Saarland University
Faculty of Natural Sciences and Technology 1
Department of Computer Science

Master Thesis

Optimal Schedulers for Time-Bounded Reachability in CTMDPs

submitted by

Markus Rabe

on September 30, 2009

Supervisor

Prof. Bernd Finkbeiner, Ph.D.

Advisor

Dr. Sven Schewe

Reviewers

Prof. Bernd Finkbeiner, Ph.D.,
Dr. Sven Schewe

**SAARLAND
UNIVERSITY**

COMPUTER SCIENCE

STATEMENT

Hereby, I confirm that this thesis is my own work and that I have documented all sources used.

Saarbrücken, September 30, 2009

Abstract

We study time-bounded reachability in continuous-time Markov decision processes for various scheduler classes. Such reachability problems play a paramount rôle in dependability analysis and the modelling of manufacturing and queueing systems. Consequently, their efficient analysis has been studied intensively, and techniques for the fast approximation of optimal control are well understood. In this thesis, we study the theoretical background of this problem and ask whether an optimal scheduler actually exists. We provide a positive answer to this question for all commonly considered scheduler classes. We provide constructive proofs in which we use the fact that optimal schedulers prefer fast actions over actions with low transition rates when time is short. Optimal schedulers therefore have a simple structure, they converge to a stable strategy as time progresses. For the scheduler classes without direct access to time, we provide a simple procedure to determine optimal schedulers.

Contents

Chapter 1

Introduction

Markov decision processes (MDPs) are a framework that incorporates both nondeterministic and probabilistic choices. They are used in a variety of applications such as the control of manufacturing processes [11, 5] or queueing systems [14]. We study a real time version of MDPs, continuous-time Markov decision processes (CTMDPs), which are a natural formalism for modelling in scheduling [4, 11] and stochastic control theory [5]. CTMDPs can also be seen as a unified framework for different stochastic model types used in dependability analysis [13, 11, 8, 6, 9].

    The analysis of CTMDPs usually concerns the different possibilities to resolve the nondeterminism by means of a scheduler (also called strategy). Typical questions cover qualitative as well as quantitative properties, such as: "Can the nondeterminism be resolved by a scheduler such that a predefined property holds?" or respectively "Which scheduler optimises a given objective function?".

    In this paper, we study the *time-bounded maximum reachability problem* [11, 3, 16, 9, 10] for CTMDPs. Time-bounded reachability is the standard control problem to construct a scheduler that controls the Markov decision process such that the likelihood of reaching a goal region within a given time bound is maximised, and to determine the probability. For CTMDPs, the answer to both questions naturally depends on the power a scheduler has to observe the run of the system—in particular if it can observe time—and on its ability to store and process this information. For the common classes of schedulers, research has focused on efficient approximation techniques [3, 9, 10], while the existence of optimal schedulers has remained open.

Overview.   Given its practical importance, the bounded reachability problem for Markov decision processes has been intensively studied [2, 3, 16, 9, 10]. However, previous research focused on *approximating* the optimal result [3, 10], leaving the question aside whether optimal schedulers exist.

    Unlike for discrete time MDPs, various classes of schedulers for CTMDPs that differ in terms of their power to observe the behaviour of the system have been discussed in the literature [9, 3]. Intuitively, the differences in these classes concern the ability to store information, and to measure time.

Figure 1.1 shows a comparison between the commonly considered scheduler classes, where schedulers that can store the history, its length, or nothing at all are marked H (for history dependent), C (for hop-counting), and P (for positional), respectively. Schedulers that can observe time are marked with a T (timed), and with TT (total time) if they have the power to revoke their decision.

Revoking decisions is a concept first discussed in [9] that extends schedulers on a different level than on what they can observe: while traditional scheduler classes require the schedulers to fix their decisions as soon as they enter a location, TT schedulers may change their decision for an action while residing in the location.

The arrows in Figure 1.1 denote inclusions between scheduler classes, which are direct implications of their definitions. The classes in the figure are ordered top down by their maximal reachability probabilities as known from the literature [3, 9].

In principle, approximating optimal schedulers is simple for all scheduler classes. For schedulers

Figure 1.1: Scheduler hierarchy

that can observe time, it suffices to discretise time and to increase the sample rate [10], and for time-abstract schedulers, it suffices to optimise the reachability within a bounded number of steps and to let this bound grow to infinity [3].

Efficient techniques to determine these rates have, for example, been discussed for uniform CTMDPs—CTMDPs with a constant transition rate—by Baier, Hermanns, Katoen, and Haverkort [3].

Contribution. This paper has contributions on two levels: The clean result on the technical level is a proof that optimal schedulers exist for all commonly considered scheduler classes, but we deem the simple insights on the conceptual level that led to these results to be of similar importance.

Markov + Time = Markov. Markov processes are mathematical models for the random evolution of *memoryless* systems, that is, systems for which the likelihood of future events, at any given moment, depends only on their present state, and not on the past. We observe that *continuous-time Markov chains and decision processes remain Markovian if we add the time that has passed to the state space.*

We use this observation in Chapter 4 to introduce *time-extended* CTMDPs, which contain the time that has passed as part of their state space. This approach has an immediate implication for all time-dependent scheduler classes: It implies without further ado that the scheduler classes TP and TH as well as the classes TTP and TTH coincide, because optimal scheduler decisions in a Markovian system (with simple objectives like time-bounded reachability) cannot depend on the history. As a result, the description of optimal time-dependent schedulers in Chapter 4 is simple.
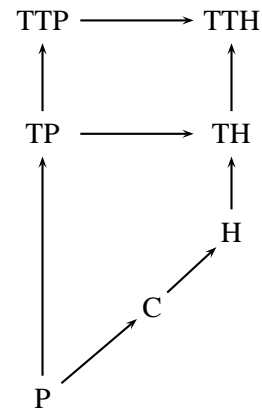
Reasoning about time-abstract scheduler classes is slightly more involved, because time-abstract schedulers do not have access to the precise time that remains for reaching the goal region. Phrased in terms of time-extended CTMDPs, these schedulers do not know precisely in which state of the time-dependent CTMDP they are, but they can infer a distribution over the states in which they could potentially be. While this argument is not used explicitly in Chapter 3, it was the driving factor in our research that led to the construction of optimal time-abstract schedulers. It also provides quick and intuitive alternative proofs for the traditional result [3] that counting and history dependent schedulers provide the same time-bounded reachability probability for uniform CTMDPs, but different ones for non-uniform CTMDPs: while the distribution over the states of the time-extended CTMDP coincides in the first case, it differs in the latter.

Optimal Schedulers. The technical contribution are simple constructive proofs for the existence of optimal time-abstract (Chapter 3) and time-dependent (Chapter 4) schedulers.

For time-abstract schedulers we build on the observation that, if time has almost run out, we can use a *greedy strategy* that optimises our chances to reach our goal in a single step. Reaching it in more steps is then used as a tie-break criterion with decreasing power for increasing distance. We show that such a scheduler exists and is indeed optimal after a certain step bound. For the time-abstract case we also provide an algorithmic solution (Section 3.4).

As a small side-result, we also extended the result that allowing for randomisation does not increase the time-bounded reachability probability for any scheduler class.

A joint publication with Sven Schewe of these results is underway [12].

Chapter 2

Continuous-Time Markov Decision Processes

A *continuous-time Markov decision process* $\mathcal{M}$ is a tuple $(L, Act, \mathbf{R}, \nu, B)$ with a finite set of locations L, a finite set of actions Act, a rate matrix $\mathbf{R} : (L \times Act \times L) \to \mathbb{R}_{\geq 0}$, an initial distribution $\nu \in Dist(L)$, and a goal region $B \subseteq L$. We define the total exit rate for a location l and an action a as $\mathbf{R}(l,a,L) = \sum_{l' \in L} \mathbf{R}(l,a,l')$. For a CTMDP we require that for all locations $l \in L$ there must be an action $a \in Act$ such that $\mathbf{R}(l,a,L) > 0$, and we call such actions *enabled*. We define $Act(l)$ to be the set of enabled actions in location l. If there is only one enabled action per location, a CTMDP $\mathcal{M}$ is a continuous-time Markov chain [7]. If multiple actions are available, we need to resolve the nondeterminism by means of a scheduler (also called strategy or policy). As usual, we assume the goal region to be absorbing, and we use $\mathbf{P}(l,a,l') = \frac{\mathbf{R}(l,a,l')}{\mathbf{R}(l,a,L)}$ to denote the time-abstract transition probability.

Note, that we explicitly distinguish between *locations* and *states*. We consider a state to be a location at a certain point of time. This notion will prove to be helpful when considering time-dependent schedulers in Chapter 4.

Uniform CTMDPs. We call a CTMDP uniform with rate λ if for each location l and action $a \in Act(l)$ the total exit rate $\mathbf{R}(l,a,L)$ is λ. In this case the probability $p_{\lambda t}(n)$ that there are exactly n discrete events (transitions) in time t is Poisson distributed: $p_{\lambda t}(n) = e^{-\lambda t} \cdot \frac{(\lambda t)^n}{n!}$.

We define the *uniformisation* $\mathcal{U}$ of a CTMDP $\mathcal{M}$ as the uniform CTMDP obtained by creating copies $l_{\mathcal{U}}$ for all locations l. We call the new copies unobservable, and the old copies observable locations. Let λ be the maximal total exit rate in $\mathcal{M}$. The new rate matrix $\mathbf{R}_{\mathcal{U}}$ extends $\mathbf{R}$ by first adding the rate $\mathbf{R}_{\mathcal{U}}(l,a,l_{\mathcal{U}}) = \lambda - \mathbf{R}(l,a,L)$ for every location $l \in L$ and action $a \in Act$ of $\mathcal{M}$, and by then copying the outgoing transitions from every observable location l to its unobservable counterpart $l_{\mathcal{U}}$, while the other components remain untouched. The intuition behind this uniformisation technique is that it enables us to distinguish whether a step would

have occurred in the original automaton or not.

Paths.   A *timed path* π in CTMDP $\mathcal{M}$ is a finite sequence in $(L \times Act \times \mathbb{R}_{\geqslant 0})^* \times L = Paths(\mathcal{M})$. We write

$$l_0 \xrightarrow{a_0,t_0} l_1 \xrightarrow{a_1,t_1} \cdots \xrightarrow{a_{n-1},t_{n-1}} l_n.$$

for a sequence π, and we require $t_{i-1} < t_i$ for all $i < n$. The t_i denote the system's time when the events happen. The corresponding *time-abstract path* is defined as $l_0 \xrightarrow{a_0} l_1 \xrightarrow{a_1} \cdots \xrightarrow{a_{n-1}} l_n$. We use $Paths_{abs}(\mathcal{M})$ to denote the set of all such projections and $|\cdot|$ to count the number of actions in a path. Concatenation of paths π, π' will be written as $\pi \circ \pi'$ if the last state of π is the first state of π'.

Schedulers.   The system's behaviour is not defined only by the CTMDP, but also by a scheduler that resolves the nondeterminism. When analysing properties of a CTMDP, such as the reachability probability, we usually quantify over a class of schedulers. We restrict all scheduler classes to those schedulers creating a measurable probability space (cf. [15]), and we consider the following common classes, which differ in their power to observe events and to revoke their decisions:

○ *Total time history-dependent* (TTH) schedulers $Paths(\mathcal{M}) \times \mathbb{R}_{\geqslant 0} \to D$
that map timed paths and the elapsed time to decisions.

○ *Total time positional* (TTP) schedulers $L \times \mathbb{R}_{\geqslant 0} \to D$
that map locations and the elapsed time to decisions.

○ *Timed history* (TH) schedulers $Paths(\mathcal{M}) \to D$
that map timed paths to decisions.

○ *Timed positional* (TP) schedulers $L \times \mathbb{R}_{\geqslant 0} \to D$
that map locations and the time until the last state change to decisions.

○ *Time-abstract history-dependent* (H) schedulers $Paths_{abs}(\mathcal{M}) \to D$
that map time-abstract paths to decisions.

○ *Time-abstract hop-counting* (C) schedulers $L \times \mathbb{N} \to D$
that map locations and the number of hops (length of the path) to decisions.

○ *Positional* (P) or memoryless schedulers $L \to D$
that map locations to decisions.

Decisions D are either randomised (R), in which case $D = Dist(Act)$ is the set of distributions over enabled actions, or are restricted to deterministic (D) choices, that is $D = Act$. Wherever necessary to distinguish randomised and deterministic versions we will add a postfix to the scheduler class, for example HD and HR.

Induced Probability Space. We build our probability space in the natural way: we first define the probability measure for cylindric sets of paths that start with

$$l_0 \xrightarrow{a_0,t_0} l_1 \xrightarrow{a_1,t_1} \cdots \xrightarrow{a_{n-1},t_{n-1}} l_n,$$

with $t_j \in I_j$ for all $j < n$, and for non-overlapping open intervals $I_0, I_1, \ldots, I_{n-1}$, to be the usual probability that a path starts with these actions for a randomised scheduler $\mathcal{S}$ that may not revoke its decisions, and such that $\mathcal{S}(l_0 \xrightarrow{a_0,t_0} \cdots \xrightarrow{a_{i-1},t_{i-1}} l_i)$ is equivalent for all $(t_0, \ldots, t_{i-1}) \in I_0 \times \ldots \times I_{i-1}$:

$$\int_{t_0 \in I_0, t_1 \in I_1, \ldots, t_{n-1} \in I_{n-1}} \prod_{i=0}^{n-1} \mathcal{S}(l_0 \xrightarrow{a_0,t_0} \cdots \xrightarrow{a_{i-1},t_{i-1}} l_i)(a_i) \cdot \mathbf{R}(l_i, a_i, l_{i+1}) \cdot e^{-\mathbf{R}(l_i, a_i, L)(t_i - t_{i-1})},$$

assuming $t_{-1} = 0$.

From this basic building block, we build our probability measure for measurable sets of paths and measurable sets of schedulers in the usual way [15]. The similar space for TT schedulers, which may revoke their decisions, is described in Section 4.2.

Time-Bounded Reachability Probability. For a given CTMDP $\mathcal{M} = (L, Act, \mathbf{R}, \nu, B)$ and a given measurable scheduler $\mathcal{S}$ that resolves the nondeterminism, we use the following notations for the probabilities:

- $Pr_{\mathcal{S}}^{\mathcal{M}}(l, t)$ is the probability of reaching the goal region B in time t when starting in location l,

- $Pr_{\mathcal{S}}^{\mathcal{M}}(t) = \sum_{l \in L} \nu(l) Pr_{\mathcal{S}}^{\mathcal{M}}(l, t)$ denotes the probability of reaching the goal region B in time t,

- $k - Pr_{\mathcal{S}}^{\mathcal{M}}(t)$ denotes the probability of reaching the goal region B in time t *and* with at most k discrete steps, and

- $PR_{\mathcal{S}}^{\mathcal{M}}(\pi, t)$ is the probability to traverse the time-abstract path π within time t, if π does not visit the goal region B, and 0 if π contains a location in B.

As usual, the supremum of the time-bounded reachability probability over a particular scheduler class is called the time-bounded reachability of $\mathcal{M}$ for this scheduler class, and we use 'max' instead of 'sup' to indicate that this value is taken for some *optimal scheduler $\mathcal{S}$* of this class.

Step Probability Vector. Given a scheduler $\mathcal{S}$ and a location l for a CTMDP $\mathcal{M}$, we define the *step probability vector $d_{l,\mathcal{S}}$* of infinite dimension. An entry $d_{l,\mathcal{S}}[i]$ for $i \geq 0$ denotes the probability to reach goal region B in up to i steps from location l (not considering any time constraints).

Chapter 3

Time-abstract Scheduling

In this chapter, we show that *optimal* schedulers exist for all natural time-abstract classes, that is, for CD, CR, HD, and HR. Moreover, we show that there are optimal schedulers that become positional after a small number of steps, which we compute with a simple algorithm. We also show that randomisation does not yield any advantage: deterministic schedulers are as good as randomised ones. Our proofs are constructive, and thus allow for the construction of optimal schedulers. This also provides the first procedure to precisely determine the time-bounded reachability probability, because we can now reduce this problem to solving the time-bounded reachability problem of Markov chains [2].

Our proof consists of two parts. We first consider the class of uniform CTMDPs, which are much simpler to treat in the time-abstract case, because we can use Poisson distributions to describe the number of steps taken within a given time bound. For uniform CTMDPs it is already known that the supremum over the bounded reachability collapses for all time-abstract scheduler classes from CD to HR [3]. It therefore suffices to show that there is a CD scheduler which takes this value.

We then show that a similar claim holds for CD and HD scheduler in the general class of not necessarily uniform CTMDPs. In this case, it also holds that there are simple optimal schedulers that converge against a positional scheduler after a finite number of steps, and that randomisation does not improve the time-bounded reachability probability. However, in the non-uniform case the time-abstract path contains more information about the remaining time than its length only, and bounded reachability of history dependent and counting schedulers usually deviate [3].

We start this section with the introduction of *greedy schedulers*, HD schedulers that favour reachability in a small number of steps over reachability with a larger number of steps; the positional schedulers against which the CD and HD schedulers converge are such greedy schedulers.

3.1   Greedy Schedulers

The natural objective when seeking optimal schedulers is to maximise time-bounded reachability $Pr_{\mathcal{S}}^{\mathcal{M}}(l,t)$ for every location l with respect to a particular scheduler class such as HD. Unfortunately, this optimisation problem is comparably complex.

However, when the remaining time t is close to 0, then increasing the likelihood of reaching the goal region in few steps dominates the impact of reaching it later. While we have no direct access to the remaining time in the time-abstract case, we can infer the distribution over the remaining time from the time-abstract history (or its length). Since the expected remaining time converges to 0 when the number of transitions goes to infinity, we can argue in a way similar to the time-dependent case.

This motivates the introduction of greedy schedulers: We call an HD scheduler *greedy*[1] if it maximises the step probability vector of every location l with respect to the lexicographic order (e.g. $(0,0.2,0.3,\dots) >_{lex} (0,0.1,0.4,\dots)$). To prove the existence of greedy schedulers, we draw from the fact that the supremum $d_l = \sup_{\mathcal{S} \in HD} d_{l,\mathcal{S}}$ obviously exists, where the supremum is to be read as a supremum with respect to the lexicographic order. An action $a \in Act(l)$ is called *greedy* for a location $l \notin B$ if it satisfies $shift(d_l) = \sum_{l' \in L} \mathbf{P}(l,a,l')d_{l'}$, where $shift(d_l)$ shifts the vector by one position (that is, $shift(d_l)[i] = d_l[i+1]$ $\forall i \in \mathbb{N}$). For locations l in the goal region B, all enabled actions $a \in Act(l)$ are greedy.

Lemma 1 *Greedy schedulers exist, and they can be described as the class of schedulers that choose a greedy action upon every reachable time-abstract path.*

Proof.   It is plain that, for every non-goal location $l \notin B$, $shift(d_l) \geq \sum_{l' \in L} \mathbf{P}(l,a,l')d_{l'}$ holds for every action a, and that equality must hold for some.

For a scheduler $\mathcal{S}$ that always chooses greedy actions, a simple inductive argument shows that $d_l[i] = d_{l,\mathcal{S}}[i]$ holds for all $i \in \mathbb{N}$, while it is easy to show that $d_l > d_{l,\mathcal{S}}$ holds if $\mathcal{S}$ deviates from greedy decisions upon a path that is possible under its own scheduling policy. $\square$

This allows in particular to fix a positional *standard greedy scheduler* by fixing a greedy action for every location.

So far, we have only shown the existence of a greedy scheduler, but not argued how to determine the set of greedy actions. If a scheduler $\mathcal{S}$ starts in a location l with a non-greedy action a, then $shift(d_{l,\mathcal{S}}) \leq \sum_{l' \in L} \mathbf{P}(l,a,l')d_{l'}$ holds true. The sum $\sum_{l' \in L} \mathbf{P}(l,a,l')d_{l'}$ corresponds to the scheduler choosing the non-greedy action a at location l and acting greedy in all further steps. Let $d_{l,a}$ denote the step probability vector of such schedulers.

We know that $d_{l,\mathcal{S}} \leq d_{l,a} < d_l$. Hence, there is not only a difference between $d_{l,\mathcal{S}}$ and d_l, this difference will not occur at a higher index as the first difference between the newly defined $d_{l,a}$ and d_l. The finite number of locations and actions thus implies

---

[1]The k-greedy schedulers introduced in [3] are greedy with respect to a different goal: they maximise the partial sum $\sum_{i=0}^{k} d_{l,\mathcal{S}}[i] \cdot p_{\lambda t}(i)$ for a given k. They correspond to the k-optimal schedulers used in this paper.

the existence of a bound k on the occurrence of this first difference between $d_{l,a}$ and d_l as well as $d_{l,S}$ and d_l. While the existence of such a k suffices to show the existence of optimal schedulers, we need an upper bound for k to actually identify greedy actions. In Appendix 4.2 we show that this constant $k < |L|$ is smaller than the CTMDP itself.

Having established such a bound k, it suffices to compare schedulers up to this bound. This provides us with the greedy actions, and also with the initial sequence $d_{l,a}[0], d_{l,a}[1], \ldots, d_{l,a}[k]$ for all locations l and actions a. Finally, we determine a positive lower bound $\mu > 0$ for the first non-zero entry of the vectors $d_l - d_{l,a}$. We call this lower bound μ the *discriminator* of the CTMDP. The intuition behind the discriminator is that it represents the minimal advantage of the greedy strategy over all other strategies.

3.2 Uniform CTMDPs

In this subsection, we show that every CD or HD scheduler for a uniform CTMDP can be transformed into a scheduler that converges to the positional standard greedy scheduler.

In the quest for an optimal CD scheduler, it is useful to consider the fact that the maximum reachability probability can be computed using the step probability vector, because the likelihood that a particular number of steps happen in time t is independent of the scheduler:

$$Pr_S^{\mathcal{M}}(t) = \sum_{l \in L} \nu(l) \sum_{i=0}^{\infty} d_{l,S}[i] \cdot p_{\lambda t}(i). \tag{3.1}$$

Moreover, the Poisson distribution $p_{\lambda t}$ has the useful property that the probability of taking k steps is falling very fast. We define the *greed bound* $n_{\mathcal{M}}$ to be a natural number, for which

$$\mu p_{\lambda t}(n) \geq \sum_{i=1}^{\infty} p_{\lambda t}(n+i) \qquad \forall n \geq n_{\mathcal{M}} \tag{3.2}$$

holds true. It suffices to choose $n_{\mathcal{M}} \geq \frac{2\lambda t}{\mu}$ since it implies $\mu p_{\lambda t}(n) \geq 2 p_{\lambda t}(n+1)$, $\forall n > n_{\mathcal{M}}$ (which yields (3.2) by simple induction). Such a greed bound implies that the decrease in likelihood of reaching the goal region in few steps caused by making a non-greedy decision after the greed bound dwarfs any potential later gain. We use this observation to improve any given CD or HD scheduler S that makes a non-greedy decision after $\geq n_{\mathcal{M}}$ steps by replacing the behaviour after this history by a greedy scheduler. Finally, we use the interchangeability of greedy schedulers to introduce a scheduler $\overline{S}$ that makes the same decisions as S on short histories and follows the standard greedy scheduling policy once the length of the history reaches the greed bound. For this scheduler, we show that $Pr_{\overline{S}}^{\mathcal{M}}(t) \geq Pr_S^{\mathcal{M}}(t)$ holds true.

Theorem 1 *For uniform CTMDPs, there is an optimal scheduler for the classes CD and HD that converges to the standard greedy scheduler after $n_{\mathcal{M}}$ steps.*

Proof. Let us consider any HD scheduler $\mathcal{S}$ that makes a non-greedy decision after a time-abstract path π of length $|\pi| \geq n_{\mathcal{M}}$ with last location l. If the path ends in, or has previously passed, the goal region, or if the probability of the history π is 0, that is, if it cannot occur with the scheduling policy of $\mathcal{S}$, then we can change the decision of $\mathcal{S}$ on every path starting with π arbitrarily—and in particular to the standard greedy scheduler—without altering the reachability probability.

If $PR_{\mathcal{S}}^{\mathcal{M}}(\pi, t) > 0$, then we change the decisions of the scheduler $\mathcal{S}$ for paths with prefix π such that they comply with the standard greedy scheduler. We call the resulting HD scheduler $\mathcal{S}'$ and analyse the change in reachability probability using Equation (3.1):

$$Pr_{\mathcal{S}'}^{\mathcal{M}}(t) - Pr_{\mathcal{S}}^{\mathcal{M}}(t) = PR_{\mathcal{S}}^{\mathcal{M}}(\pi, t) \cdot \sum_{i=0}^{\infty} (d_l[i] - d_{l,\mathcal{S}_{\pi}}[i]) \cdot p_{\lambda t}(|\pi| + i),$$

where $\mathcal{S}_{\pi} : \pi' \mapsto \mathcal{S}(\pi \circ \pi')$ is the HD scheduler which prefixes its input with the path π and then calls the scheduler $\mathcal{S}$. The greedy criterion implies $d_l > d_{l,\mathcal{S}_{\pi}}$ with respect to the lexicographic order, and we can apply Equation 3.2 to deduce that the difference $Pr_{\mathcal{S}'}^{\mathcal{M}}(t) - Pr_{\mathcal{S}}^{\mathcal{M}}(t)$ is non-negative.

Likewise, we can concurrently change the scheduling policy to the standard greedy scheduler for all paths of length $\geq n_{\mathcal{M}}$ for which the scheduler $\mathcal{S}$ makes non-greedy decisions. In this way, we obtain a scheduler $\mathcal{S}''$ that makes non-greedy decisions only in the first $n_{\mathcal{M}}$ steps, and yields a (not necessarily strictly) better time-bounded reachability probability than $\mathcal{S}$.

Since all greedy schedulers are interchangeable without changing the bounded reachability probability (and even without altering the step probability vector), we can modify $\mathcal{S}''$ such that, after $\geq n_{\mathcal{M}}$ steps, it does not only follow any greedy scheduling policy, but complies with the standard greedy scheduler, resulting in another scheduler $\overline{\mathcal{S}}$ with the same time-bounded reachability probability as $\mathcal{S}''$. Note that $\overline{\mathcal{S}}$ is counting if $\mathcal{S}$ is counting.

Hence, the supremum over the bounded reachability of all CD/HD schedulers is equivalent to the supremum over the bounded reachability of CD/HD schedulers that deviate from the standard greedy scheduler only in the first $n_{\mathcal{M}}$ steps. This class is finite, and the supremum over the bounded reachability is therefore the maximal bounded reachability obtained by one of its representatives. $\qquad\square$

Hence, we have shown the existence of a—simple—optimal time-bounded CD scheduler. Using the fact that the suprema over the time-bounded reachability probability coincide for CD, CR, HD, and HR scheduler [3], we can infer that such a scheduler is optimal for all of these classes.

Corollary 1 $\max\limits_{\mathcal{S} \in CD} Pr_{\mathcal{S}}^{\mathcal{M}}(t) = \max\limits_{\mathcal{S} \in CR} Pr_{\mathcal{S}}^{\mathcal{M}}(t) = \max\limits_{\mathcal{S} \in HD} Pr_{\mathcal{S}}^{\mathcal{M}}(t) = \max\limits_{\mathcal{S} \in HR} Pr_{\mathcal{S}}^{\mathcal{M}}(t)$ *holds for all uniform CTMDPs $\mathcal{M}$.* $\qquad\square$

The existential proof above does not directly lead to a construction though. In Section 3.4 we present a method to obtain the optimal scheduler.

3.3   Non-uniform CTMDPs

Reasoning over non-uniform CTMDPs is harder than reasoning over uniform CT-MDPs, because the likelihood of seeing exactly k steps does not adhere to the simple Poisson distribution, but depends on the precise history. Even if two paths have the same length, they may refer to different probability distributions over the time passed so far. Knowing the time-abstract history therefore provides a scheduler with more information about the system's state than merely its length. As a result, it is simple to construct example CTMDPs, for which history dependent and counting schedulers can obtain different time-bounded reachability probabilities [3].

In this subsection, we extend the results from the previous subsection to general CTMDPs. We show that simple optimal CD/HD scheduler exist, and that randomisation does not yield an advantage:

$$\max_{S \in CD} Pr_S^{\mathcal{M}}(t) = \max_{S \in CR} Pr_S^{\mathcal{M}}(t) \qquad \text{and} \qquad \max_{S \in HD} Pr_S^{\mathcal{M}}(t) = \max_{S \in HR} Pr_S^{\mathcal{M}}(t).$$

To obtain this result, we work on the uniformisation $\mathcal{U}$ of $\mathcal{M}$ instead of working on $\mathcal{M}$ itself. We argue that the behaviour of a general CTMDP $\mathcal{M}$ can be viewed as the observable behaviour of its uniformisation $\mathcal{U}$, using a scheduler that does not *see* the new transitions and locations. Schedulers from this class can then be replaced by (or viewed as) schedulers that do not *use* the additional information. And finally, we can approximate schedulers that do not use the additional information by schedulers that do not use it initially, where initially means until the number of visible steps—and hence in particular the number of steps—exceeds the greed bound $n_{\mathcal{U}}$ of the uniformisation $\mathcal{U}$ of $\mathcal{M}$. Comparable to the argument from the proof of Theorem 1, we show that we can restrict our attention to the standard greedy scheduler after this initial phase, which leads again to a situation where considering a finite class of schedulers suffices to obtain the optimum.

Lemma 2 *The greedy decisions and the step probability vector coincide for the observable and unobservable copy of each location in the uniformisation $\mathcal{U}$ of any CTMDP $\mathcal{M}$.*

Proof.   The observable and unobservable copy of each location reach the same successors under the same actions with the same transition rate.   □

We can therefore choose a positional *standard greedy scheduler* whose decisions coincide for the observable and unobservable copy of each location.

For the *uniformisation $\mathcal{U}$* of a CTMDP $\mathcal{M}$, we define the function $vis : Paths_{abs}(\mathcal{U}) \to Paths_{abs}(\mathcal{M})$ that maps a path π of $\mathcal{U}$ to the corresponding path in $\mathcal{M}$, the *visible path*, by deleting all unobservable locations and their preceding transitions from π. (Note that all paths in $\mathcal{U}$ start in an observable location.) We call a scheduler *n-visible* if its decisions only depend on the visible path and coincide for the observable and unobservable copy of every location for all paths containing up to n visible steps. We call a scheduler *visible* if it is n-visible for all $n \in \mathbb{N}$.

We call a HD/HR scheduler an (n-)visible HD/HR scheduler if it is (n-)visible, and we call an (n-)visible HD/HR scheduler a visible CD/CR scheduler if its decisions depend only on the length of the visible path, and an n-visible CD/CR scheduler if its decisions depend only on the length of the visible path for all paths containing up to n visible steps. The respective classes are denoted with according prefixes, for example, n-vCD. Note that (n-)visible counting schedulers are not counting.

It is a simple observation that we can study visible CD, CR, HD, and HR schedulers on the uniformisation $\mathcal{U}$ of a CTMDP $\mathcal{M}$ instead of studying CD, CR, HD, and HR schedulers on $\mathcal{M}$.

Lemma 3 $\mathcal{S} \mapsto \mathcal{S} \circ vis$ *is a bijection from visible CD, CR, HD, or HR schedulers for the uniformisation $\mathcal{U}$ of a CTMDP $\mathcal{M}$ onto CD, CR, HD, or HR schedulers, respectively, for $\mathcal{M}$ that preserves the time-bounded reachability probability:* $Pr_{\mathcal{S}}^{\mathcal{U}}(t) = Pr_{\mathcal{S} \circ vis}^{\mathcal{M}}(t).$ □

At the same time, copying the argument from the proof of Theorem 1, an $n_{\mathcal{U}}$-visible CD or HD scheduler $\mathcal{S}$ can be adjusted to the $n_{\mathcal{U}}$-visible CD or HD scheduler $\overline{\mathcal{S}}$ that deviates from $\mathcal{S}$ only in that it complies with the standard greedy scheduler for $\mathcal{U}$ after $n_{\mathcal{U}}$ visible steps, without decreasing the time-bounded reachability probability. These schedulers are visible schedulers from a finite sub-class, and hence some representative of this class takes the optimal value.

Lemma 4 *The following equations hold for the uniformisation $\mathcal{U}$ of a CTMDP $\mathcal{M}$:*

$$\max_{\mathcal{S} \in n_{\mathcal{U}} - vCD} Pr_{\mathcal{S}}^{\mathcal{U}}(t) = \max_{\mathcal{S} \in vCD} Pr_{\mathcal{S}}^{\mathcal{U}}(t) \quad and \quad \max_{\mathcal{S} \in n_{\mathcal{U}} - vHD} Pr_{\mathcal{S}}^{\mathcal{U}}(t) = \max_{\mathcal{S} \in vHD} Pr_{\mathcal{S}}^{\mathcal{U}}(t).$$

Proof. We have shown in Theorem 1 that turning to the standard greedy scheduling policy after $n_{\mathcal{U}}$ or more steps can only increase the time-bounded reachability probability. This implies in particular that we can turn to the standard greedy scheduler after $n_{\mathcal{U}}$ *visible* steps.

The scheduler resulting from this adjustment does not only remain $n_{\mathcal{U}}$-visible, it becomes a visible CD and HD scheduler, respectively. Moreover, it is a scheduler from the finite subset of CD or HD schedulers, respectively, whose behaviour may only deviate from the standard scheduler within the first $n_{\mathcal{U}}$ visible steps. □

We can therefore construct optimal CD and HD scheduler for every CTMDP $\mathcal{M}$. To prove that optimal CD and HD scheduler are also optimal CR and HR scheduler, respectively, we first prove the simpler lemma that this holds for k-bounded reachability.

Lemma 5 *k-optimal CD or HD schedulers are also k-optimal CR or HR schedulers, respectively.*

Proof. For a CTMDP $\mathcal{M}$ we can turn an arbitrary CR or HR scheduler $\mathcal{S}$ into a CD or HD scheduler $\mathcal{S}'$ with a time and k-bounded reachability probability that is at

least as good as the one of $\mathcal{S}$ by first determinising the scheduler decisions from the $k+1$st step onwards—this has obviously no impact on k-bounded reachability—and then determinising the remaining randomised choices.

Replacing a single randomised decision on a path π (for history dependent schedulers) or on a set of paths Π (for counting schedulers) that end(s) on a location l is safe, because the time and k-bounded reachability probability of a scheduler is an affine combination—the affine combination defined by $\mathcal{S}(\pi)$ and $\mathcal{S}(|\pi|,l)$, respectively—of the $|Act(l)|$ schedulers resulting from determinising this single decision. Hence, we can pick one of them whose time and k-bounded reachability probability is at least as high as the one of $\mathcal{S}$.

As the number of these randomised decisions is finite ($\leq k|L|$ for CR, and $\leq k^{|L|}$ for HR schedulers), this results in a deterministic scheduler after a finite number of improvements. $\qquad\square$

Theorem 2 *Optimal CD schedulers are also optimal CR schedulers.*

Proof. First, the probability that the goal region B is reached in more than k steps converges to 0, independent of the scheduler. Together with Lemma 5, this implies

$$\sup_{\mathcal{S}\in CR} Pr_{\mathcal{S}}^{\mathcal{M}}(t) = \lim_{n\to\infty} \sup_{\mathcal{S}\in CR} n-Pr_{\mathcal{S}}^{\mathcal{M}}(t) = \lim_{n\to\infty} \sup_{\mathcal{S}\in CD} n-Pr_{\mathcal{S}}^{\mathcal{M}}(t) \leq \max_{\mathcal{S}\in CD} Pr_{\mathcal{S}}^{\mathcal{M}}(t),$$

while $\geq$ is implied by $CD \subseteq CR$. $\qquad\square$

Analogously, we can prove the similar theorem for history dependent schedulers:

Theorem 3 *Optimal HD schedulers are also optimal HR schedulers.* $\qquad\square$

3.4 A Practical Approach

In this section we present a procedure to construct an optimal scheduler in the time-abstract case. For the sake of simplicity we assume a uniform CTMDP $\mathcal{M}$.

There is an obvious method how to construct an optimal scheduler: Compute and compare the reachability probabilities for all finitely many history dependent schedulers that may choose freely until step $n_{\mathcal{M}}$ and act greedy afterwards. In order to compute these values, we construct a continuos time Markov chain (CTMC)—a CTMDP without nondeterminism—for each scheduler $\mathcal{S}$ and compute its reachability probability.

As we know that in uniform CTMDPs the classes of counting schedulers and history dependent schedulers yield the same reachability probabilities, it suffices to consider the CTMDP which encodes the step number in its state space (up to step $n_{\mathcal{M}}$) in the natural way and fix the decisions according to $\mathcal{S}$ in order to obtain a CTMC.

It is plain that this CTMC yields the same reachability probability as the original CTMDP $\mathcal{M}$ under scheduler $\mathcal{S}$: $Pr_{\mathcal{S}}^{\mathcal{M}}(t)$. These probabilities can be computed and
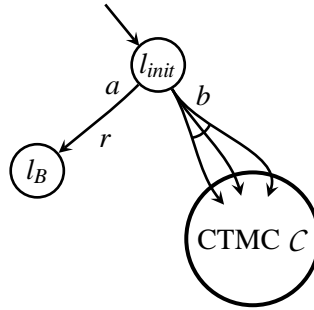
Figure 3.1: A simple CTMDP

compared by means of the methods of Aziz et al. [2]. Unfortunately, the complexity of this result is unknown which prevents us from assessing this method.

We can, however, determine a lower bound by reducing the core question of CSL (continuous stochastic logic, cf. [2, 3]) model checking in CTMCs to the comparison of actions in CTMDPs:

Theorem 4 *The search for an optimal scheduler is at least as hard as checking the formula $\mathcal{P}_{<p}(\Diamond^{[0,t]}B)$ a CTMC for a rational number p.*

Proof.   Given a CTMC $\mathcal{C} = (L, Act, \mathbf{R}, \nu, B)$, with $Act = \{b\}$, and a simple CSL formula $\Phi = \mathcal{P}_{<p}(\Diamond^{[0,t]}B)$ we construct a CTMDP $\mathcal{M}_{\mathcal{C}}$ for which the (only) decision which action to take corresponds to the evaluation of Φ. We assume a unique initial state l_{init} with $\nu(l_{init}) = 1$.

The CTMDP $\mathcal{M}_{\mathcal{C}} = (L', Act', \mathbf{R}', \nu', B')$ is constructed as follows: $L' = L \dot{\cup} \{s_B\}$, $Act' = \{a, b\}$, $\nu'(l_{init}) = 1$, and $B' = B \dot{\cup} \{l_B\}$. Further, $\mathbf{R}'$ extends $\mathbf{R}$ by the entries for action a which are all zero up to the entry $\mathbf{R}'(l_{init}, a, l_B) = r$ for some constant $r \in \mathbb{R}$. Figure 3.1 illustrates the construction. Constant r is chosen such that the time-bounded reachability probability of reaching l_B with action a is exaclty p. This constant r clearly exists and therefore the search for the optimal scheduler (out of the two available) directy corresponds to the evaluation of the formula Φ. $\square$

Chapter 4

Time-dependent Scheduling

In this chapter we make use of a simple but illuminative shift in our view on the control problem for a CTMDP $\mathcal{M}$: We consider the time that has passed as part of the state-space[1] of a time-extended CTMDP (tCTMDP), turning the time-bounded reachability problem to reach B in time t_0 into an ordinary reachability problem to reach $B \times [0,t_0]$ in a tCTMDP $\mathcal{M}_{t_0}$.

This extension has obviously no effect on the Markovian character of the tCTMDP. In particular for a TT scheduler $\mathcal{S}$, which can revoke its decisions, the probability $Pr_{\mathcal{S}}^{\mathcal{M}}((l,t))$ to reach the goal region $B \times [0,t_0]$ from a state (l,t) in the time extended CTMDP $\mathcal{M}_{t_0}$ is independent of the history.

For a traditional time-dependent schedulers $\mathcal{S}$, the probability $Pr_{\mathcal{S}}^{\mathcal{M}}((l,t))$ to reach a location in the goal region from a state (l,t) is memoryful in general, as the decisions made by the scheduler $\mathcal{S}$ depend on the time that the location l was entered. However, the behaviour becomes memoryless if we focus on the points of time at which a discrete transition took place.

In both cases it is simple to translate positional schedulers for the resulting time-extended CTMDP $\mathcal{M}_{t_0}$ to equivalent TTP/TP schedulers for the original CTMDP $\mathcal{M}$.

4.1 Timed Schedulers

For every TP scheduler $\mathcal{S}$, we have $Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}((l,t)) = 1$ for all goal states $(l,t) \in B \times [0,t_0]$, as we have reached the goal region in time in this case, and $Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}((l,t)) = 0$ for all locations $l \in L$ and all $t > t_0$, because the goal region cannot be reached in time any longer if it has not been visited before.

For a measurable deterministic positional scheduler $\mathcal{S}$ and a non-goal locations $l' \notin B$ and times $t \in [0,t_0]$, we will reach the goal region in time (provided we have not reached it before), if we reach it in time with or after the following transition.

---

[1]Adding the time to the state-space leads to a construction that recalls the semantics of timed automata [1], although the treatment of time is much simpler here, because we have only one clock and no resets.

Hence,

$$Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}\big((l,t)\big) = \sum_{l' \in L} \mathbf{R}\big(l, \mathcal{S}((l,t)), l'\big) \int_t^{\infty} Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}(l', \tau) e^{-\mathbf{R}(l, \mathcal{S}((l,t)), L)\tau} d\tau$$

holds true, where $Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}\big((l,t)\big)$ denotes the probability of reaching the goal region when the location l is *entered* at time t. Different to tCTMDPs for TT schedulers (cf. Section 4.2), tCTMDPs for traditional T schedulers therefore have a discrete flavour.

Naturally, this shift in our way of looking at the problem has no influence on the probability of reaching our objective, and the following equations must hold:

$$\sup_{\mathcal{S} \in TP} Pr_{\mathcal{S}}^{\mathcal{M}}(l, t_0 - t) = \sup_{\mathcal{S} \in P} Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}\big((l,t)\big)$$

and

$$\sup_{\mathcal{S} \in TP} Pr_{\mathcal{S}}^{\mathcal{M}}(t_0) = \sum_{l \in L} \nu(l) \sup_{\mathcal{S} \in P} Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}\big((l,0)\big).$$

The hard part—both for T and TT schedulers—is to show that an optimal measurable scheduler exists. We start with a proof that there are optimal schedulers in the class of randomised history and time-dependent schedulers that are deterministic and positional.

Theorem 5 $\max_{\mathcal{S} \in TP} Pr_{\mathcal{S}}^{\mathcal{M}}(t) = \sup_{\mathcal{S} \in TH} Pr_{\mathcal{S}}^{\mathcal{M}}(t)$, *and randomisation does not improve the result.*

Proof. The formulas given above for positional schedulers, as well as similar formulas for history dependent schedulers, are clearly dominated by the functions defined by

$$Pr_P^{\mathcal{M}_{t_0}}\big((l,t)\big) = \max_{a \in Act(l)} \sum_{l' \in L} \mathbf{R}(l, a, l') \int_t^{\infty} Pr_P^{\mathcal{M}_{t_0}}(l', \tau) e^{-\mathbf{R}(l, a, L)\tau} d\tau.$$

For an extension to randomised schedulers, the maximum over the actions needs to be replaced by a supremum over the distributions in an intermediate step, but as suprema over affine combinations over a finite set of values are taken in one of these values, the same function is dominating the functions for measurable history dependent randomised schedulers as well.

The hard part is to show that a measurable scheduler exists that takes these maxima, that is, that no non-measurable change between different actions is required. To prove this, we show how to construct a measurable deterministic positional scheduler $\mathcal{S}$ that always chooses actions a that take the maximum value.

To determine suitable scheduler decisions for a location l for such a scheduler $\mathcal{S}$, we disintegrate $[0, t_0]$ into measurable sets $\{T_a \mid a \in Act(l)\}$, such that $\mathcal{S}$ only makes

decisions that maximise $\sum_{l' \in L} \mathbf{R}(l,a,l') \int_t^\infty Pr_P^{\mathcal{M}_{t_0}}\big((l',\tau)\big)e^{-\mathbf{R}(l,a,L)\tau}d\tau$. (For positions outside of $[0,t_0]$, that is, for times behind our time bound t_0, the behaviour of the scheduler does not matter, and $\mathcal{S}(l,t)$ can be fixed to any constant decision $a \in Act(l)$ for all $t > t_0$.)

We start with fixing an arbitrary order $\succ$ on the actions in $Act(l)$, and introduce, for each point $t \in [0,t_0]$, an order $\succcurlyeq_t$ on the actions determined by the value of $\sum_{l' \in L} \mathbf{R}(l,a,l') \int_t^\infty Pr_P^{\mathcal{M}_{t_0}}\big((l',\tau)\big)e^{-\mathbf{R}(l,a,L)\tau}d\tau$, using $\succ$ as a tie-breaker.

1. For the action a in $Act(l)$ that is minimal with respect to $\succ$, we start by fixing the open set $O_a = [0,t_0]$ of points in time where the scheduler does not make a decision $a' \succ a$ (where open set in this proof refers to sets open in $[0,t_0]$).

2. We then define the set T_a as the points $t \in O_a$ in time, for which the action a is maximal with respect to $\succcurlyeq_t$.

   Then T_a is an open measurable set with a countable fringe, and for all points $t \in \overline{T_a} \smallsetminus T_a$ it holds that a maximises $\sum_{l' \in L} \mathbf{R}(l,b,l') \int_t^\infty Pr_P^{\mathcal{M}_{t_0}}\big((l',\tau)\big)e^{-\mathbf{R}(l,b,L)\tau}d\tau$ among all actions $b \in Act(l)$, though not strictly. (A detailed description why the continuity of $\sum_{l' \in L} \mathbf{R}(l,b,l') \int_t^\infty Pr_P^{\mathcal{M}_{t_0}}\big((l',\tau)\big)e^{-\mathbf{R}(l,b,L)\tau}d\tau$ for all actions $b \in Act(l)$ implies that T_a is open, measurable, and has a countable fringe is supplied in Section 4.2.)

3. We fix $\mathcal{S}(l,t) = a$ for all $t \in \overline{T_a} \cap O_a$.

4. If there is a next smaller (with respect to $\succ$) action $a' = \max\{a'' \prec a\}$, than we fix the new open set $O_{a'} = O_a \smallsetminus \overline{T_a}$ for a', and proceed with Step 2.

Repeating this for all non-goal locations $l \notin B$, and fixing arbitrary decisions for the goal locations (independent of the time passed) provides the sought measurable deterministic time-dependent positional scheduler that dominates all history dependent randomised time-dependent schedulers. $\qquad\square$

Theorem 6 $\max\limits_{\mathcal{S} \in TTP} Pr_\mathcal{S}^{\mathcal{M}}(t) = \sup\limits_{\mathcal{S} \in TTH} Pr_\mathcal{S}^{\mathcal{M}}(t)$, *and randomisation does not improve the result.*

4.2 Total Time Schedulers

In this section we describe the small differences that occur when we allow for schedulers that have the capability to revoke their decisions.

Probability Space. As a first adjustment, we have to build a probability space that covers this generalisation. Such spaces are not hard to build (cf. [9, 10] for locally uniform CTMDPs): We can simply define measures for simple types of these schedulers, and complete the measure space in the usual way.

That is, we start with defining the probability measure for sets of paths

$$l_0 \xrightarrow{a_0,t_0} l_1 \xrightarrow{a_1,t_1} \cdots \xrightarrow{a_{n-1},t_{n-1}} l_n$$

with $0 < t_0 < t_1 < t_2 < \ldots < t_{n-1}$, such that $t_0 \in I_0$, $t_1 \in I_1$, $\ldots$, $t_{n-1} \in I_{n-1}$, for disjoint open intervals $I_0, I_1, \ldots, I_{n-1}$, and schedulers that revoke their decisions in finitely many points $r_1, r_2, \ldots, r_m$, but whose decisions do not depend on the times $t_0, t_1, \ldots, t_{n-1}$.

For such simple sets of paths and schedulers, we can compute the probability to obtain a path in this cylindric set as

$$\int_{t_0 \in I_0, t_1 \in I_1, \ldots, t_{n-1} \in I_{n-1}} \prod_{i=0}^{n-1} \mathbf{R}(l_i, a_i, l_{i+1}) \prod_{i=1}^{m+n} e^{-\mathbf{R}(l_i', a_i', L)(t_i' - t_{i-1}')},$$

where

- $t_0' = 0$,

- $t_1' < t_2' < \ldots < t_{m+n}'$ is the chain of points in time that contains $t_0 < t_1 < t_2 < \ldots < t_{n-1}$ and $r_1, r_2, \ldots, r_m$,

- l_i' is the location the CTMDP is in for the interval (t_{i-1}', t_i'), and

- a_i' is the decision the scheduler would make in the time interval (t_{i-1}', t_i'), which is also the decision it makes at the times t_j of the discrete transitions.

The extension to randomised schedulers is trivial.

These probabilities for cylindric sets then become the basic building blocks of our probability space: As usual, we can build a σ-algebra over these sets, and complete the resulting simple measure space. Note that this definition does not raise the requirement of locally uniform schedulers that was considered necessary previously (cf. [9, 10]), although using locally uniform schedulers admittedly simplifies representing the measure of these cylindric sets of traces to the same integral used in Chapter 2.

Optimal TT Schedulers. Based on the resulting probabilistic space, we argue as in Section 4.1 that we can consider tCTMDPs instead of the standard ones, and that the resulting tCTMDPs remain Markovian. This suggests a proof for the existence of optimal TT schedulers comparable to the prove for time-dependent schedulers that cannot revoke their decisions.

The main difference to the proof in Section 4.1 is that TT schedulers can revoke their decision in any point of time, and the resulting tCTMDP $\mathcal{M}_{t_0}$ is Markovian in any state, rather than only in any discrete entry point. This takes away the discrete flavour from the T scheduler case. Comparable to the case of T schedulers, we know that

○ $Pr_S^{\mathcal{M}_{t_0}}\big((l,t)\big) = 1$ holds for all goal states $l \in B$ and all $t \le t_0$,

○ $Pr_S^{\mathcal{M}_{t_0}}\big((l,t)\big) = 0$ holds for all locations $l \in L$ and all $t > t_0$, and

○ $Pr_S^{\mathcal{M}_{t_0}}\big((l,t_0)\big) = 0$ holds for all non-goal locations $l \notin B$.

holds for every scheduler. For a measurable positional scheduler S, we now have that

$$\dot{Pr}_S^{\mathcal{M}_{t_0}}\big((l,t)\big) = \sum_{l' \in L} \mathbf{R}\big(l, S((l,t)), l'\big) \cdot \Big(Pr_S^{\mathcal{M}_{t_0}}\big((l,t)\big) - Pr_S^{\mathcal{M}_{t_0}}\big((l',t)\big) \Big)$$

for all non-goal locations $l \notin B$, and all $t \in [0,t_0]$ holds true, where $\dot{Pr}_S^{\mathcal{M}_{t_0}}\big((l,t)\big)$ is the derivation of $Pr_S^{\mathcal{M}_{t_0}}\big((l,t)\big)$ to the second argument, that is, to the time.

Naturally, our shift in the way we look at the problem has again no influence on the probability of reaching our objective, and the following equations must hold:

$$\sup_{S \in TTP} Pr_S^{\mathcal{M}}(l, t_0 - t) = \sup_{S \in P} Pr_S^{\mathcal{M}_{t_0}}\big((l,t)\big)$$

and

$$\sup_{S \in TTP} Pr_S^{\mathcal{M}}(t_0) = \sum_{l \in L} \nu(l) \sup_{S \in P} Pr_S^{\mathcal{M}_{t_0}}\big((l,0)\big).$$

The hard part is again to show that an optimal measurable scheduler exists.

Theorem 6 $\max_{S \in TTP} Pr_S^{\mathcal{M}}(t_0) = \sup_{S \in TTH} Pr_S^{\mathcal{M}}(t_0)$, and randomisation does not improve the result.

Proof. The formulas discussed above provide us with simple differential equations, and the functions that we yield for positional schedulers, as well as those that we would get for history dependent schedulers, are clearly dominated by the functions defined by the differential equation

$$\dot{Pr}_S^{\mathcal{M}_{t_0}}\big((l,t)\big) = \min_{a \in Act(l)} \sum_{l' \in L} \mathbf{R}(l, a, l') \cdot \Big(Pr_S^{\mathcal{M}_{t_0}}\big((l,t)\big) - Pr_S^{\mathcal{M}_{t_0}}\big((l',t)\big) \Big)$$

for all $t \in [0,t_0]$.

For an extension to randomised schedulers, the minimum over the actions needs to be replaced by an infimum over the distributions in an intermediate step, but as the infima over affine combinations of a finite set of values takes its minimum in one of these values, the same differential equations defines a dominating function.

Just like in the proof of Theorem 5, the hard part of the proof is to show that there is a measurable scheduler S that always chooses an action a that minimises this value. This guarantees $\sum_{l \in L} \nu(l) Pr_S^{\mathcal{M}}(l,0) = \sup_{S \in TTH} Pr_S^{\mathcal{M}}(t_0)$. We can construct

such a scheduler similarly to the construction of an optimal scheduler in the proof of Theorem 5.

To construct the scheduler decisions for a location l for a measurable scheduler $\mathcal{S}$, we disintegrate $[0,t_0]$ into measurable sets $\{T_a \mid a \in Act(l)\}$, such that $\mathcal{S}$ only makes decisions that minimise $\sum_{l' \in L} \mathbf{R}(l,a,l') \cdot \left(Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}((l,t)) - Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}((l',t)) \right)$. (For positions outside of $[0,t_0]$, that is, for times behind the time bound t_0, the behaviour of the scheduler does not matter and $\mathcal{S}(l,t)$ can be fixed to any constant decision $a \in Act(l)$ for all $t > t_0$.)

We start with fixing an arbitrary order $\succ$ on the actions in $Act(l)$, and introduce, for each point $t \in [0,t_0]$, an order $\succeq_t$ on the actions determined by the value of $\sum_{l' \in L} \mathbf{R}(l,a,l') \cdot \left(Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}((l,t)) - Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}((l',t)) \right)$, using $\succ$ as a tie-breaker.

1. For the action a in $Act(l)$ that is maximal with respect to $\succ$, we start by fixing the open set $O_a = [0,t_0]$ of points in time where the scheduler does not make a decision $a' \succ a$ (where open set in this proof refers to sets open in $[0,t_0]$).

2. We then define the set T_a as the points $t \in O_a$ in time, for which the action a is minimal with respect to $\succeq_t$.

   Being minimal with respect to $\succeq_t$ requires the value

   $$\sum_{l' \in L} \mathbf{R}(l,a,l') \cdot \left(Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}((l,t)) - Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}((l',t)) \right)$$

   to be strictly smaller for a compared to the respective value of all other actions $a' \prec a$, which implies that this sum is also strictly smaller for all t' in some ε-environment of t in $[0,t_0]$.

   As O_a is open, such an ε-environment is a subset of O_a, which implies that T_a is open.

   We now first fix one such ε-environment $\mathcal{E}_t \subseteq O_a$ for all $t \in T_a$, and then fix an arbitrary sequence $\mathcal{E}'_0, \mathcal{E}'_1, \ldots$ from these open sets such that, for all $i \in \mathbb{N}$, there is no $t \in T_a$ such that $\int_{\mathcal{E}_t \smallsetminus \bigcup_{j<i} \mathcal{E}'_j} d\tau > 2 \int_{\mathcal{E}'_i \smallsetminus \bigcup_{j<i} \mathcal{E}'_j} d\tau$.

   Now, the open set $\mathcal{E}_a = \bigcup_{i \in \mathbb{N}} \mathcal{E}'_i$ is measurable, and since obviously $\lim_{i \to \infty} \int_{\mathcal{E}'_i \smallsetminus \bigcup_{j<i} \mathcal{E}'_j} d\tau = 0$ holds true, $\int_{\mathcal{E}_t \smallsetminus \bigcup_{i \in \mathbb{N}} \mathcal{E}'_i} d\tau = 0$ follows for all $t \in T_a$, which implies $\mathcal{E}_t \subseteq \overline{\mathcal{E}_a}$ (because the measurable open set $\mathcal{E}_a \smallsetminus \overline{\mathcal{E}_a}$ is a 0 set, and hence empty).

   Since $\mathcal{E}_a$ is a countable union of open intervals, $|\overline{\mathcal{E}_a} \smallsetminus \mathcal{E}_a| \leq |\mathbb{N}|$ is countable, and hence a 0 set. Furthermore, we have shown $\mathcal{E}_a \subseteq T_a \subseteq \overline{\mathcal{E}_a}$ holds true, and hence that $\overline{T_a} = \overline{\mathcal{E}_a}$ is measurable.

3. Moreover, for the points of time in the fringe of T_a, the continuity of $\sum_{l' \in L} \mathbf{R}(l,a',l') \cdot \left(Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}((l,t)) - Pr_{\mathcal{S}}^{\mathcal{M}_{t_0}}((l',t)) \right)$ for every action $a' \in Act(l)$ guarantees that a still minimises this value, although not necessarily strictly. Hence, we can fix $\mathcal{S}(l,t) = a$ for all $t \in \overline{T_a} \cap O_a$.

4. If there is a next smaller (with respect to $\succ$) action $a' = \max\{a'' \prec a\}$, we fix the new open set $O_{a'} = O_a \smallsetminus \overline{T_a}$ for a', and proceed with step 2.

This way, we can construct a *measurable* scheduler that provides the optimal solution. $\qquad\square$

Bibliography

[1] Rajeev Alur and David L. Dill. A Theory of Timed Automata. *Theoretical Computer Science*, 126(2):183–235, 1994.

[2] Adnan Aziz, Kumud Sanwal, Vigyan Singhal, and Robert Brayton. Model-checking continuous-time Markov chains. *Transactions on Computational Logic*, 1(1):162–170, 2000.

[3] Christel Baier, Holger Hermanns, Joost-Pieter Katoen, and Boudewijn R. Haverkort. Efficient computation of time-bounded reachability probabilities in uniform continuous-time Markov decision processes. *Theoretical Computer Science*, 345(1):2–26, 2005.

[4] J. Bruno, P. Downey, and G. N. Frederickson. Sequencing Tasks with Exponential Service Times to Minimize the Expected Flow Time or Makespan. *Journal of the ACM*, 28(1):100–113, 1981.

[5] Eugene A. Feinberg. Continuous Time Discounted Jump Markov Decision Processes: A Discrete-Event Approach. *Mathematics of Operations Research*, 29(3):492–524, 2004.

[6] H. Hermanns. *Interactive Markov Chains and the Quest for Quantified Quality. LNCS 2428.* Springer-Verlag, 2002.

[7] Vidyadhar G. Kulkarni. *Modeling and Analysis of Stochastic Systems*. Chapman & Hall, Ltd., London, UK, 1995.

[8] M. A. Marsan, G. Balbo, G. Conte, S. Donatelli, and G. Franceschinis. Modelling with Generalized Stochastic Petri Nets. *SIGMETRICS Performance Evaluation Review*, 26(2):2, 1998.

[9] Martin R. Neuhäußer, Mariëlle Stoelinga, and Joost-Pieter Katoen. Delayed Nondeterminism in Continuous-Time Markov Decision Processes. In *Proceedings of FOSSACS '09*, pages 364–379, 2009.

[10] Martin R. Neuhäußer and Lijun Zhang. Time-Bounded Reachability in Continuous-Time Markov Decision Processes. Technical report, 2009.

[11] Martin L. Puterman. *Markov Decision Processes: Discrete Stochastic Dynamic Programming*. Wiley-Interscience, April 1994.

[12] Markus Rabe and Sven Schewe. Optimal Schedulers for Time-Bounded Reachability in CTMDPs. Reports of SFB/TR 14 AVACS 55, SFB/TR 14 AVACS, October 2009. http://www.avacs.org.

[13] William H. Sanders and John F. Meyer. Reduced Base Model Construction Methods for Stochastic Activity Networks. In *Proceedings of PNPM'89*, pages 74–84, 1989.

[14] Linn I. Sennott. *Stochastic Dynamic Programming and the Control of Queueing Systems*. Wiley-Interscience, 1999.

[15] Nicolás Wolovick and Sven Johr. A Characterization of Meaningful Schedulers for Continuous-Time Markov Decision Processes. In *Proceedings of FORMATS'06*, pages 352–367, 2006.

[16] L. Zhang, H. Hermanns, E. M. Hahn, and B. Wachter. Time-bounded model checking of infinite-state continuous-time Markov chains. In *Proceedings of ACSD'08*, pages 98–107, 2008.

Appendices

Omitted Proofs

Greedy Schedulers

The proof of the existence of a positional greedy scheduler in Section 3.1 is not constructive, because we do not provide means to compute a constant k that satisfies $d_l \neq d_{l,a} \Rightarrow \exists k' \leq k.\ d_l[k] > d_{l,a}[k]$. Moreover, suitable constants μ and $n_{\mathcal{M}}$ can only be computed once k is known. Without being able to provide a suitable constant, we could therefore provide an algorithm that converges to optimal time-abstract schedulers, but we would not be able to determine whether an optimal solution has already been reached.

The techniques we exploit in this appendix to show that $k < |L|$ is in fact smaller than the uniform CTMDP $\mathcal{M}$ itself draw from linear algebra, and are, while simple, a bit unusual in this context. We first turn to the simpler notion of Markov chains, resolving the nondeterminism in accordance with the positional standard greedy scheduler $\mathcal{S}$ whose existence was shown in Section 3.1.

We first lift the step probability vector from locations to distributions, where $d_v = \sum_{l \in L} v(l) d_l$ is, for a distribution $v : L \to [0,1]$, the normal affine combination of the step probability vectors of the individual locations. We call two distributions $v, v' : L \to [0,1]$ equivalent, if their step probability vectors $d_v = d_{v'}$ are equal, and i-step equivalent if they are equal up to position i ($\forall j \leq i.\ d_v[j] = d_{v'}[j]$). We immediately extend these definitions to arbitrary vectors $v : L \to \mathbb{R}$.

We then define the vector spaces D_i of multitudes of differences $v - v'$ of i-step equivalent distributions $v, v' : L \to [0,1]$. That is, the differences between i-step equivalent distributions form a spanning set of D_i.

As a basis for our construction, D_0 is simple to construct: it is the vector space that contains the multitudes of differences $v - v'$ of distributions $v, v' : L \to [0,1]$ that are equally likely in the goal region. The first implies the restriction $\sum_{l \in L} \delta(l) = 0$, and the latter the restriction $\sum_{l \in B} \delta(l) = 0$ for all $\delta \in D_0$. On the other hand, all vectors that satisfy these side conditions are clearly multitudes of differences $v - v'$ of 0-step equivalent distributions $v, v' : L \to [0,1]$, such that D_0 is completely described by this description.

Once we have computed the vector space D_i, we can compute the vector space O_i that contains a vector δ if it is a multitude of differences $v - v'$ of distributions $v, v' : L \to [0,1]$, such that $shift(d_v)$ and $shift(d_{v'})$ are i-step equivalent. The transition

from step probability vectors to the *shift* of them is a simple linear operation, which essentially transforms the distributions according to the transition matrix, but drops all weight that is already accumulated in the goal region. Hence, we can obtain O_i from D_i by a simple linear transformation of the vector space.

Now, D_{i+1} obviously consists of those vectors in D_i that are also in O_i, and $D_{i+1} = D_i \cap O_i$ can be obtained by an intersection of two vector spaces.

Now $D_0 \supseteq D_1 \supseteq \ldots \supseteq D_{|L|-2} \supseteq \ldots$ obviously holds, and $D_i = D_{i+1}$ implies $O_i = O_{i+1}$, and hence $D_i = D_\infty$. As D_0 is a $|L| - 2$ dimensional vector space, and inequality implies the loss of at least one dimension, a fixed point is reached after at most $|L| - 2$ steps. That is, two distributions are equivalent, if, and only if, they are $(|L| - 2)$-step equivalent.

Having established this, we apply it on the distribution $v_{l,a}$ obtained in one step from a position $l \notin B$ when choosing the action a, as compared to the distribution v_l obtained when choosing the action $\mathcal{S}(l)$ defined by the positional greedy scheduler.

Now, $d_l > d_{l,a}$ holds if, and only if, $shift(d_l) = d_{v_l} > d_{v_{l,a}} = shift(d_{l,a})$, which implies $d_{v_l}[k'] > d_{v_{l,a}}[k']$ for some $k' \leq |L| - 2$, and hence $d_l[k] > d_{l,a}[k]$ for some $k < |L|$.