Multiple Objectives and Cost Bounds in MDP

Arnd Hartmanns, Sebastian Junges, Joost-Pieter Katoen, Tim Quatmann

Markov decision processes [15] (MDPs) with *rewards* or *costs* are a popular model to describe planning problems under uncertainty. Planning algorithms typically find strategies that minimise the expected costs to fulfil a task. However, many scenarios call for minimising the probability to run out of resources before reaching the goal: while it is *beneficial* for a plane to reach its destination with low *expected* fuel consumption, it is *essential* to reach its destination with the *fixed* available amount of fuel. Strategies that optimise solely for the probability to fulfil a task are often very expensive. Decision makers have to trade the success probability against the costs. Different types of costs such as time, energy, money, or capacity have to be considered at once [2], [5], [13], [16]. This makes many planning problems inherently multi-objective [3], [5]. In this extended abstract, we summarise our results on practically efficient multi-objective model checking methods for multiple cost bounds implemented in the STORM tool [7] and presented at TACAS 2018 [11].

Related work: The analysis of single-objective cost-bounded reachability in MDPs is an active area of research in both the AI and formal method communities. We build on [10], where three different practically efficient cost-bounded model checking approaches are explored and compared, in particular including the sequential epoch analysis of [12]. For multi-objective analysis, the model checking community focuses on probabilities and expected costs [4], [8], and our implementation uses the value iteration-based approach of [9].

Problem Statement: Given an MDP M with multiple cost structures over non-negative costs and $\ell \geq 2$ objectives of the form "maximise the probability to reach a state in the goal set G_i such that the cumulative cost for the j_i-th cost structure is at most b_i" (written $\mathcal{P}^{\max}(\langle j_i \rangle_{\leq b_i} G_i)$), obtain the Pareto-optimal points to compute their probabilities and visualise the underlying trade-offs between the objectives.

Approach: A naive approach for this problem is to first unfold M by encoding the incurred costs in the state space [1], [14], resulting in the unfolded model M_{unf}, and then apply existing multi-objective model checking algorithms (e.g. [9]) for *unbounded* reachability probabilities on M_{unf}. Figure 1 illustrates this approach for $\ell = 2$ objectives with cost bounds $b_0 = 1$ and $b_1 = 2$. Each so-called *epoch model* $M^{x,y}$ reflects a copy of the given MDP M for cost *epoch* $(x, y) \in \{\perp, 0, \ldots, b_0\} \times \{\perp, 0, \ldots, b_1\}$. The epochs encode the costs that can still be incurred to satisfy the objectives. Transitions that incur non-zero costs are redirected to a successive epoch. If the i-th component of the current epoch is $\perp$, the corresponding cost bound has already been exceeded. For the i-th objective we consider the probability to eventually reach a goal state in an epoch where the i-th entry is not $\perp$.

Our approach [11] avoids explicitly considering the potentially large unfolding of M. We analyse the epoch models one after the other: for $M^{\boldsymbol{e}}$, only the results for epoch models $M^{\boldsymbol{e}'}$ with $\boldsymbol{e}' \leq \boldsymbol{e}$ are relevant. This naturally extends approaches for the single-objective case [10], [12]. The similarity of the epoch models $M^{x,y}$ enables an efficient implementation. We embed this approach into the value-iteration based multi-objective model checking framework of [9].

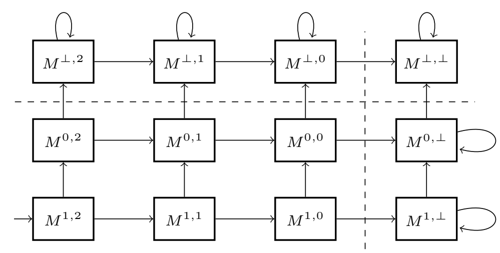


Fig. 1. Illustration of unfolding M_{unf}

Besides the objectives mentioned above, the approach also supports lower cost bounds, minimal probabilities, conjunctions of cost bounds (e.g. "maximise the probability to reach the target within the given time *and* energy limits"), and cost-bounded expected costs (e.g. "minimise the expected energy consumption within the first minute").

Visualisations: Users need to consider the tradeoffs between objectives and make informed decisions for system design. The results of multi-objective model checking are typically presented as a Pareto curve as in Fig. 2. However, Pareto set visualisations alone may not provide sufficient information, about, e.g., which objectives are aligned or conflicting. Cost bounds add an extra dimension for each cost structure, making the underlying tradeoffs significantly harder to discover and understand. However, for each Pareto-optimal scheduler, our method has implicitly computed the probabilities of the objectives for all reachable epochs as well, i.e. for all cost values below the bounds b_i required in the query. We visualise this information for better insights into the behaviour of each

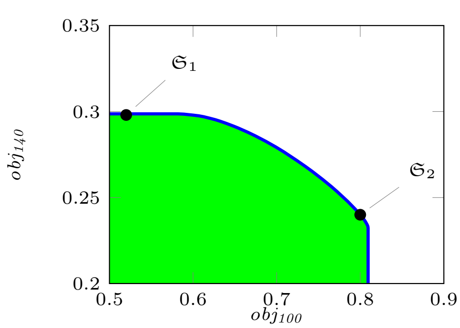


Fig. 2. Pareto curve for $multi(obj_{100}, obj_{140})$

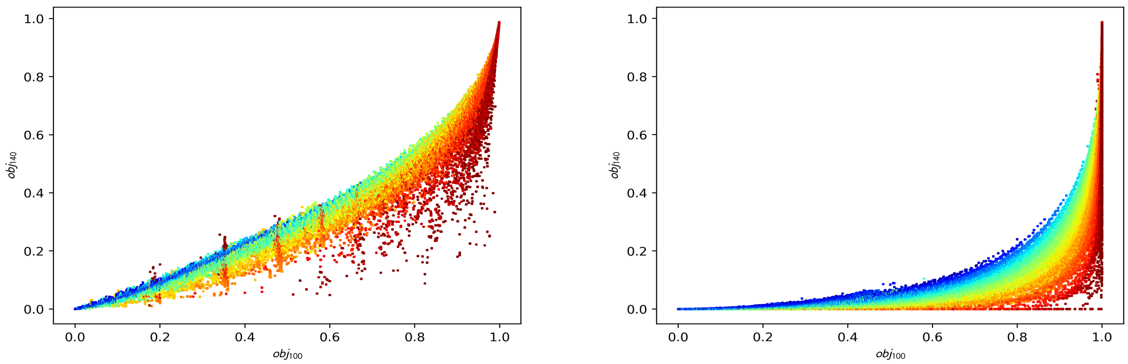
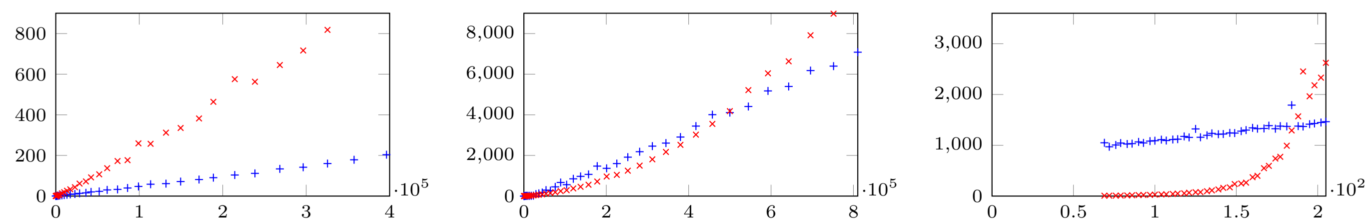


Fig. 3. Remaining cost budget (red is high, blue is low) vs. the probabilities of the two objectives



(a) Rover (16 states)

(b) Resources (94 states)

(c) Wlan6 ($5 \cdot 10^6$ states)

Fig. 4. Runtime (y-axis) of SEQ (+) and UNF (×) for increasing cost bounds (x-axis)

scheduler, its robustness w.r.t. the bounds, and its preferences for certain objectives depending on the remaining budget for each quantity. In Fig. 3, we show one such heatmap-style visualisation, comparing the two Pareto-optimal schedulers $\mathfrak{S}_1$ and $\mathfrak{S}_2$ from Fig. 2. The probabilities of two objectives using the same cost structure are mapped to the x and y axes while the colour indicates the remaining cost value. Comparing the two schedulers, we see that the left one achieves higher goal probabilities for medium remaining budget, while the right one is safer overall at the cost of performance.

Experiments: The discussed approach is implemented in STORM [7] v1.2, and available via [6]. Comparing the naive unfolding approach (UNF) with the sequential approach (SEQ) on different models, we observe that SEQ is less sensitive to increases in the magnitude of the cost bounds, as illustrated in Fig. 4. We refer to [11] for more details on the experiments.

REFERENCES

[1] Andova, S., Hermanns, H., Katoen, J.P.: Discrete-time rewards model-checked. In: FORMATS. LNCS, vol. 2791, pp. 88–104. Springer (2003)
[2] Bresina, J.L., Jónsson, A.K., Morris, P.H., Rajan, K.: Activity planning for the Mars exploration rovers. In: ICAPS. pp. 40–49. AAAI (2005)
[3] Bryce, D., Cushing, W., Kambhampati, S.: Probabilistic planning is multi-objective. Tech. rep., Arizona State Univ., CSE (2007)
[4] Chatterjee, K., Majumdar, R., Henzinger, T.A.: Markov decision processes with multiple objectives. In: STACS. LNCS, vol. 3884, pp. 325–336. Springer (2006)
[5] Cheng, L., Subrahmanian, E., Westerberg, A.W.: Multiobjective decision processes under uncertainty: applications, problem formulations, and solution strategies. Industrial & Engineering Chemistry Research 44(8), 2405–2415 (2005)
[6] Dehnert, C., Junges, S., Katoen, J.P., Quatmann, T., Volk, M.: Storm source files. zenodo (2018), https://doi.org/10.5281/zenodo.1181896
[7] Dehnert, C., Junges, S., Katoen, J.P., Volk, M.: A Storm is coming: A modern probabilistic model checker. In: CAV (2). LNCS, vol. 10427. Springer (2017)
[8] Etessami, K., Kwiatkowska, M., Vardi, M.Y., Yannakakis, M.: Multi-objective model checking of Markov decision processes. LMCS 4(4) (2008)
[9] Forejt, V., Kwiatkowska, M., Parker, D.: Pareto curves for probabilistic model checking. In: ATVA. LNCS, vol. 7561, pp. 317–332. Springer (2012)
[10] Hahn, E.M., Hartmanns, A.: A comparison of time- and reward-bounded probabilistic model checking techniques. In: SETTA. LNCS, vol. 9984, pp. 85–100 (2016)
[11] Hartmanns, A., Junges, S., Katoen, J.P., Quatmann, T.: Multi-cost bounded reachability in MDP. In: TACAS (2). LNCS, vol. 10806, pp. 320–339. Springer (2018)
[12] Klein, J., Baier, C., Chrszon, P., Daum, M., Dubslaff, C., Klüppelholz, S., Märcker, S., Müller, D.: Advances in probabilistic model checking with PRISM: variable reordering, quantiles and weak deterministic Büchi automata. STTT pp. 1–16 (2017)
[13] Lacerda, B., Parker, D., Hawes, N.: Multi-objective policy generation for mobile robots under probabilistic time-bounded guarantees. In: ICAPS. pp. 504–512. AAAI Press (2017)
[14] Laroussinie, F., Sproston, J.: Model checking durational probabilistic systems. In: FoSSaCS. LNCS, vol. 3441, pp. 140–154. Springer (2005)
[15] Puterman, M.L.: Markov Decision Processes. John Wiley and Sons (1994)
[16] Roijers, D.M., Vamplew, P., Whiteson, S., Dazeley, R.: A survey of multi-objective sequential decision-making. J. Artif. Intell. Res. 48, 67–113 (2013)