $\Delta' \approx_{rbc} \Theta'$.
(ii) If $\Delta \overset{\lambda}{\Longrightarrow} \Delta'$ with $\lambda \in \mathbb{R}^+$ then $\Theta \overset{\lambda}{\Longrightarrow} \Theta'$ such that $\Delta' \approx_{rbc} \Theta'$.
(iii) If $\Delta \overset{\delta}{\Longrightarrow} \Delta'$ then $\Theta \overset{\delta}{\Longrightarrow} \Theta'$ such that $\Delta' \approx_{rbc} \Theta'$. □

Example 5. Consider the two MAs s and u from Figure 2, discussed in the Introduction. Suppose $s \approx_{rbc} u$. Then by compositionality we have $s \mid T \approx_{rbc} u \mid T$, where T is the process $\tau.\delta.\bar{a}.succ + \tau.\delta.\bar{b}.succ$. Let Δ denote the point distribution $\overline{\mathbf{0} \mid succ}$. Since $s \mid T \Longrightarrow \Delta$, we have $(s \mid T) \Downarrow_{succ}^{\geq 1}$.

However, the weak derivatives of $u \mid T$ under the evolution relation are very few, and one can easily check that none will have exactly the barbs of Δ because if $(u \mid T) \Downarrow_{succ}^{\geq p}$ then p is at most $\frac{1}{2}$. It follows that s and u are indeed behaviourally different. □

Theorem 3 (Soundness). *In a finitary par-closed MA, if $\Delta \approx_{bis} \Theta$ then $\Delta \approx_{rbc} \Theta$.* □

In order to obtain *completeness*, the converse of Theorem 3, we need to ensure that the MA under consideration can provide sufficient contexts in order to probe the behaviour of systems. For this purpose, we use the language mCCS.

Theorem 4 (Completeness). *In finitary mCCS, $\Delta \approx_{rbc} \Theta$ implies $\Delta \approx_{bis} \Theta$.* □

5 Conclusion and Related Work

The thesis underlying this paper is that bisimulations should be considered as a proof methodology for demonstrating behavioural equivalence between systems, rather than

[1] Here we use the notation $P \not\Downarrow_a^{>0}$ to mean that $P \Downarrow_a^{\geq p}$ does not hold for any $p > 0$.

providing the definition of the extensional behavioural equivalence itself. We have adapted the well-known *reduction barbed congruence* used for a variety of process calculi [13,19,9], to obtain a touchstone extensional behavioural equivalence for a minor variation of the Markov automata, MAs, originally defined in [8]. Incidently there are also minor variations on the formulation of reduction barbed congruence, often called *contextual equivalence* or *barbed congruence*, in the literature. See [9,21] for a discussion of the differences.

Then we have defined a novel notion of (weak) bisimulations which provide both a sound and complete coinductive proof methodology for establishing the equivalence between such automata. These results were achieved within the context of a rich language, mCCS, for defining MAs. Of particular significance is the presence of insistent actions and a compositional operator which is sensitive to the passage of time; this combination is reminiscent of synchronous CCS [16], although similar compositional operators have already been used for certain varieties of Markov processes [3]. We should point out that our interpretation of mCCS is somewhat simplistic, in that unlike IMC in [10] it does not take into account the multiplicities of action occurrences. However, our interpretation is sufficient for the purposes of this paper. If we were interested in, for example, developing an algebraic theory for mCCS then a more refined interpretation would be required; this could easily be adapted from [10].

There are already quite a few variations on the theme of bisimulations for PAs which can be used to establish behavioural equivalences between MAs [24,18,15,6,11]. A characteristic of our formulation is that it allows bisimulations to relate states to distributions rather than simply states, thus differentiating it from most of these. One exception is [8], where properties of subdistributions are also used in defining their bisimulations. However, our bisimulation $\approx_{bis}$ is different from the bisimulation of [8], denoted by $\approx_{MA}$ here, because the former is defined for full distributions while the latter is for subdistributions. Even if we restrict $\approx_{MA}$ to full distributions, they are still different. For example, we have $A \approx_{bis} \mathbf{0}$ but $A \not\approx_{MA} \mathbf{0}$, where $A \Leftarrow \tau{:}A$. We conjecture that in general $\approx_{bis}$ is strictly coarser than $\approx_{MA}$ (restricted to full distributions), but they coincide for non-divergent systems [7].

Our approach to Markov processes is based directly on that of [8,10], in which external actions are considered instantaneous, and time can only pass when no more internal activity can be performed. Moreover it is only timed actions which are subject to Markovian behaviour. However, there is a large literature on a more general framework in which Markovian behaviour applies to all actions. See [12] or Chapter 3 of [1] for a representative exposition. It would be interesting to see if our notion of bisimulation could be adapted to such a framework.

Acknowledgement. We thank Christian Eisentraut for the interesting discussion on clarifying the relationship between $\approx_{bis}$ and $\approx_{MA}$.

References

1. Aldini, A., Bernardo, M., Corradini, F.: A Process Algebraic Approach to Software Architecture Design. Springer, Heidelberg (2010)
2. Bernardo, M., Gorrieri, R.: A tutorial on empa: A theory of concurrent processes with nondeterminism, priorities, probabilities and time. Theor. Comput. Sci. 202(1-2), 1–54 (1998)

3. Bravetti, M., Bernardo, M.: Compositional asymmetric cooperations for process algebras with probabilities, priorities, and time. ENTCS 39(3) (2000)
4. Deng, Y., Du, W.: Probabilistic barbed congruence. ENTCS 190(3), 185–203 (2007)
5. Deng, Y., van Glabbeek, R., Hennessy, M., Morgan, C.: Testing finitary probabilistic processes (extended abstract). In: Bravetti, M., Zavattaro, G. (eds.) CONCUR 2009. LNCS, vol. 5710, pp. 274–288. Springer, Heidelberg (2009)
6. Desharnais, J., Gupta, V., Jagadeesan, R., Panangaden, P.: Weak bisimulation is sound and complete for pCTL*. Information and Computation 208(2), 203–219 (2010)
7. Eisentraut, C.: Personal communication (2011)
8. Eisentraut, C., Hermanns, H., Zhang, L.: On probabilistic automata in continuous time. In: Proc. LICS 2010, pp. 342–351 (2010)
9. Fournet, C., Gonthier, G.: A hierarchy of equivalences for asynchronous calculi. J. Log. Algebr. Program. 63(1), 131–173 (2005)
10. Hermanns, H.: Interactive Markov Chains: The Quest for Quantified Quality. LNCS, vol. 2428, p. 129. Springer, Heidelberg (2002)
11. Hermanns, H., Parma, A., Segala, R., Wachter, B., Zhang, L.: Probabilistic logical characterization. Inf. Comput. 209, 154–172 (2011)
12. Hillston, J.: A Compositional Approach to Performance Modelling. Distinguished Dissertations in Computer Science. Cambridge University Press, New York (2005)
13. Honda, K., Tokoro, M.: On asynchronous communication semantics. In: Zatarain-Cabada, R., Wang, J. (eds.) ECOOP-WS 1991. LNCS, vol. 612, pp. 21–51. Springer, Heidelberg (1992)
14. Jeffrey, A., Rathke, J.: Contextual equivalence for higher-order pi-calculus revisited. LMCS 1(1:4) (2005)
15. Larsen, K.G., Skou, A.: Bisimulation through probabilistic testing (preliminary report). In: Proc. POPL 1989, pp. 344–352. ACM, New York (1989)
16. Milner, R.: Calculi for synchrony and asynchrony. Theor. Comput. Sci. 25, 267–310 (1983)
17. Milner, R.: Communication and Concurrency. Prentice-Hall, Englewood Cliffs (1989)
18. Philippou, A., Lee, I., Sokolsky, O.: Weak bisimulation for probabilistic systems. In: Palamidessi, C. (ed.) CONCUR 2000. LNCS, vol. 1877, pp. 334–349. Springer, Heidelberg (2000)
19. Rathke, J., Sobocinski, P.: Deriving structural labelled transitions for mobile ambients. In: van Breugel, F., Chechik, M. (eds.) CONCUR 2008. LNCS, vol. 5201, pp. 462–476. Springer, Heidelberg (2008)
20. Sangiorgi, D., Kobayashi, N., Sumii, E.: Environmental bisimulations for higher-order languages. In: Proc. LICS 2007, pp. 293–302. IEEE Computer Society, Los Alamitos (2007)
21. Sangiorgi, D., Walker, D.: The π-calculus: a Theory of Mobile Processes. Cambridge University Press, Cambridge (2001)
22. Segala, R.: Modeling and verification of randomized distributed real-time systems. Technical Report MIT/LCS/TR-676, PhD thesis, MIT, Dept. of EECS (1995)
23. Segala, R.: Testing probabilistic automata. In: Sassone, V., Montanari, U. (eds.) CONCUR 1996. LNCS, vol. 1119, pp. 299–314. Springer, Heidelberg (1996)
24. Segala, R., Lynch, N.A.: Probabilistic simulations for probabilistic processes. Nord. J. Comput. 2(2), 250–273 (1995)