Algorithmic Verification of Recursive Probabilistic State Machines

Kousha Etessami[1] and Mihalis Yannakakis[2]

[1] School of Informatics, University of Edinburgh
[2] Department of Computer Science, Columbia University

Abstract. Recursive Markov Chains (RMCs) ([EY05]) are a natural abstract model of procedural probabilistic programs and related systems involving recursion and probability. They succinctly define a class of denumerable Markov chains that generalize multi-type branching (stochastic) processes. In this paper, we study the problem of model checking an RMC against a given ω-regular specification. Namely, given an RMC A and a Büchi automaton B, we wish to know the probability that an execution of A is accepted by B. We establish a number of strong upper bounds, as well as lower bounds, both for *qualitative* problems (is the probability $= 1$, or $= 0$?), and for *quantitative* problems (is the probability $\geq p$?, or, approximate the probability to within a desired precision). Among these, we show that qualitative model checking for general RMCs can be decided in PSPACE in $|A|$ and EXPTIME in $|B|$, and when A is either a single-exit RMC or when the total number of entries and exits in A is bounded, it can be decided in polynomial time in $|A|$. We then show that quantitative model checking can also be done in PSPACE in $|A|$, and in EXPSPACE in $|B|$. When B is deterministic, all our complexities in $|B|$ come down by one exponential.

For lower bounds, we show that the qualitative model checking problem, even for a fixed RMC, is EXPTIME-complete. On the other hand, even for reachability analysis, we showed in [EY05] that our PSPACE upper bounds in A can not be improved upon without a breakthrough on a well-known open problem in the complexity of numerical computation.

1 Introduction

Recursive Markov Chains (RMCs) are a natural abstract model of procedural probabilistic programs. They succinctly define a natural class of denumerable Markov chains that generalize multi-type branching (stochastic) processes. Informally, an RMC consists of a collection of finite state component Markov chains (MC) that can call each other in a potentially recursive manner. Each component MC has a set of *nodes* (ordinary states), a set of *boxes* (each mapped to a component MC), a well-defined interface consisting of a set of *entry* and *exit* nodes (nodes where it may start and terminate), and a set of probabilistic transitions connecting the nodes and boxes. A transition to a box specifies the entry node and models the invocation of the component MC associated with the

N. Halbwachs and L. Zuck (Eds.): TACAS 2005, LNCS 3440, pp. 253–270, 2005.

box; when (and if) the component MC terminates at an exit, execution of the calling MC resumes from the corresponding exit of the box.

RMCs are a probabilistic version of Recursive State Machines (RSMs). RSMs ([AEY01, BGR01]) and closely related models like Pushdown Systems (PDSs) (see, e.g., [EHRS00, BR00]) have been studied extensively in recent research on model checking and program analysis, because of their applications to verification of sequential programs with procedures. Recursive Markov Chains generalize other well-studied models involving probability and recursion: *Stochastic Context-Free Grammars* (SCFGs) have been extensively studied, mainly in natural language processing (NLP) (see [MS99]). *Multi-Type Branching Processes* (MT-BPs), are an important family of stochastic processes with many applications in a variety of areas (see, e.g., [Har63]). Both SCFG's and MT-BP's are essentially equivalent to *single-exit* RMC's: the special case of RMC's in which all components have one exit. Probabilistic models of programs and systems are of interest for several reasons. First, a program may use randomization, in which case the transition probabilities reflect the random choices of the algorithm. Second, we may want to model and analyse a program or system under statistical conditions on its behaviour (e.g., based on profiling statistics or on statistical assumptions), and to determine the induced probability of properties of interest

We introduced RMCs in [EY05] and developed some of their basic theory, focusing on algorithmic reachability analysis: what is the probability of reaching a given state starting from another? In this paper we study the more general problem of model checking an RMC against an ω-regular specification: given an RMC A and a Büchi automaton B, what is the probability that an execution of A is accepted by B? The techniques we develop in this paper for model checking go far beyond what was developed in [EY05] for reachability analysis.

General RMCs are intimately related to probabilistic Pushdown Systems (pPDSs), and there are efficient translations between RMCs and pPDSs. There has been some recent work on model checking of pPDSs ([EKM04, BKS05]). As we shall describe shortly, our results yield substantial improvements, when translated to the setting of pPDSs, on the best algorithmic upper and lower bounds known for ω-regular model checking of pPDSs.

We now outline the main results in this paper. We are given an RMC A and a property in the form of a (non-deterministic) Büchi automaton (BA) B, whose alphabet corresponds to (labels on) the vertices of A. Let $P_A(L(B))$ denote the probability that an execution of A is accepted by B (i.e., satisfies the property). The *qualitative* model checking problems are: (1) determine whether almost all executions of A satisfy the property B (i.e. is $P_A(L(B)) = 1?$, this corresponds to B being a desirable correctness property), and (2) whether almost no executions of A satisfy B (i.e. is $P_A(L(B)) = 0?$, corresponding to B being an undesirable error property). In the *quantitative* model checking problems we wish to compare $P_A(L(B))$ to a given rational threshold p, i.e., is $P_A(L(B)) \geq p?$, or alternatively, we may wish to approximate $P_A(L(B))$ to within a given number of bits of precision. Note that in general $P_A(L(B))$ may be irrational or may not even be expressible by radicals [EY05]. Hence it cannot be computed exactly.

Qualitative:

	reachability	det. Büchi	nondet. Büchi
1-exit	P	P	P in RMC, EXPTIME in Büchi
Bd	P	P	P in RMC, EXPTIME in Büchi
general	PSPACE	PSPACE	PSPACE in RMC, EXPTIME in Büchi

Quantitative:

	reachability	det. Büchi	nondet. Büchi
1-exit	PSPACE	PSPACE	PSPACE in RMC, EXPSPACE in Büchi
Bd	P	P in RMC for fixed Büchi	P in RMC, for fixed Büchi
general	PSPACE	PSPACE	PSPACE in RMC, EXPSPACE in Büchi

Fig. 1. Complexity of Qualitative and Quantitative problems

We show that the qualitative model checking problems can be solved in PSPACE in $|A|$ and EXPTIME in $|B|$. More specifically, in a first phase the algorithm analyzes the RMC A by itself (using PSPACE only in $|A|$). In a second phase, it further analyses A in conjunction with B, using polynomial time in A and exponential time in B. If the automaton B is deterministic then the time is polynomial in B. Furthermore, if A is a single-exit RMC (which corresponds to SCFG's and MT-BP's), then the first phase, and hence the whole algorithm, can be done in polynomial time in A. Another such case, where we can model-check qualitatively in polynomial time in A, is when the total number of entries and exits in A is bounded (we call them Bd-RMCs). In terms of probabilistic program abstractions, this class of RMC's corresponds to programs with a bounded number of distinct procedures, each of which has a bounded number of input/output parameter values. The internals of the components of the RMCs (i.e., the procedures) can be arbitrarily large and complex.

For quantitative model checking, we show that deciding whether $P_A(L(B)) \geq p$, given a rational $p \in [0, 1]$, can be decided in PSPACE in $|A|$ and EXPSPACE in $|B|$. When B is deterministic, the space is polynomial in both A and B. Moreover, for A a Bd-RMC, and when B is fixed, there is an algorithm that runs in P-time in $|A|$; however, in this case (unlike the others) the exponent of the polynomial depends on B. Table 1 summarizes our complexity upper bounds (the "reachability" columns are from [EY05]; all the other results are new).

For lower bounds, we prove that the qualitative model checking problem, even for a fixed, single entry/exit RMC, is already EXPTIME-complete. On the other hand, even for reachability analysis, we showed in [EY05] that our PSPACE upper bounds in A, even for the quantitative 1-exit problem, and the general qualitative problem, can not be improved without a breakthrough on the complexity of the *square root sum* problem, a well-known open problem in the complexity of exact numerical computation (see Section 2.2).

Related Work. Model checking of flat Markov chains has received extensive attention both in theory and practice (e.g. [CY95, Kwi03, PZ93, Var85]). It is known that model checking of a Markov chain A with respect to a Büchi automaton B is PSPACE-complete, and furthermore the probability $P_A(L(B))$ can

be computed exactly in time polynomial in A and exponential in B. Recursive Markov chains were introduced recently in [EY05], where we developed some of their basic theory and investigated the termination and reachability problems; we summarize the main results in Section 2.2. Recursion introduces a number of new difficulties that are not present in the flat case. For example, in the flat case, the qualitative problems depend only on the structure of the Markov chain (which transitions are present) and not on the precise values of the transition probabilities; this is not any more the case for RMC's and numerical issues have to be dealt with even in the qualitative problem. Furthermore, unlike the flat case, the desired probabilities cannot be computed exactly.

The closely related model of probabilistic Pushdown Systems (pPDS) was introduced and studied recently in [EKM04, BKS05]. They largely focus on model checking against branching-time properties, but they also study deterministic ([EKM04]) and non-deterministic ([BKS05]) Büchi automaton specifications. There are efficient (linear time) translations between RMCs and pPDSs, similar to translations between RSMs and PDSs (see [AEY01, BGR01]). Our upper bounds, translated to pPDSs, improve those obtained in [EKM04, BKS05] by an exponential factor in the general setting, and by more for specific classes like single-exit and Bd-RMCs. Specifically, [BKS05], by extending results in [EKM04], show that qualitative model checking for pPDSs can be done in PSPACE in the size of the pPDS and 2-EXPSPACE in the size of the Büchi automaton, while quantitative model checking can be decided in EXPTIME in the size of the pPDS and in 3-EXPTIME in the size of the Büchi automaton. They do not obtain stronger complexity results for the class of pBPAs (equivalent to single-exit RMCs). Also, the class of Bd-RMCs has no direct analog in pPDSs, as the total number of entries and exits of an RMC gets lost in translation to pPDSs. Reference [EE04] is a survey paper that predates this paper and summarizes only the results in prior papers [EKM04, EY05, BKS05].

The paper is organized as follows. Section 2 gives necessary definitions and background on RMC's from [EY05]. Section 3 shows how to construct from an RMC, A, a flat Markov chain M'_A which in some sense "summarizes" A; this chain plays a central role analogous to the "summary graph" for RSMs [AEY01, BGR01]. Section 4 addresses the qualitative model checking problems, presenting both upper and lower bounds. Section 5 addresses the quantitative model checking problem; a fundamental "unique fixed point theorem" is proved for RMC's, and plays a crucial role in our quantitative algorithms.

Due to space limitations, we have removed almost all proofs from this paper.

2 Definitions and Background

A *Recursive Markov Chain (RMC)*, A, is a tuple $A = (A_1, \ldots, A_k)$, where each *component chain* $A_i = (N_i, B_i, Y_i, En_i, Ex_i, \delta_i)$ consists of:

- A set N_i of *nodes*
- A subset of *entry* nodes $En_i \subseteq N_i$, and a subset of *exit* nodes $Ex_i \subseteq N_i$.
- A set B_i of *boxes*. Let $B = \cup_{i=1}^{k} B_i$ be the (disjoint) union of all boxes of A.

- A mapping $Y_i : B_i \mapsto \{1, \ldots, k\}$ assigns a component to every box. Let $Y = \cup_{i=1}^{k} Y_i$ be $Y : B \mapsto \{1, \ldots, k\}$ where $Y|_{B_i} = Y_i$, for $1 \leq i \leq k$.
- To each box $b \in B_i$, we associate a set of *call ports*, $Call_b = \{(b, en) \mid en \in En_{Y(b)}\}$, and a set of *return ports*, $Return_b = \{(b, ex) \mid ex \in Ex_{Y(b)}\}$.
- A transition relation δ_i, where transitions are of the form $(u, p_{u,v}, v)$ where:
  1. the source u is either a non-exit node $u \in N_i \setminus Ex_i$, or a return port $u = (b, ex) \in Return_b$, where $b \in B_i$.
  2. The destination v is either a non-entry node $v \in N_i \setminus En_i$, or a call port $v = (b, en) \in Call_b$, where $b \in B_i$.
  3. $p_{u,v} \in \mathbb{R}_{>0}$ is the probability of transition from u to v. (We assume $p_{u,v}$ is rational.)
  4. *Consistency of probabilities*: for each u, $\sum_{\{v' | (u, p_{u,v'}, v') \in \delta_i\}} p_{u,v'} = 1$, unless u is a call port or exit node; neither have outgoing transitions, in which case $\sum_{v'} p_{u,v'} = 0$.

We will use the term *vertex* of A_i to refer collectively to its set of nodes, call ports, and return ports, and we denote this set by Q_i, and we let $Q = \bigcup_{i=1}^{k} Q_i$ be the set of all vertices of the RMC A. That is, the transition relation δ_i is a set of probability-weighted directed edges on the set Q_i of vertices of A_i. Let $\delta = \cup_i \delta_i$ be the set of all transitions of A.

An RMC A defines a global denumerable Markov chain $M_A = (V, \Delta)$ as follows. The global *states* $V \subseteq B^* \times Q$ are pairs of the form $\langle \beta, u \rangle$, where $\beta \in B^*$ is a (possibly empty) sequence of boxes and $u \in Q$ is a *vertex* of A. More precisely, the states $V \subseteq B^* \times Q$ and transitions Δ are defined inductively as follows:

1. $\langle \epsilon, u \rangle \in V$, for $u \in Q$. (ϵ denotes the empty string.)
2. if $\langle \beta, u \rangle \in V$ and $(u, p_{u,v}, v) \in \delta$, then $\langle \beta, v \rangle \in V$ and $(\langle \beta, u \rangle, p_{u,v}, \langle \beta, v \rangle) \in \Delta$
3. if $\langle \beta, (b, en) \rangle \in V$ and $(b, en) \in Call_b$, then $\langle \beta b, en \rangle \in V$, & $(\langle \beta, (b, en) \rangle, 1, \langle \beta b, en \rangle) \in \Delta$.
4. if $\langle \beta b, ex \rangle \in V$ and $(b, ex) \in Return_b$, then $\langle \beta, (b, ex) \rangle \in V$ & $(\langle \beta b, ex \rangle, 1, \langle \beta, (b, ex) \rangle) \in \Delta$.

Item 1 corresponds to the possible initial states, 2 corresponds to a transition within a component, 3 is when a new component is entered via a box, 4 is when the process exits a component and control returns to the calling component.

Some states of M_A are *terminating*, i.e., have no outgoing transitions. Namely, states $\langle \epsilon, ex \rangle$, where ex is an exit. We want M_A to be a proper Markov chain, so we consider terminating states as *absorbing*, with a self-loop of probability 1.

A *trace* (or *trajectory*) $t \in V^\omega$ of M_A is an infinite sequence of states $t = s_0 s_1 s_2 \ldots$ such that for all $i \geq 0$, there is a transition $(s_i, p_{s_i, s_{i+1}}, s_{i+1}) \in \Delta$, with $p_{s_i, s_{i+1}} > 0$. Let $\Omega \subseteq V^\omega$ denote the set of traces of M_A. For a state $s = \langle \beta, v \rangle \in V$, let $Q(s) = v$ denote the vertex at state s. Generalizing this to traces, for a trace $t \in \Omega$, let $Q(t) = Q(s_0)Q(s_1)Q(s_2) \ldots \in Q^\omega$. We will consider M_A with *initial states* from $Init = \{\langle \epsilon, v \rangle \mid v \in Q\}$. More generally we may have a probability distribution $p_{\text{init}} : V \mapsto [0, 1]$ on initial states (we usually assume p_{init} has support only in $Init$, and we always assume it has finite support). This induces a probability distribution on traces generated by random walks on M_A. Formally, we have a probability space $(\Omega, \mathcal{F}, \mathbf{Pr}_\Omega)$, parametrized

by p_{init}, where $\mathcal{F} = \sigma(\mathcal{C}) \subseteq 2^{\Omega}$ is the σ-field generated by the set of *basic cylinder sets*, $\mathcal{C} = \{C(x) \subseteq \Omega \mid x \in V^*\}$, where for $x \in V^*$ the cylinder at x is $C(x) = \{t \in \Omega \mid t = xw,\ w \in V^{\omega}\}$. The probability distribution $\mathbf{Pr}_{\Omega} : \mathcal{F} \mapsto [0,1]$ is determined uniquely by the probabilities of cylinder sets, which are:

$$\mathbf{Pr}_{\Omega}(C(s_0 s_1 \dots s_n)) = p_{\text{init}}(s_0) p_{s_0,s_1} p_{s_1,s_2} \cdots p_{s_{n-1},s_n}$$

See, e.g., [Bil95]. RMCs where every component has at most one exit are called *1-exit* RMCs. RMCs where the total number of entries and exits is bounded by a constant c, (i.e., $\sum_{i=1}^{k} |En_i| + |Ex_i| \leq c$) are called *Bounded total entry-exit* RMCs (Bd-RMCs, for short).

2.1 The Central Questions for Model Checking of RMCs

We first define reachability probabilities that play an important role in our analysis. Given a vertex $u \in Q_i$ and an exit $ex \in Ex_i$, both in the same component A_i, let $q^*_{(u,ex)}$ denote the probability of eventually reaching the state $\langle \epsilon, ex \rangle$, starting at the state $\langle \epsilon, u \rangle$. Formally, we have $p_{\text{init}}(\langle \epsilon, u \rangle) = 1$, and $q^*_{(u,ex)} \doteq \mathbf{Pr}_{\Omega}(\{t = s_0 s_1 \dots \in \Omega \mid \exists\, i\ , s_i = \langle \epsilon, ex \rangle\})$. As we shall see, the probabilities $q^*_{(u,ex)}$ will play an important role in obtaining other probabilities.

Recall that a Büchi automaton $B = (\Sigma, S, q_0, R, F)$, has an alphabet Σ, a set of states S, an initial state $q_0 \in S$, a transition relation $R \subseteq S \times \Sigma \times S$, and a set of accepting states $F \subseteq S$. A *run* of B is a sequence $\pi = q_0 v_0 q_1 v_1 q_2 \dots$ of alternating states and letters such that for all $i \geq 0$ $(q_i, v_i, q_{i+1}) \in R$. The ω-word associated with run π is $w_\pi = v_0 v_1 v_2 \dots \in \Sigma^{\omega}$. The run π is *accepting* if for infinitely many i, $q_i \in F$. Define the ω-language $L(B) = \{w_\pi \mid \pi \text{ is an accepting run of } B\}$. Note that $L(B) \subseteq \Sigma^{\omega}$. Let $\mathcal{L} : Q \mapsto \Sigma$, be a given Σ-labelling of the vertices v of RMC A. $\mathcal{L}$ naturally generalizes to $\mathcal{L} : Q^{\omega} \mapsto \Sigma^{\omega}$: for $w = v_0 v_1 v_2 \dots \in Q^{\omega}$, $\mathcal{L}(w) = \mathcal{L}(v_0)\mathcal{L}(v_1)\mathcal{L}(v_2)\dots$. Given RMC A, with initial state $s_0 = \langle \epsilon, u \rangle$, and given a BA B over the alphabet Σ, let $P_A(L(B))$ denote the probability that a trace of M_A is in $L(B)$. More precisely: $P_A(L(B)) \doteq \mathbf{Pr}_{\Omega}(\{t \in \Omega \mid \mathcal{L}(Q(t)) \in L(B)\})$. One needs to show that the sets $\{t \in \Omega \mid \mathcal{L}(Q(t)) \in L(B)\}$ are measurable (in $\mathcal{F}$). This is not difficult (see similar proofs in [CY95, Var85]). The *model checking* problems for ω-regular properties of RMCs are:

(1) *Qualitative* model checking problems: Is $P_A(L(B)) = 1$? Is $P_A(L(B)) = 0$?
(2) *Quantitative* model checking problems: given $p \in [0,1]$, is $P_A(L(B)) \geq p$? Also, we may wish to approximate $P_A(L(B))$ to within a given number of bits of precision.

Note, with a routine for the problem $P_A(L(B)) \geq p$?, we can approximate $P_A(L(B))$ to within i bits using binary search with i calls to the routine. Thus, for quantitative model checking the first problem entails the second. Note that probabilistic reachability is a special case of model checking: given vertex u of RMC A and a subset of vertices F, the probability that the RMC starting at u visits some vertex in F (in some stack context) is equal to $P_A(L(B))$, where we let the labelling $\mathcal{L}$ map vertices in F to 1 and the other vertices to 0, and B is

the 2-state automaton that accepts strings that contain a 1. Similarly, for the *repeated reachability* problem, where we are interested whether a trajectory from u infinitely often visits a vertex of F, we can let B be the (2-state deterministic) automaton that accepts strings with an infinite number of 1's.

To simplify the descriptions of our results, we assume henceforth that $\Sigma = Q$, the vertices of A. This is w.l.o.g. since the problem can be reduced to this case by relabelling the RMC A and modifying the automaton B (see, e.g., [CY95]), however care is needed when measuring complexity separately in RMC, A, and BA, B, since typically B and Σ are small in relation to A. Our complexity results are all with respect to the standard sizes of A and B.

2.2 Basic RMC Theory and Reachability Analysis (From [EY05])

We recall some of the basic theory of RMCs developed in [EY05], where we studied reachability analysis. Considering the probabilities $q^*_{(u,ex)}$ as unknowns, we can set up a system of (non-linear) polynomial equations, such that the probabilities $q^*_{(u,ex)}$ are the *Least Fixed Point* (LFP) solution of this system. Use a variable $x_{(u,ex)}$ for each unknown probability $q^*_{(u,ex)}$. We will often find it convenient to index the variables $x_{(u,ex)}$ according to a fixed order, so we can refer to them also as $x_1, \ldots, x_n$, with each $x_{(u,ex)}$ identified with x_j for some j. We thus have a vector of variables: $\mathbf{x} = (x_1\ x_2 \ldots x_n)^T$.

Definition 1. *Given RMC $A = (A_1, \ldots, A_k)$, define the system of polynomial equations, S_A, over the variables $x_{(u,ex)}$, where $u \in Q_i$ and $ex \in Ex_i$, for $1 \leq i \leq k$. The system contains one equation $x_{(u,ex)} = P_{(u,ex)}(\mathbf{x})$, for each variable $x_{(u,ex)}$. $P_{(u,ex)}(\mathbf{x})$ denotes a multivariate polynomial with positive rational coefficients. There are 3 cases, based on the "type" of vertex u:*

1. *Type I: $u = ex$. In this case: $x_{(ex,ex)} = 1$.*
2. *Type II: either $u \in N_i \setminus \{ex\}$ or $u = (b, ex')$ is a return port. In these cases:*
$$x_{(u,ex)} = \sum_{\{v \mid (u, p_{u,v}, v) \in \delta\}} p_{u,v} \cdot x_{(v,ex)}.$$
3. *Type III: $u = (b, en)$ is a call port. In this case:*
$$x_{((b,en),ex)} = \sum_{ex' \in Ex_{Y(b)}} x_{(en,ex')} \cdot x_{((b,ex'),ex)}$$

In vector notation, we denote $S_A = (x_j = P_j(\mathbf{x}) \mid j = 1, \ldots, n)$ by: $\mathbf{x} = P(\mathbf{x})$.

Given A, we can construct $\mathbf{x} = P(\mathbf{x})$ in P-time: $P(\mathbf{x})$ has size $O(|A|\theta^2)$, where θ denotes the maximum number of exits of any component. For vectors $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, define $\mathbf{x} \preceq \mathbf{y}$ to mean that $x_j \leq y_j$ for every coordinate j. For $D \subseteq \mathbb{R}^n$, call a mapping $H : \mathbb{R}^n \mapsto \mathbb{R}^n$ *monotone* on D, if: for all $\mathbf{x}, \mathbf{y} \in D$, if $\mathbf{x} \preceq \mathbf{y}$ then $H(\mathbf{x}) \preceq H(\mathbf{y})$. Define $P^1(\mathbf{x}) = P(\mathbf{x})$, and $P^k(\mathbf{x}) = P(P^{k-1}(\mathbf{x}))$, for $k > 1$. Let $\mathbf{q}^* \in \mathbb{R}^n$ denote the n-vector of probabilities $q^*_{(u,ex)}$, using the same indexing as used for $\mathbf{x}$. Let $\mathbf{0}$ denote the all 0 n-vector. Define $\mathbf{x}^0 = \mathbf{0}$, and $\mathbf{x}^k = P(\mathbf{x}^{k-1}) = P^k(\mathbf{0})$, for $k \geq 1$. The map $P : \mathbb{R}^n \mapsto \mathbb{R}^n$ is monotone on $\mathbb{R}^n_{\geq 0}$.

Theorem 1. *([EY05], see also [EKM04]) $\mathbf{q}^* \in [0,1]^n$ is the Least Fixed Point solution,* LFP(P)*, of $\mathbf{x} = P(\mathbf{x})$. Thus, $\mathbf{q}^* = P(\mathbf{q}^*)$ and $\mathbf{q}^* = \lim_{k \to \infty} \mathbf{x}^k$, and for all $k \geq 0$, $\mathbf{x}^k \preceq \mathbf{x}^{k+1} \preceq \mathbf{q}^*$, and for all $\mathbf{q}' \in \mathbb{R}^n_{\geq 0}$, if $\mathbf{q}' = P(\mathbf{q}')$, then $\mathbf{q}^* \preceq \mathbf{q}'$.*

There are already 1-exit RMCs for which the probability $q^*_{(en,ex)}$ is irrational and not "solvable by radicals" ([EY05]). Thus, we can't compute probabilities exactly. Given a system $x = P(x)$, and a vector $q \in [0,1]^n$, consider the following sentence in the *Existential Theory of Reals* (which we denote by $\boldsymbol{ExTh}(\mathbb{R})$):

$$\varphi \equiv \exists x_1, \ldots, x_m \bigwedge_{i=1}^{m} P_i(x_1, \ldots, x_m) = x_i \wedge \bigwedge_{i=1}^{m} 0 \leq x_i \wedge \bigwedge_{i=1}^{m} x_i \leq q_i$$

φ is true precisely when there is some $z \in \mathbb{R}^m$, $0 \preceq z \preceq q$, and $z = P(z)$. Thus, if we can decide the truth of this sentence, we could tell whether $q^*_{(u,ex)} \leq p$, for some rational p, by using the vector $q = (1, \ldots, p, 1, \ldots)$. We will rely on decision procedures for $\boldsymbol{ExTh}(\mathbb{R})$. It is known that $\boldsymbol{ExTh}(\mathbb{R})$ can be decided in PSPACE and in exponential time, where the time exponent depends (linearly) only on the number of variables; thus for a fixed number of variables the algorithm runs in polynomial time [Can88, Ren92, BPR96]. As a consequence:

Theorem 2. *([EY05]) Given RMC A and rational ρ, there is a PSPACE algorithm to decide whether $q^*_{(u,ex)} \leq \rho$, with running time $O(|A|^{O(1)} \cdot 2^{O(m)})$ where m is the number of variables in the system $x = P(x)$ for A. Moreover $q^*_{(u,ex)}$ can be approximated to within j bits of precision within PSPACE and with running time at most j times the above.*

For Bd-RMCs, as shown in [EY05] it is possible to construct efficiently a system of equations in a *bounded* number of variables, whose LFP yields the entry-exit probabilities $q^*_{(en,ex)}$. Since $\boldsymbol{ExTh}(\mathbb{R})$ is decidable in P-time when the number of variables is bounded, this yields:

Theorem 3. *([EY05]) Given Bd-RMC, A & rational $p \in [0,1]$, there is a P-time algorithm to decide whether, for a vertex u & exit ex, $q^*_{(u,ex)} \geq p (or < p)$.*

For 1-exit RMCs (SCFGs), the qualitative termination/reachability problem can be solved efficiently, via an eigenvalue characterization and other techniques.

Theorem 4. *([EY05]) There is a P-time algorithm that for a 1-exit RMC, vertex u and exit ex, decides which of the following holds:(1) $q^*_{(u,ex)} = 0$,(2) $q^*_{(u,ex)} = 1$,or (3) $0 < q^*_{(u,ex)} < 1$.*

Hardness, such as NP-hardness, is not known for RMC reachability. However, in [EY05] we gave strong evidence of "difficulty": the square-root sum problem is P-time reducible to deciding whether $q^*_{(u,ex)} \geq p$, in a 1-exit RMC, and to deciding whether $q^*_{(u,ex)} = 1$ for a 2-exit RMC. *Square-root sum* is the following decision problem: given $(d_1, \ldots, d_n) \in \mathbb{N}^n$ and $k \in \mathbb{N}$, decide whether $\sum_{i=1}^{n} \sqrt{d_i} \leq k$. It is solvable in PSPACE, but it has been a major open problem since the 1970's (see, e.g., [GGJ76, Tiw92]) whether it is solvable even in NP.

As a practical efficient numerical algorithm for computing the probabilities $q^*_{(u,ex)}$, it was proved in [EY05] that a multi-dimensional Newton's method converges monotonically to the LFP of $\mathbf{x} = P(\mathbf{x})$, and constitutes a rapid acceleration of iterating $P^k(\mathbf{0})$, $k \to \infty$.

3 The Conditioned Summary Chain M'_A

For an RMC A, suppose we somehow have the probabilities $q^*_{(u,ex)}$ "in hand". Based on these, we construct a *conditioned summary chain*, M'_A, a finite Markov chain that will allow us to answer repeated reachability questions. Extensions of M'_A will later be a key to model checking RMCs. Since probabilities $q^*_{(u,ex)}$ are potentially irrational, we can not compute M'_A exactly. However, M'_A will be important in our correctness arguments, and we will in fact be able to compute the "structure" of M'_A, i.e., what transitions have non-zero probability. The structure of M'_A will be sufficient for answering various "qualitative" questions.

We will assume, w.l.o.g., that each RMC has one initial state $s_0 = \langle \epsilon, en_{\text{init}} \rangle$, with en_{init} the only entry of a component that does not contain any exits. Any RMC can readily be converted to an "equivalent" one in this form.

Before describing M'_A, let us recall from [AEY01], the construction of a "summary graph", $H_A = (Q, E_{H_A})$, which ignores probabilities and is based only on information about reachability in the underlying RSM of A. Let R be the binary relation between entries and exits of components such that $(en, ex) \in R$ precisely when there exists a path from $\langle \epsilon, en \rangle$ to $\langle \epsilon, ex \rangle$, in the underlying graph of M_A. The edge set E_{H_A} is defined as follows. For $u, v \in Q$, $(u, v) \in E_{H_A}$ iff one of the following holds:

1. u is not a call port, and $(u, p_{u,v}, v) \in \delta$, for $p_{u,v} > 0$.
2. $u = (b, en)$ is a call port, and $(en, ex) \in R$, and $v = (b, ex)$ is a return port.
3. $u = (b, en)$ is a call port, and $v = en$ is the corresponding entry.

For each vertex $v \in Q_i$, let us define the probability of *never exiting*: $\text{ne}(v) = 1 - \sum_{ex \in Ex_i} q^*_{(v,ex)}$. Call a vertex v *deficient* if $\text{ne}(v) > 0$, i.e. there is a nonzero probability that if the RMC starts at v it will never terminate (reach an exit of the component).

We define $M'_A = (Q_{M'_A}, \delta_{M'_A})$ as follows. The set of states $Q_{M'_A}$ of M'_A is the set of deficient vertices: $Q_{M'_A} = \{v \in Q \mid \text{ne}(v) > 0\}$. For $u, v \in Q_{M'_A}$, there is a transition $(u, p'_{u,v}, v)$ in $\delta_{M'_A}$ if and only if one of the following conditions holds:

1. $u, v \in Q_i$ and $(u, p_{u,v}, v) \in \delta_i$, and $p'_{u,v} = \frac{p_{u,v} \cdot \text{ne}(v)}{\text{ne}(u)}$.
2. $u = (b, en) \in \mathit{Call}_b$, $v = (b, ex) \in \mathit{Return}_b$, $q^*_{(en,ex)} > 0$, & $p'_{u,v} = \frac{q^*_{(en,ex)} \text{ne}(v)}{\text{ne}(u)}$.
3. $u = (b, en) \in \mathit{Call}_b$ and $v = en$, and $p'_{u,v} = \frac{\text{ne}(v)}{\text{ne}(u)}$. We call these transitions, from a call port to corresponding entry, special *red* transitions.

Note that in all three cases, $p'_{u,v}$ is well-defined (the denominator is nonzero) and it is positive. Recall that we assumed that the initial vertex en_{init} is the entry of a component A_0, and A_0 has no exits. Thus for all $v \in Q_0$, $\text{ne}(u) = 1$, and thus $Q_0 \subseteq Q_{M'_A}$, and if $(u, p_{u,v}, v) \in \delta_0$, then $(u, p_{u,v}, v) \in \delta_{M'_A}$.

Proposition 1. *Probabilities on transitions out of each state in $Q_{M'_A}$ sum to 1.*

M'_A is an ordinary (flat) Markov chain. Let $(\Omega', \mathcal{F}', \mathbf{Pr}_{\Omega'})$ denote the probability space on traces of M'_A. We now define a mapping $\rho : \Omega \mapsto \Omega' \cup \{\star\}$, that maps

every trace t of the original (infinite) Markov chain M_A, either to a unique trajectory $\rho(t) \in \Omega'$ of the MC M'_A, or to the special symbol $\star$. Trajectories mapped to $\star$ will be precisely those that go through missing vertices $u \in Q$ that are not in $Q_{M'_A}$, i.e., with $\mathrm{ne}(u) = 0$. We show the total probability of all these trajectories is 0, i.e., $\mathbf{Pr}_\Omega(\rho^{-1}(\star)) = 0$, and moreover, M'_A preserves the probability measure of M_A: for all $D \in \mathcal{F}'$, $\rho^{-1}(D) \in \mathcal{F}$, and $\mathbf{Pr}_{\Omega'}(D) = \mathbf{Pr}_\Omega(\rho^{-1}(D))$. We define ρ in two phases. We first define a map $\rho^H : \Omega \mapsto Q^\omega$, where every trajectory $t \in \Omega$ is mapped to an infinite path $\rho^H(t)$ in the summary graph H_A. Thereafter, we let $\rho(t) = \rho^H(t)$ if all vertices of $\rho^H(t)$ are in M'_A, and let $\rho(t) = \star$ otherwise. We define ρ^H for a trace $t = s_0 s_1 \dots s_i \dots$, sequentially based on prefixes of t, as follows. By assumption, $s_0 = \langle \epsilon, en_{\mathrm{init}} \rangle$. ρ^H maps s_0 to en_{init}. Suppose $s_i = \langle \beta, u \rangle$, and, inductively, suppose that ρ^H maps $s_0 \dots s_i$ to $e_{\mathrm{init}} \dots u$. First, suppose u is not a call port, and that $s_{i+1} = \langle \beta, v \rangle$, then $s_0 \dots s_i s_{i+1}$ maps to $e_{\mathrm{init}} \dots uv$. Next, suppose $u = (b, en)$ is a call port and $s_{i+1} = \langle \beta b, en \rangle$. If the trace eventually returns from this call (i.e., there exists $j > i+1$, such that $s_j = \langle \beta b, ex \rangle$ and $s_{j+1} = \langle \beta, (b, ex) \rangle$, and such that each of the states $s_{i+1} \dots s_j$, have βb as a prefix of the call stack), then $s_0 \dots s_j$ is mapped by ρ^H to $en_{\mathrm{init}} \dots u(b, ex)$. If the trace never returns from this call, then $s_0 \dots s_i s_{i+1}$ maps to $en_{\mathrm{init}} \dots u\ en$. This concludes the definition of ρ^H. We show that the mapping ρ is measure preserving.

Lemma 1. $\mathbf{Pr}_\Omega(\rho^{-1}(\star)) = 0$. *Moreover, for all* $D \in \mathcal{F}'$, $\rho^{-1}(D) \in \mathcal{F}$ *and* $\mathbf{Pr}_\Omega(\rho^{-1}(D)) = \mathbf{Pr}_{\Omega'}(D)$.

Let $H'_A = (Q_{H'_A}, E_{H'_A})$ be the underlying directed graph of M'_A. In other words, $Q_{H'_A} = Q_{M'_A}$, and $(u, v) \in E_{H'_A}$ iff $(u, p'_{u,v}, u) \in \delta_{M'_A}$. We show we can compute H'_A in P-time for 1-exit RMCs and Bd-RMCs, and in PSPACE for arbitrary RMCs. The basic observation is: the structure of M'_A depends only on qualitative facts about the probabilities $q^*_{(en,ex)}$ and $\mathrm{ne}(u)$, for $u \in Q$.

Proposition 2. *For a RMC A (respectively, 1-exit or Bd-RMC), and* $u \in Q$, *we can decide whether* $\mathrm{ne}(u) > 0$ *in PSPACE (respectively, P-time).*

Proof. Suppose u is in a component A_i where $Ex_i = \{ex_1, \dots, ex_k\}$. Clearly, $\mathrm{ne}(u) > 0$ iff $\sum_{j=1}^k q^*_{(u,ex_j)} < 1$. Consider the following sentence, φ, in $\boldsymbol{ExTh}(\mathbb{R})$.

$$\varphi \equiv \exists x_1, \dots, x_n \bigwedge_{i=1}^n P_i(x_1, \dots, x_n) = x_i \wedge \bigwedge_{i=1}^n 0 \leq x_i \wedge \sum_{j=1}^k x_{(u,ex_j)} < 1$$

Since $\mathbf{q}^*$ is the LFP solution of $\mathbf{x} = P(\mathbf{x})$, φ is true in the reals if and only if $\sum_{j=1}^k q^*_{(u,ex_j)} < 1$. This query can be answered in PSPACE. In the special case of a 1-exit RMC, we have $Ex_i = \{ex_1\}$, and $\mathrm{ne}(u) > 0$ iff $q^*_{(u,ex_1)} < 1$. As mentioned in section 2.2, this can be answered in P-time for 1-exit RMCs ([EY05]). Similarly, for Bd-RMCs the question can be answered in P-time by the techniques developed in [EY05]. □

Corollary 1. *For a RMC A (respectively, 1-exit or Bd-RMC), we can compute H'_A in PSPACE (respectively, in polynomial time).*

Proof. Recall that $u \in Q_{H'_A}$ precisely when $u \in Q$ and $\mathrm{ne}(u) > 0$. Thus we can determine the set of nodes with the said complexities, respectively. The transitions of type 1 and 3 in the definition of M'_A are immediately determined. For the type 2 transitions, where $u = (b, en)$ and $v = (b, ex)$, in order to determine whether to include the corresponding summary edge (u, v) we need to decide whether $q^*_{(en,ex)} > 0$. This can be done in polynomial time by invoking the reachability algorithm for RSM's [AEY01, BGR01]. □

4 Qualitative Model Checking

Upper Bounds. Given an RMC $A = (A_1, \ldots, A_k)$ and a (nondeterministic) Büchi automaton $B = (\Sigma, S, q_0, R, F)$ whose alphabet Σ is the vertex set of A, we wish to determine whether $P_A(L(B)) = 1$, $= 0$, or is in-between. We will construct a finite Markov chain $M'_{A,B}$ such that $P_A(L(B))$ is equal to the probability that a trajectory of $M'_{A,B}$ starting from a given initial state reaches one of a designated set of "accepting" bottom SCCs.

First, let $B' = (\Sigma, 2^S, \{q_0\}, R', F')$ be the deterministic automaton obtained by the usual subset construction on B. In other words, states of B' are subsets $T \subseteq S$, and the transition function $R' : (2^S \times \Sigma) \mapsto 2^S$ is given by: $R'(T_1, v) = \{q' \in S \mid \exists q \in T_1 \; s.t. \; (q, v, q') \in R\}$. (We make no claim that $L(B) = L(B')$.)

Next we define the standard *product* RMC, $A \otimes B'$, of the RMC A, and the deterministic Büchi automaton B'. $A \otimes B'$ has the same number of components as A. Call these $A'_1, \ldots, A'_k$. The vertices in component A'_i are pairs (u, T), where $u \in Q_i$ and $T \in 2^S$, and (u, T) is an entry (exit) iff u is an entry (exit). The transitions of A'_i are as follows: there is a transition $((u, T), p_{u,v}, (v, R'(T, v)))$ in A'_i iff there is a transition $(u, p_{u,v}, v)$ in A_i.

Define $M'_{A,B}$ as $M'_{A,B} = M'_{A \otimes B'}$. Thus $M'_{A,B}$ is the conditioned summary chain of RMC $A \otimes B'$. For qualitative analysis on $M'_{A,B}$, we need the underlying graph $H'_{A,B}$. Importantly for the complexity of our algorithms, we do not have to explicitly construct $A \otimes B'$ to obtain $H'_{A,B}$. Observe that states of $M'_{A,B} = (Q \times 2^S, \delta_{M'_{A,B}})$ are pairs (v, T) where v is a state of M'_A, and T a state of B'. The initial state of $M'_{A,B}$ is $(v_0, \{q_0\})$, where v_0 is the initial state of M'_A and q_0 of B. The transitions of $M'_{A,B}$ from a state (v, T) are as follows:

- Case 1: v is not a call port. Then for every transition $(v, p'_{v,v'}, v') \in \delta_{M'_A}$, we have a corresponding transition $((v, T), p'_{v,v'}, (v', R'(T, v'))) \in \delta_{M'_{A,B}}$.
- Case 2: v is a call port, $v = (b, en)$ where v is vertex in component A_i and box b is mapped to component A_j. If there is a *red* transition $(v, p_{v,en}, en) \in \delta_{M'_A}$ then there is a *red* transition $((v, T), p_{v,en}, (en, R'(T, en)) \in \delta_{M'_{A,B}}$ with the same probability.
- Case 3: If v has a summary transition $(v, p_{v,v'}, v')$ in M'_A, where $v' = (b, ex)$, then we have summary transitions of the form $((v, T), p'', (v', T'))$ in $M'_{A,B}$

to states of the form (v', T') iff there exists a path in M_A from $\langle \epsilon, en \rangle$ to $\langle \epsilon, ex \rangle$ which, viewed as a string, drives B' from T to T'; the probability p'' of the transition is $p'' = p' \cdot \mathrm{ne}(v')/\mathrm{ne}(v)$ where p' is the probability of all such v-v' paths that drive B' from T to T'.

$M'_{A,B}$ is a well-defined Markov chain, which is a refinement of M'_A. That is, every trajectory of $M'_{A,B}$ projected on the first component is a trajectory of M'_A and the projection preserves probabilities. We can define a mapping σ from the trajectories t of the original (infinite) Markov chain M_A to the trajectories of $M'_{A,B}$, or the special symbol $\star$, in a similar manner as we defined the mapping ρ from trajectories of M to M'_A. For a trajectory t of M_A, it is easy to see that if $\rho(t) \neq \star$ then also $\sigma(t) \neq \star$. Thus, with probability 1 a trajectory of M_A is mapped to one of $M'_{A,B}$. Furthermore, we can show along similar lines the analogue of Lemma 2, i.e. the mapping σ preserves probabilities.

Consider a product graph (without probabilities) $M'_A \otimes B$ between the Markov chain M'_A and the given nondeterministic BA B (not B') as follows: $M'_A \otimes B$ has nodes (v, q), for all vertices v of M'_A and states q of B, and an edge $(v, q) \to (v', q')$ if either (i) $v \to v'$ is an ordinary edge or a red edge of M'_A and q has a transition to q' on input v', or (ii) $v \to v'$ is a summary edge and the RMC has a path from v to v' that corresponds to a run of B from q to q'; if any such run goes through an accepting state then we mark the edge $(v, q) \to (v', q')$ as an *accepting* edge. Also, call a node (v, q) *accepting* if $q \in F$ is an accepting state of B.

With every transition (edge) of $M'_{A,B}$ and every edge of $M'_A \otimes B$ we associate a string γ over Σ (the vertex set of A) that caused the edge to be included; i.e., if edge $(v, T) \to (v', T')$ of $M'_{A,B}$ (respectively, edge $(v, q) \to (v', q')$ of $M'_A \otimes B$) corresponds to an ordinary or red edge of M'_A then $\gamma = v'$. If it corresponds to a summary edge then we let γ be any string that corresponds to a $v - v'$ path that drives B' from T to T' (resp., for which B has a path from q to q'; if the edge $(v, q) \to (v', q')$ is marked as accepting then we pick a path that goes through an accepting state of B). In the case of a summary edge, there may be many strings γ as above; we just pick anyone of them.

Let t be any trajectory of M_A starting from $\langle \epsilon, v \rangle$, for some vertex v of M'_A and let r be a corresponding run of B starting from a state q. With probability 1, t maps to a trajectory $t' = \rho(t)$ of M'_A. The mapping ρ can be extended to pairs (t, r), where r is a run of B on t, i.e., the pair (t, r) is mapped to a run $r' = \rho(t, r)$ of $M'_A \otimes B$. If r is an accepting run of B then r' goes infinitely often through an accepting node or an accepting edge. The converse does not hold necessarily: a non-accepting run r of B corresponding to a trajectory t may be mapped to a run r' of $M'_A \otimes B$ that traverses infinitely often an accepting edge.

If B is a deterministic BA, then $M'_{A,B}$ and $M'_A \otimes B$ are clearly the same (except that in $M'_A \otimes B$ we did not include the probabilities of the edges). In this case, the analysis is simpler. Let us say that a bottom strongly connected component (SCC) of $M'_{A,B}$ (and $M'_A \otimes B$) is *accepting* iff it contains an accepting node or an accepting edge.

Theorem 5. *For a RMC A and a deterministic BA B, the probability $P_A(L(B))$ that a trajectory of A is accepted by B is equal to the probability that a trajectory of $M'_{A,B}$ starting from the initial node (v_0, q_0) reaches an accepting bottom SCC.*

Suppose now that B is nondeterministic. We will follow the approach of [CY95] for flat Markov chains, except that here we have to deal with recursive calls and with the summary edges of the constructed Markov chain $M'_{A,B}$ which correspond to sets of paths in the original chain M_A rather than single steps. This complicates things considerably.

Let v be a vertex of M'_A and $q \in F$ an accepting state of B. Let $D(v, q)$ be the subgraph of $M'_{A,B}$ induced by the node $(v, \{q\})$ and all nodes reachable from it . We say that the pair (v, q) is *special of type 1* if some bottom SCC C of $D(v, q)$ contains a state (v, T) with $q \in T$. We associate with such a pair (v, q) a string $\gamma(v, q) \in \Sigma^*$ that is the concatenation of the strings associated with the edges of $D(v, q)$ on a path from $(v, \{q\})$ to a node of C. (There may be many such paths; just pick anyone.) Let $v = (b, en)$ be a vertex of M'_A that is a call port of a box b of A and let $q \notin F$ be a non-accepting state of B. Define a graph $D(v, q)$ as follows. The graph contains a root node vq and a subgraph of $M'_{A,B}$ consisting of the nodes reachable from vq after we add the following edges. We add an edge from vq to a node $(v', \{q'\})$ of $M'_{A,B}$, where $v' = (b, ex)$ is a return port of the same box b as v, iff there is a path γ from $\langle \epsilon, en \rangle$ to $\langle \epsilon, ex \rangle$ such that B has a run from q to q' on γ that goes through an accepting state; we label the edge $vq \to (v', \{q'\})$ with such a string γ. The graph $D(v, q)$ consists of the root vq and the subgraph of $M'_{A,B}$ induced by all the nodes that are reachable from vq after adding the above edges. We call the pair (v, q) *special of type 2* if some bottom SCC C of $D(v, q)$ contains a state (v, T) with $q \in T$. As in the previous case, we associate with the pair (v, q) a string $\gamma(v, q) \in \Sigma^*$ that is the concatenation of the strings associated with the edges of $D(v, q)$ on a path from vq to a node of C. Special pairs have the following important properties.

Lemma 2. *Suppose (v, q) is special and that RMC A starts at $\langle \epsilon, v \rangle$ and first performs the transitions in $\gamma(v, q)$. Then with probability 1 such a trajectory t of the RMC is accepted by B with initial state q. Specifically, there is a corresponding accepting run r of B such that $\rho(t, r)$ is a run of $M'_A \otimes B$ starting from (v, q) that infinitely repeats node (v, q) if (v, q) is special of type 1, or repeats an accepting edge out of (v, q) if (v, q) is special of type 2.*

Lemma 3. *Suppose there is non-zero probability that a trajectory of the RMC A starting at any vertex $u \in M'_A$ has a corresponding run in $M'_A \otimes B$ starting from any node (u, p) which repeats an accepting state (v, q) infinitely often or repeats an accepting edge $(v, q) \to (v', q')$ infinitely often. Then (v, q) is special.*

Proposition 3. *$P_A(L(B)) > 0$ iff from (v_0, q_0) in $M'_A \otimes B$ we can reach a special (v, q).*

Call a bottom SCC of the flat Markov chain $M'_{A,B}$ *accepting* if it contains a state (v, T), with some $q \in T$ such that (v, q) is special; otherwise call it *rejecting*.

Theorem 6. *$P_A(L(B))$ is equal to the probability that a trajectory of $M'_{A,B}$ starting from the initial state $(v_0, \{q_0\})$ reaches an accepting bottom SCC.*

Thus, $P_A(L(B)) = 1$ iff all bottom SCCs of $M'_{A,B}$ reachable from $(v_0, \{q_0\})$ are accepting, and $P_A(L(B)) = 0$ iff no reachable bottom SCC is accepting (i.e., by Prop. 3, there is no path in $M'_A \otimes B$ from $(v_0, \{q_0\})$ to a special node (v, q)).

As with M'_A and H'_A, let $H'_{A,B}$ denote the underlying directed graph of $M'_{A,B}$. For the qualitative problem, we only need (1) to construct $H'_{A,B}$ and thus only need to know which nodes and edges are present, and (2) to determine which pairs (v, q) are special, and hence which bottom SCCs are accepting. Thus we first have to identify the vertices u of the RMC A for which $\mathrm{ne}(u) > 0$, which can be done in PSPACE for general RMCs and P-time for 1-exit RMCs and for Bd-RMCs. Then, the edges of $H'_{A,B}$ can be determined by the standard reachability algorithm for RSMs ([AEY01]). This works by first constructing the genuine product of the underlying RSM of A (ignoring probabilities on transitions) together with the Büchi automaton B'. This defines a new RSM $A \otimes B'$ (no probabilities), whose size is polynomial in A and B', and thus is exponential in the original non-deterministic Büchi automaton B. The time required for reachability analysis for RSMs is polynomial ([AEY01]). Thus, once we have identified the deficient vertices of the RMC, the rest of the construction of $H'_{A,B}$ takes time polynomial in A and B'.

To determine which pairs (v, q) are special, we construct for each candidate (v, q) the graph $D(v, q)$. For (v, q) with $q \in F$, this is immediate from $H'_{A,B}$. For (v, q) with $q \notin F$ and $v = (b, en)$ a call port of a box b, we test for each return port $v' = (b, ex)$ of the box and each state q' of B whether there should be an edge $vq \to (v', \{q'\})$; this involves a call to the RSM algorithm of [AEY01] to determine whether there is a path in the RSM $A \otimes B$ from (en, q) to (ex, q') (with empty stack) that goes through a vertex whose second component is an accepting state of B. Once we determine these edges, we can construct $D(v, q)$. This takes time polynomial in A and B'. Then compute the SCCs of $D(v, q)$, examine the bottom SCCs and check if one of them contains (v, T) with $q \in T$.

Finally, once we have identified the special pairs, we examine the reachable bottom SCCs of $H'_{A,B}$ and determine which ones are accepting and which are rejecting. The dependence of the time complexity on the size of the given RMC A is polynomial except for the identification of the vertices u for which $\mathrm{ne}(u) > 0$. The dependence on $|B|$ is exponential because of the subset construction. If B is deterministic to begin with, we avoid the exponential blow-up and thus have polynomial complexity in B. Thus we have:

Theorem 7. *Given RMC A & Büchi automaton B, we can decide whether $P_A(L(B)) = 0$, $P_A(L(B)) = 1$, or $0 < P_A(L(B)) < 1$ in PSPACE in A, and EXPTIME in B. For a 1-exit RMC or Bd-RMC, the time complexity is polynomial in $|A|$. And, if B is deterministic, the time complexity in $|B|$ is also polynomial.*

Lower Bounds. We show conversely that the exponential time complexity of qualitative model checking for a nondeterministic BA is in general unavoidable.

Theorem 8. *Deciding whether a given RMC A satisfies a property specified by a Büchi automaton B with probability $= 1$, (i.e., whether $P_A(L(B)) = 1$)) is EXPTIME-complete. Furthermore, this holds even if the RMC is fixed and each component has 1 entry and 1 exit. Moreover, the qualitative "emptiness" problem, namely deciding whether $P_A(L(B)) = 0$, is also EXPTIME-complete, again even when the RMC is fixed and each component has 1 entry and 1 exit.*

5 Quantitative Model Checking

As mentioned, the transition probabilities of the chain $M'_{A,B}$ cannot be computed exactly, but instead have to be determined implicitly. To do quantitative model checking in PSPACE in $|A|$, it will be crucial to use $\boldsymbol{ExTh}(\mathbb{R})$ to uniquely identify $\mathrm{LFP}(P)$ for the systems $x = P(x)$. The following key theorem enables this.

Theorem 9. *(Unique Fixed Point Theorem) The equations $x = P(x)$ have a unique solution q^* that satisfies $\sum_{ex} q^*_{(u,ex)} < 1$ for every deficient vertex u, and $\sum_{ex} q^*_{(u,ex)} \leq 1$ for every other vertex u. (Of course, $q^* = \mathrm{LFP}(P)$.)*

Theorem 10. *Given RMC, A, and BA, B, and a rational value $p \in [0,1]$, we can decide whether $P_A(L(B)) \geq p$ in PSPACE in $|A|$ and in EXPSPACE in B, specifically in space $O(|A|^{c_1} 2^{c_2|B|})$ for some constants c_1, c_2. Furthermore, if B is deterministic we can decide this in PSPACE in both A and B.*

Proof. We make crucial use of Theorem 9, and we combine this with use of the summary chain $M'_{A,B}$, and queries to $\boldsymbol{ExTh}(\mathbb{R})$. Observe that by Theorem 6, all we need to do is "compute" the probability that a trajectory of $M'_{A,B}$, starting from the initial state $(v_0, \{q_0\})$ reaches an accepting bottom SCC. We can not compute $M'_{A,B}$ exactly, however, we will be able to identify the transition probabilities uniquely inside a $\boldsymbol{ExTh}(\mathbb{R})$ query, and will, inside the same query identify the probability of reaching an accepting bottom SCC.

Let $\mathbf{q}^* = \mathrm{LFP}(P)$ be the solution vector of probabilities for the system $\mathbf{x} = P(\mathbf{x})$ associated with RMC A. Recall that by Proposition 2, we can compute in PSPACE in $|A|$ the set $Q' = \{u \in Q \mid \mathrm{ne}(u) > 0\}$ of deficient vertices. We do this as a first step. Consider next the following quantifier-free formula, where $c(u)$ is the index of the component of a vertex u:

$$\varphi_1(\mathbf{x}) \equiv \mathbf{x} = P(\mathbf{x}) \wedge 0 \preceq \mathbf{x} \wedge \bigwedge_{u \in Q'} \sum_{ex \in Ex_{c(u)}} x_{(u,ex)} < 1 \wedge \bigwedge_{u \in Q \setminus Q'} \sum_{ex \in Ex_{c(u)}} x_{(u,ex)} = 1$$

By Theorem 9, the only vector $\mathbf{x}$ in $\mathbb{R}^n$ for which $\varphi_1(\mathbf{x})$ holds true is $\mathbf{q}^*$. In other words, φ_1 uniquely identifies $\mathrm{LFP}(P)$. Recall that $\mathrm{ne}(u) = 1 - \sum_{ex \in Ex_{c(u)}} q^*_{(u,ex)}$. Now, let $\mathbf{y}$ be a vector of variables indexed by vertices of A, and let $\varphi_2(\mathbf{x}, \mathbf{y}) \equiv \bigwedge_{u \in Q} y_u = 1 - \sum_{ex \in Ex_{c(u)}} x_{(u,ex)}$. The only vector of reals $(\mathbf{x}, \mathbf{y})$ that satisfies $\varphi_1 \wedge \varphi_2$ is the one where $x_{(u,ex)} = q^*_{(u,ex)}$ and $y_u = \mathrm{ne}(u)$. Recall the construction of $M'_{A,B}$. The states of $M'_{A,B}$ are pairs (v, T), where $v \in Q'$, and $T \subseteq S$ is a set of states of B. The transitions of $M'_{A,B}$ come in three varieties.

Case 1: v is not a call port, and $(v, p'_{v,v'}, v') \in \delta_{M'_A}$. Then we have a corresponding transition $((v,T), p'_{v,v'}, (v', R'(T,v'))) \in \delta_{M'_{A,B}}$, where $p'_{v,v'} = p_{v,v'} \operatorname{ne}(v')/\operatorname{ne}(v)$, and thus $p'_{v,v'} \operatorname{ne}(v) = p_{v,v'} \operatorname{ne}(v')$. Associate a variable $z_{v,v'}$ with each such probability $p'_{v,v'}$, and define the formula: $\varphi_3(\mathbf{y}, \mathbf{z}) \equiv \bigwedge_{(v,v') \in \text{Case1}} z_{v,v'} y_v = p_{v,v'} y_{v'}$.

Case 2: v is a call port, $v = (b, en)$ where v is vertex in component A_i and box b is mapped to component A_j, and $v' = en$, and there is a *red* transition $(v, p'_{v,v'}, v') \in \delta_{M'_A}$. Then there is a *red* transition $((v,T), p'_{v,v'}, (v', R'(T,v')) \in \delta_{M'_{A,B}}$ with the same probability. Here $p'_{v,v'} = \operatorname{ne}(v')/\operatorname{ne}(v)$, and thus $p'_{v,v'} \operatorname{ne}(v) = \operatorname{ne}(v')$. Associate a variable $z_{v,v'}$ with each such probability $p'_{v,v'}$, and define: $\varphi_4(\mathbf{y}, \mathbf{z}) \equiv \bigwedge_{(v,v') \in \text{Case2}} z_{v,v'} y_v = y_{v'}$.

Case 3: v is a call port that has a summary transition $(v, p'_{v,v'}, v')$ in M'_A to a vertex $v' = (b, ex)$, then we have summary transitions of the form $((v,T), p'', (v', T'))$ in $M'_{A,B}$ to the following set of states of the form (v', T'): If there exists a path of M_A that starts at the entry en of A_j and ends at the exit ex (with empty call stack) which, viewed as a string drives B' from T to T', then we include the edge $((v,T), p'_{(v,T),(v',T')}, (v', T'))$ in $\delta_{M'_{A,B}}$, where $p'_{(v,T),(v',T')} = q^*_{((en,T),(ex,T'))} \cdot ne(v')/ne(v)$, and where $q^*_{((en,T),(ex,T'))}$ is the probability of reaching $\langle \epsilon, (ex, T') \rangle$ from $\langle \epsilon, (en, T) \rangle$ in the product RMC $A \otimes B'$. First, compute $A \otimes B'$ and its associated equations $\mathbf{w} = P^{\otimes}(\mathbf{w})$ explicitly. Note that $|A \otimes B'| = O(|A||B'|)$. Let $Q^{\otimes}$ be the set of vertices of $A \otimes B'$. We can compute the set $Q'^{\otimes}$ of vertices v of $A \otimes B'$, for which $\operatorname{ne}(v) > 0$ in PSPACE in $|A \otimes B'|$. Consider now the quantifier-free formula:

$$\varphi_5(\mathbf{w}) \equiv \mathbf{w} = P^{\otimes}(\mathbf{w}) \wedge 0 \preceq \mathbf{w} \wedge \bigwedge_{u \in Q'^{\otimes}} \sum_{ex \in Ex_{c(u)}} w_{(u,ex)} < 1 \wedge \bigwedge_{u \in Q^{\otimes} \setminus Q'^{\otimes}} \sum_{ex \in Ex_{c(u)}} w_{(u,ex)} = 1$$

By Theorem 9, LFP($P^{\otimes}$), is the only vector in $\mathbb{R}^n$ for which $\varphi_5(\mathbf{w})$ holds true. In other words, φ_5 uniquely identifies LFP($P^{\otimes}$). Now, associate a variable $z_{(v,T),(v',T')}$ with each probability $p'_{(v,T),(v',T')}$, where $v = (b, en)$ and $v' = (b, ex)$, and define: $\varphi_6(\mathbf{y}, \mathbf{w}, \mathbf{z}) \equiv \bigwedge_{((v,T),(v',T')) \in \text{Case3}} z_{(v,T),(v',T')} y_v = w_{((en,T),(ex,T'))} y_{v'}$.

Observe, $\bigwedge_{j=1}^{6} \varphi_j$ has a unique solution, and the values of variables $\mathbf{z}$ in this solution identify the probabilities p' on transitions of $M'_{A,B}$. By the methods of section 4, we compute the underlying graph $H'_{A,B}$ of $M'_{A,B}$ and compute the SCCs of $H'_{A,B}$ that contain either an accepting node or an accepting edge. Let us define a revised finite Markov chain, $M''_{A,B}$, in which we remove all SCCs in $M'_{A,B}$ that contain an accepting node or edge, and replace them by a new absorbing node v^*, with a probability 1 transition to itself. Furthermore, in $M''_{A,B}$ we also remove all nodes that can not reach v^*, and all transitions into those nodes. (Technically, some nodes of $M''_{A,B}$ may no longer have full probability on the transitions leaving them, but that is ok for our purposes.)

Now, recall from Markov chain theory (see, e.g., [Bil95]) that for such a finite (sub-)Markov chain $M''_{A,B}$, there is a *linear* system of equations $\mathbf{t} = F(\mathbf{t})$, over variables t_{u,v^*}, where u is any node of $M''_{A,B}$, and where the coefficients in

the linear system $F(\mathbf{t})$ are the probabilties p' on transitions of $M''_{A,B}$ such that the least fixed point solution, $\mathrm{LFP}(F)$, of $\mathbf{t} = F(\mathbf{t})$ assigns to variable t_{u,v^*} the probability that v^* is reachable from u. (In particular, one of the linear equations is $t_{v^*,v^*} = 1$.) Moreover, because we have eliminated from $M''_{A,B}$ all nodes that can not reach v^*, $\mathrm{LFP}(F)$ is the *unique* solution to this system. Thus consider the formula: $\varphi_7(\mathbf{w}, \mathbf{t}) \equiv \mathbf{t} = F(\mathbf{t})$. Thus the formula $\bigwedge_{j=1}^{7} \varphi_j$ has a unique solution in the reals, and the values assigned to variables $t_{(u,v^*)}$ in this solution identify the probability of reaching an accepting SCC from node u in $M'_{A,B}$.

For initial node $u^* = (v_0, \{q_0\})$ of $M'_{A,B}$, and $p \in [0,1]$, the following sentence, ψ, is true in $\mathbb{R}$ iff $P_A(L(B)) \geq p$: $\psi \equiv \exists \mathbf{x}, \mathbf{y}, \mathbf{z}, \mathbf{w}, \mathbf{t} \ \bigwedge_{j=1}^{7} \varphi_j \ \wedge \ t_{u^*,v^*} \geq p$. □

Theorem 11. *For a fixed BA, B, given a Bd-RMC, A, and a rational value $p \in [0,1]$, we can decide whether $P_A(L(B)) \geq p$ in time polynomial in $|A|$.*

Proof. (idea) The proof is a modification of Theorem 10. We extend a technique developed in [EY05]. We use variables only for entry-exit pairs of A and $A \otimes B'$, express all other variables as rational functions of those, and transform the system to one of polynomial constraints in a bounded number of variables. □

References

[AEY01] R. Alur, K. Etessami, and M. Yannakakis. Analysis of recursive state machines. In *Proc. of 13th Int. Conf. on Computer-Aided Verification*, pages 304–313, 2001.

[BGR01] M. Benedikt, P. Godefroid, and T. Reps. Model checking of unrestricted hierarchical state machines. In *Proc. of ICALP'01*, volume 2076 of *LNCS*, pages 652–666, 2001.

[Bil95] P. Billingsley. *Probability and Measure*. J. Wiley and Sons, 3rd edition, 1995.

[BKS05] T. Brázdil, A. Kučera, and O. Stražovský. Decidability of temporal properties of probabilistic pushdown automata. In *Proc. of 22nd STACS'05*. Springer, 2005.

[BPR96] S. Basu, R. Pollack, and M. F. Roy. On the combinatorial and algebraic complexity of quantifier elimination. *J. of the ACM*, 43(6):1002–1045, 1996.

[BR00] T. Ball and S. Rajamani. Bebop: A symbolic model checker for boolean programs. In *SPIN'2000*, volume 1885 of *LNCS*, pages 113–130, 2000.

[Can88] J. Canny. Some algebraic and geometric computations in PSPACE. In *Prof. of 20th ACM STOC*, pages 460–467, 1988.

[CY95] C. Courcoubetis and M. Yannakakis. The complexity of probabilistic verification. *Journal of the ACM*, 42(4):857–907, 1995.

[EE04] J. Esparza and K. Etessami. Verifying probabilistic procedural programs. In *Proc. FSTTCS'04*, 2004. (Invited survey paper).

[EHRS00] J. Esparza, D. Hansel, P. Rossmanith, and S. Schwoon. Efficient algorithms for model checking pushdown systems. In *12th CAV*, volume 1855, pages 232–247. Springer, 2000.

[EKM04] Javier Esparza, Antonín Kučera, and Richard Mayr. Model checking probabilistic pushdown automata. In *Proc. of 19th IEEE LICS'04*, 2004.

[EY05] K. Etessami and M. Yannakakis. Recursive markov chains, stochastic grammars, and monotone systems of non-linear equations. In *Proc. of 22nd STACS'05*. Springer, 2005. (Tech. Report, U. Edinburgh, June 2004).

[GGJ76] M. R. Garey, R. L. Graham, and D. S. Johnson. Some NP-complete geometric problems. In *8th ACM STOC*, pages 10–22, 1976.

[Har63] T. E. Harris. *The Theory of Branching Processes*. Springer-Verlag, 1963.

[Kwi03] M. Kwiatkowska. Model checking for probability and time: from theory to practice. In *Proc. 18th IEEE LICS*, pages 351–360, 2003.

[MS99] C. Manning and H. Schütze. *Foundations of Statistical Natural Language Processing*. MIT Press, 1999.

[PZ93] A. Pnueli and L. D. Zuck. Probabilistic verification. *Inf. and Comp.*, 103(1):1–29, 1993.

[Ren92] J. Renegar. On the computational complexity and geometry of the first-order theory of the reals. parts i,ii, iii. *J. of Symbolic Computation*, pages 255–352, 1992.

[Tiw92] P. Tiwari. A problem that is easier to solve on the unit-cost algebraic ram. *Journal of Complexity*, pages 393–397, 1992.

[Var85] M. Vardi. Automatic verification of probabilistic concurrent finite-state programs. In *Proc. of 26th IEEE FOCS*, pages 327–338, 1985.