On a Notion of Resilience for Markov Decision Processes with Reachability Objectives ⋆

Xiaoming Duan * **Nasim Baharisangari** ** **Rui Yan** ***
Zhe Xu ** **Melkior Ornik** ****

* *Department of Automation, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: xiaomingduan.zju@gmail.com)*
** *School for Engineering of Matter, Transport, and Energy, Arizona State University, Tempe, AZ 85287, USA (e-mail: {nbaharis, xzhe1}@asu.edu)*
*** *Department of Computer Science, University of Oxford, Oxford OX1 3QD, UK (e-mail: rui.yan@cs.ox.ac.uk.)*
**** *Department of Aerospace Engineering and the Coordinated Science Laboratory, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA (e-mail: mornik@illinois.edu)*

Abstract: We propose and study a notion of resilience for Markov decision processes (MDPs) with the almost-sure reachability objective to action losses. Given an MDP with an initial state and a set of target states, we define the resilience degree of the MDP as the minimum number of actions that must be removed so that the target states cannot be reached almost surely from the initial state. This notion measures the level of tolerance of an MDP to action losses under the reachability objective. We first preprocess the MDP to remove irrelevant states and actions and construct a reduced transition diagram. Then, we show that computing the resilience degree is an NP-hard problem and provide an exact solution based on the mixed-integer linear programming.

Keywords: Markov decision processes, decision making, reachability analysis, resilience

1. INTRODUCTION

Markov decision processes (MDPs) are a widely adopted modeling formalism for sequential decision making under uncertainties (Puterman, 2014). They find applications in robotic planning (Antonyshyn et al., 2022), portfolio optimization (Bäuerle and Rieder, 2011), probabilistic verification (Forejt et al., 2011) and so on. Reachability analysis is a classic problem for MDPs where the goal is to synthesize policies so as to optimize the probability of reaching a set of targets (de Alfaro, 1999). In this work, we propose and study a notion of resilience to action failures for MDPs under the almost-sure reachability objective. This notion is related to structural properties of MDPs and quantifies the degree to which they can withstand action losses while still achieving objectives.

Our problem formulation is motivated by the work (Bouvier and Ornik, 2022) and (Bucić et al., 2018). In the former, the resilient control for linear systems is considered, and the authors synthesize control laws in the presence of lost control over some actuators. The latter formulates a design problem for the space of control signals so that systems can always stay safe despite the existence of failed actuators. Since the action failures can be interpreted as adversarial attacks, our problem also has connections with reachability games (Chatterjee and Henzinger, 2012) and sabotage games (Löding and Rohde, 2003; van Benthem, 2005). In zero-sum reachability games, two players select actions at each state (by turns or concurrently) so as to maximize/minimize the probability that the state transitions into a set of target states. In these games, the actions available to the players are fixed a priori. In a sabotage game, one player tries to reach the targets by selecting actions at each state while the other obstructs the transitioning process by removing available actions at states. The game is usually turn-based and played on graphs without probabilistic transitions. In terms of model uncertainties, robust control of MDPs under uncertain transition models (Nilim and Ghaoui, 2005; Wiesemann et al., 2013) or rewards (Xu and Mannor, 2009) has been studied extensively in the literature. However, none of these work investigates consequences of changes in the action space. Yet another closely related line of work focuses on network survivability problems (Coudert et al., 2007; Kuipers, 2012; Coudert et al., 2016) where cuts with the minimum number of (labeled) edges between source nodes and destination nodes in a graph are computed. These problems are usually formulated as label/color cut problems and are computationally hard (Zhang et al., 2011). Our problem is different because breaking probabilistic reachability does not require a cut in the transition graph.

⋆ The work of X. Duan is sponsored by Shanghai Pujiang Program under grant 22PJ1404900.

Our objective in this paper is to propose and study a notion of resilience for MDPs to action losses under the almost-sure reachability objective. The main contributions are as follows. We first propose to use the minimum number of actions that must be disabled so that the reachability objective cannot be achieved as the degree of resilience for MDPs. Then, we preprocess the MDPs for computing such a resilience degree by removing irrelevant states and actions. We further prove that the optimization problem is NP-hard and provide an exact solution method based on the mixed-integer linear programming.

The rest of the paper is organized as follows. We review the basics of MDP reachability problems and formulate the problem of interest in Section 2. The main results are presented in Section 3. We provide a numerical example in Section 4. Finally, Section 5 concludes the paper.

1.1 Notation

Let $\mathbb{R}$ and $\mathbb{R}^n$ be the set of real numbers and n-dimensional real vectors, respectively. We use $\mathbb{N}_0$ to denote the set of nonnegative integers. For a finite set B, its cardinality is denoted by $|B|$. The support of a probability mass function $f : X \to [0, 1]$ is $\mathrm{Supp}(f) = \{x \in X \mid f(x) > 0\}$.

2. PRELIMINARIES AND PROBLEM OF INTEREST

2.1 Markov decision processes and almost-sure reachability

A Markov decision process (MDP) $\mathcal{M}$ is a tuple $\mathcal{M} = (S, \mathcal{A}, P, s_{\text{init}})$ where

(1) S is a finite set of states;
(2) $\mathcal{A} = \cup_{s\in S} A(s)$ is a finite set of actions and $A(s)$ consists of actions available at state s;
(3) $P : S \times \mathcal{A} \times S \to [0, 1]$ is the transition kernel that satisfies $\sum_{s'\in S} P(s' \mid s, a) = 1$ for $s \in S$ and $a \in A(s)$;
(4) $s_{\text{init}} \in S$ is the initial state.

A *policy* for an MDP is a sequence of decision rules that resolve the action selection at each state. A *Markovian* policy $\pi = (d_0, d_1, \cdots)$ for an MDP $\mathcal{M}$ is a sequence of distribution functions where $d_t : S \times \mathcal{A} \to [0, 1]$ and $\sum_{a\in A(s)} d_t(a \mid s) = 1$ for all $s \in S$ and $t \geq 0$. A Markovian policy π is *deterministic* and *stationary* if all the distribution functions in the sequence are the same and their support consists of a single action, i.e., $\pi = (d, d, \cdots)$ and $d(a \mid s) = 1$ for all $s \in \mathcal{S}$ and some $a \in A(s)$.

An instance $\Lambda = (S, \mathcal{A}, P, s_{\text{init}}, \mathcal{T})$ of the reachability problem consists of an MDP $\mathcal{M} = (S, \mathcal{A}, P, s_{\text{init}})$ and a set of target states $\mathcal{T} \subset S$ in the state space of $\mathcal{M}$, and one seeks a policy π so that the targets $\mathcal{T}$ can be reached from the initial state s_{init} with the maximum probability. It is well known that there always exist deterministic and stationary optimal policies π for a reachability problem (Baier and Katoen, 2008, Lemma 10.102). In this paper, we focus on the almost-sure reachability objective where the maximum probability of reaching $\mathcal{T}$ in Λ is 1 and adopt the following assumption.

Assumption 1. (Almost-sure reachability). There exists a policy for $\Lambda = (S, \mathcal{A}, P, s_{\text{init}}, \mathcal{T})$ so that the target set $\mathcal{T}$ can be reached with probability (w.p.) 1.

2.2 Problem of interest

Let Λ be a reachability problem that satisfies Assumption 1. We are interested in analyzing a resilience property of the MDP $\mathcal{M}$ with respect to action failures.

Definition 2. (Resilience degree to action failures). Let Λ be a reachability problem that satisfies Assumption 1. The MDP $\mathcal{M}$ is p-resilient to action failures if p is the solution to the following optimization problem

$$\begin{aligned} &\underset{q\in\mathbb{N}_0}{\text{minimize}} \quad q \\ &\text{subject to} \quad \underset{\mathcal{A}'\subset\mathcal{A},|\mathcal{A}'|=q}{\text{minimize}} \ \Pr^{\max}_{\Lambda'}(\text{Reach}(\mathcal{T})) < 1, \end{aligned} \tag{1}$$

where $\Pr^{\max}_{\Lambda'}(\text{Reach}(\mathcal{T}))$ is the maximum probability of reaching $\mathcal{T}$ in a modified reachability instance $\Lambda' = (S, \mathcal{A} \setminus \mathcal{A}', P, s_{\text{init}}, \mathcal{T})$ [1].

In words, given a reachability instance Λ satisfying Assumption 1, Definition 2 asks for a minimum number p so that when p actions are removed from the action set, the target states can no longer be reached almost surely. The resilience degree p is a quantitative measure of how resilient an MDP is to action losses under almost-sure reachability objectives. Specifically, in a p-resilient MDP, the target states can still be reached w.p. 1 when arbitrary $p-1$ actions fail. Note that if Λ does not satisfy Assumption 1, then it is 0-resilient immediately. In the rest of the paper, we are interested in computing the resilience degree of a given MDP. We illustrate Definition 2 and reveal one nontrivial aspect of problem (1) in the following example.

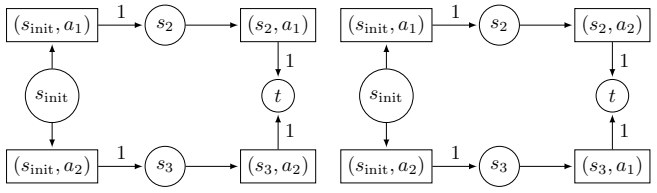
Example 3. (Coupling among actions). The transition diagrams of two MDPs with the same state space $S = \{s_{\text{init}}, s_2, s_3, t\}$ and action space $\mathcal{A} = \{a_1, a_2\}$ are shown in Fig. 1, where the circles represent states and rectangles represent actions. Directed edges exist between a state s and its associated actions $A(s)$ and between actions and possible successor states. The numbers on the edges indicate transition probabilities. In Fig. 1, although the two MDPs have exactly the same transition structure, their resilience degrees are different. This difference is caused by the *action coupling* at different states, i.e., it is possible that $A(s) \cap A(s') \neq \emptyset$ for some $s, s' \in S$, and removing an action in the intersection disables the action from both states. In Fig. 1(a), the target state t is still reachable w.p. 1 after removing either action a_1 or a_2. In contrast, removing either a_1 or a_2 leads to 0 probability of reaching for the MDP in Fig. 1(b).

3. MAIN RESULTS

3.1 MDP preprocessing and smaller problem instance

We first note that in problem (1), for an $\mathcal{A}' \subset \mathcal{A}$ with $|\mathcal{A}'| = q$ and $\Pr^{\max}_{\Lambda'}(\text{Reach}(\mathcal{T})) < 1$, if $\mathcal{A}'$ consists of actions that will never be used in achieving almost-sure reachability in Λ, then q cannot be optimal since removing those actions from $\mathcal{A}'$ makes q smaller. Therefore, we can

[1] We also need to modify the transition kernel P when there is a change in the action space $\mathcal{A}$ (removing all transition probabilities associated with the removed actions). However, to simply notation and emphasize the change of the action space, we write Λ' as is with the understanding that P changes accordingly.



(a) A 2-resilient MDP (b) A 1-resilient MDP

Fig. 1. The transition diagrams of two MDPs with different resilience degrees.

preprocess the MDP $\mathcal{M}$ by removing actions that are never used in the optimal policy for almost-sure reachability from s_{init}. This preprocessing step leads to a potentially smaller problem instance and helps with the computation of the resilience degree.

Given a reachability instance Λ, computing the set of states starting from which the target states can be reached almost surely is a well-studied problem, and there exist efficient algorithms for it (Baier and Katoen, 2008, Algorithm 45) (Forejt et al., 2011, Algorithm 4). Let $R \subset S$ be the set of such states, which contains the target set $\mathcal{T}$. Then, by Assumption 1, we have that $s_{\text{init}} \in R$. Moreover, the set of states $S \setminus R$ can never be reached starting from the initial state x_{init} under the optimal policy for almost-sure reachability. Therefore, we can safely discard all the states $S \setminus R$ and their associated actions. On the other hand, since we only care about reaching the set of targets from s_{init}, those states in R that cannot be reached from s_{init} and their actions can also be removed.

In terms of the actions, at each state $s \in R$, let the set of actions $A^{\max}(s)$ be

$$A^{\max}(s) = \{a \in A(s) \mid \text{Supp}(P(\cdot \mid s, a)) \subset R\}.$$

Then, by the Bellman optimality criterion (Baier and Katoen, 2008, Theorem 10.100), if an action $a \in A(s)$ is optimal at a state $s \in R$ for almost-sure reachability, then it must be that $a \in A^{\max}(s)$ (the converse is not necessarily true), i.e., any action $a \in A(s) \setminus A^{\max}(s)$ will never be selected. Let $\mathcal{A}^{\max} = \cup_{s \in S} A^{\max}(s)$. Then, we only need to consider the set of actions $\mathcal{A}^{\max}$, and all actions in $\mathcal{A} \setminus \mathcal{A}^{\max}$ can be safely discarded without affecting the solution to problem (1).

In summary, the preprocessing, which can be performed efficiently, consists of the following two steps

(1) remove all states $S \setminus R$, and states in R that are not reachable from s_{init} and their associated actions;
(2) remove all actions in $\mathcal{A} \setminus \mathcal{A}^{\max}$.

We note that after the preprocessing, the reduced MDP only contains states whose maximum probability of reaching the targets is 1. Then we can solve (1) on the reachability instance with potentially smaller state and action spaces. Without loss of generality, we assume all the MDPs in the rest of the paper have already been preprocessed.

3.2 NP-hardness from minimum hitting set problem

In this subsection, we provide a complexity result for problem (1). Specifically, we show that problem (1) is NP-hard by a reduction from the minimum hitting set problem (Garey and Johnson, 1979, SP8).

Definition 4. (Minimum hitting set problem). Given a collection C of subsets of a finite set E and a positive integer $K \leq |E|$, decide whether there exists a subset $E' \subset E$ with $|E'| \leq K$ such that E' contains at least one element from each subset in C.

It is known that the minimum hitting set problem is NP-complete. In the next theorem, we show that our problem can be reduced from the minimum hitting set problem.

Theorem 5. (NP-hardness of (1)). For a reachability problem $\Lambda = (S, \mathcal{A}, P, s_{\text{init}}, \mathcal{T})$ and an integer $K \leq |\mathcal{A}|$, deciding whether there exists a subset $\mathcal{A}' \subset \mathcal{A}$ with $\mathcal{A}' \leq K$ such that $\Pr^{\max}_{\Lambda'}(\text{Reach}(\mathcal{T})) < 1$ is NP-complete.

Proof. The decision problem is in NP since for any given $\mathcal{A}'$, whether $\mathcal{T}$ can be reached w.p. 1 in $\Lambda' = (S, \mathcal{A} \setminus \mathcal{A}', P, s_{\text{init}}, \mathcal{T})$ can be verified in polynomial time using the algorithm (Baier and Katoen, 2008, Algorithm 45).

To show the NP-completeness of the decision problem of (1), we perform a reduction from the minimum hitting set problem inspired by the one in (Jha et al., 2002).

Let $C = \{C_1, \cdots, C_m\}$, $E = \{e_1, \cdots, e_n\}$ and an integer K be a minimum hitting set problem where $C_i \subset E$ for $1 \leq i \leq m$. We define a reachability problem Λ as follows. For $1 \leq i \leq m$, we define a set of states S_i for each $C_i \in C$ where $S_i = \{s_{i,1}, \cdots s_{i,|C_i|-1}\}$. If $|C_i| = 1$, then $S_i = \emptyset$. Then, the state space S of Λ is $S = \{s_{\text{init}}, t\} \cup \cup_{i=1}^{m} S_i$ where t is the target state. There are m actions available at s_{init}, i.e., $A_{s_{\text{init}}} = \{a_{0,1}, \cdots, a_{0,m}\}$. For $1 \leq i \leq m$, an action $a_{0,i}$ takes s_{init} to $s_{i,1}$ w.p. 1 if $|C_i| > 1$, and it takes s_{init} to t w.p. 1 if $|C_i| = 1$. At each state $s_{i,j}$ for $1 \leq i \leq m$ and $1 \leq j \leq |C_i| - 1$, there is a single action $a_{i,j}$. The action $a_{i,j}$ takes $s_{i,j}$ to $s_{i,j+1}$ w.p. 1 if $j \leq |C_i| - 2$, and it takes $s_{i,j}$ to t w.p. 1 if $j = |C_i| - 1$. Essentially, there are m parallel paths from s_{init} to t in the transition graph of Λ. It is easy to check that a minimum hitting set with size K exists if and only if there exists a subset $\mathcal{A}' \subset \mathcal{A}$ with $|\mathcal{A}'| \leq K$ such that t cannot be reached from s_{init} w.p. 1 in Λ'. Since the minimum hitting set problem is NP-complete, the decision problem of (1) is also NP-complete, which implies NP-hardness of (1).

An illustration of the reduction in the proof of Theorem 5 is shown in Fig. 2.

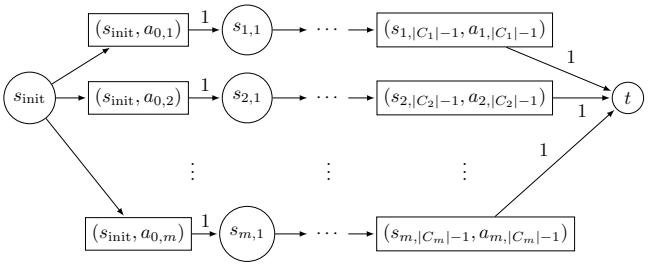


Fig. 2. An illustration of the reduction from the minimum hitting set problem to the resilience degree problem.

Remark 6. (Connections with label cut problems). In (Jha et al., 2002), NP-hardness of now known as the label cut problem (Zhang et al., 2011) is proved by a reduction from the minimum hitting set problem. In a label cut problem, one is given a graph with a source, a destination, and labeled edges, and the goal is to find a cut for the graph

that separates the source from the destination and uses the minimum number of labels (cf. classic minimum cut problem). Our problem is different from label cut problems since we consider probabilistic reachability. For example, in Fig. 3, to disconnect all paths from s_{init} to t, one has to cut both actions a_1 and a_2. However, removing any one of these actions results in reachability probability less than 1 from s_{init}. In fact, one can see that a solution to the label cut serves as an upper bound for the resilience degree.

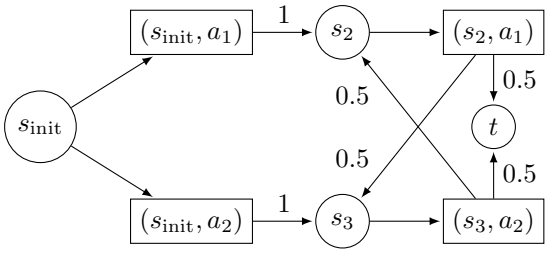


Fig. 3. The transition diagram of an MDP where the label cut number is different from the resilience degree.

3.3 An exact solution based on MILP

In this subsection, we first provide an exact solution to (1) by solving multiple mixed-integer linear programs (MILPs). Then, we develop an algorithm to find the resilience degree of a given MDP.

Consider the following MILP:

$$\underset{x_s \in \mathbb{R}}{\text{minimize}} \quad x_{s_{\text{init}}} \tag{2a}$$

subject to

$$x_s \geq \sum_{s' \in S} P(s' \,|\, s, a) x_{s'} - y_a, \text{ for all } s \in S \setminus \mathcal{T}, a \in A(s), \tag{2b}$$

$$x_s = 1, \text{ for all } s \in \mathcal{T}, \tag{2c}$$

$$x_s \geq 0, \text{ for all } s \in S \setminus \mathcal{T}, \tag{2d}$$

$$y_a \in \{0, 1\}, \text{ for all } a \in \mathcal{A}, \tag{2e}$$

$$\sum_{a \in \mathcal{A}} y_a \leq K, \tag{2f}$$

where $K \in \mathbb{N}_0$ is a positive parameter that represents the maximum number of actions that can be removed. There are a total of $|S|$ continuous variables and $|\mathcal{A}|$ binary variables in (2). The variables y_a for $a \in \mathcal{A}$ indicate whether the corresponding actions are removed from the action set. The constraint (2b) is a modified Bellman optimality condition. When $y_a = 1$, the action a is removed from the available set of actions, and this constraint is inactive since $\sum_{s' \in S} P(s' \,|\, s, a) x_{s'} \leq 1$ and $x_s \geq 0$. Constraints (2c) and (2d) are standard constraints in the LP formulation for computing the maximum reachability probabilities (Courcoubetis and Yannakakis, 1998, Proposition 3.2). Note that when $K = 0$, we indeed recover the standard LP. In the next theorem, we show that solving (2) is equivalent to solving the decision problem for (1).

Theorem 7. (Equivalent MILP formulation for (1)). For a reachability problem $\Lambda = (S, \mathcal{A}, P, s_{\text{init}}, \mathcal{T})$ that satisfies Assumption 1, there exists a subset $\mathcal{A}' \subset \mathcal{A}$ with $|\mathcal{A}'| \leq K$ such that $\Pr^{\max}_{\Lambda'}(\text{Reach}(\mathcal{T})) < 1$ if and only if the solution to (2) satisfies $x_{s_{\text{init}}} < 1$ with parameter K.

Proof. Suppose there exists a subset $\mathcal{A}' \subset \mathcal{A}$ with $|\mathcal{A}'| \leq K$ such that $\Pr^{\max}_{\Lambda'}(\text{Reach}(\mathcal{T})) < 1$. We let $y_a = 1$ for $a \in \mathcal{A}'$ and $y_a = 0$ otherwise. Under this specific setup, (2) is essentially computing the maximum reachability probabilities for a modified instance $(S, \mathcal{A} \setminus \mathcal{A}', P, s_{\text{init}}, \mathcal{T})$. To see this, for $y_a = 1$ (action a is removed), the constraint (2b) becomes redundant due to the existence of the constraint $x_s \geq 0$ and the fact that

$$\sum_{s' \in S} P(s' \,|\, s, a) x_{s'} - y_a \leq \sum_{s' \in S} P(s' \,|\, s, a) - y_a = 0,$$

where we used that $x_s \leq 1$. Since $\Pr^{\max}_{\Lambda'}(\text{Reach}(\mathcal{T})) < 1$, the solution to (2) satisfies $x_{s_{\text{init}}} < 1$ when we fix $y_a = 1$ for $a \in \mathcal{A}'$. Thus, the optimal solution to (2) must also satisfy $x_{s_{\text{init}}} < 1$.

For the other direction, suppose for some parameter K, the solution to (2) satisfies $x_{s_{\text{init}}} < 1$. We construct $\mathcal{A}' = \{a \in \mathcal{A} \,|\, y_a = 1\}$. By the argument above, we know that (2) is computing the maximum reachability probabilities for $(S, \mathcal{A} \setminus \mathcal{A}', P, s_{\text{init}}, \mathcal{T})$. Thus, we have $\Pr^{\max}_{\Lambda'}(\text{Reach}(\mathcal{T})) < 1$.

By Theorem 7, if we solve (2) multiple times with an increasing sequence of K's and stop whenever the solution to (2) satisfies $x_{s_{\text{init}}} < 1$, then we obtain the optimal solution to (1), i.e., the resilience degree has been found. To figure out the number of times (2) needs to be solved, we notice that both the number of actions $|A(s_{\text{init}})|$ available at s_{init} and $|\{a \in A(s) \,|\, s \in S, \sum_{t \in \mathcal{T}} P(t \,|\, s, a) > 0\}|$ are obvious upper bounds on K. That is because removing either set of these actions eliminates all paths from s_{init} to $\mathcal{T}$ in the transition diagram. We let

$$\bar{K} = \min\{|A(s_{\text{init}})|, \\ |\{a \in A(s) \,|\, s \in S, \sum_{t \in \mathcal{T}} P(t \,|\, s, a) > 0\}|\}.$$

The following lemma suggests a structural property of the MDP after removing a set of actions.

Lemma 8. (Structural property). For a preprocessed reachability problem $\Lambda = (S, \mathcal{A}, P, s_{\text{init}}, \mathcal{T})$ that satisfies Assumption 1, if for some $\mathcal{A}' \subset \mathcal{A}$, there holds that $\Pr^{\max}_{\Lambda'}(\text{Reach}(\mathcal{T})) < 1$, then there exists at least one sink end component other than $\mathcal{T}$ in Λ'.

Proof. We postpone the proof and the introduction of relevant concepts to Appendix A.

By Lemma 8, we know that to produce an optimal $\mathcal{A}'$, we need to create at least one sink end component (EC). These could be single states or an EC as a whole that do not have actions that lead to transitions to other states. This is in contrast with the discussion in Section 3.1 regarding the preprocessed MDPs in which every state can reach $\mathcal{T}$ with probability 1. The detailed procedure for computing the resilience degree of an MDP is summarized in Algorithm 1.

4. NUMERICAL EXAMPLE

In this section, we showcase the performance of Algorithm 1 on a numerical example where a mobile agent navigates from the starting point to the destination in a grid world while avoiding obstacles. The grid world

Algorithm 1: Computing resilience degree of an MDP

Input: A preprocessed reachability instance Λ
Output: The resilience degree of Λ
for $K = 1 : \bar{K}$ **do**
  Solve (2) with parameter K and obtain $\mathbf{x} \in \mathbb{R}^{|S|}$
  if $x_{s_{\text{init}}} < 1$ **then**
    return K
  end
end

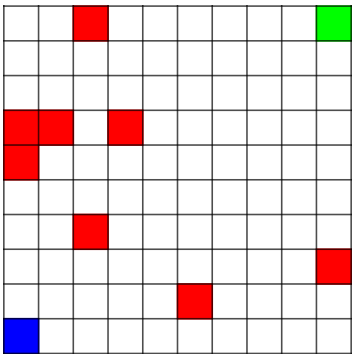
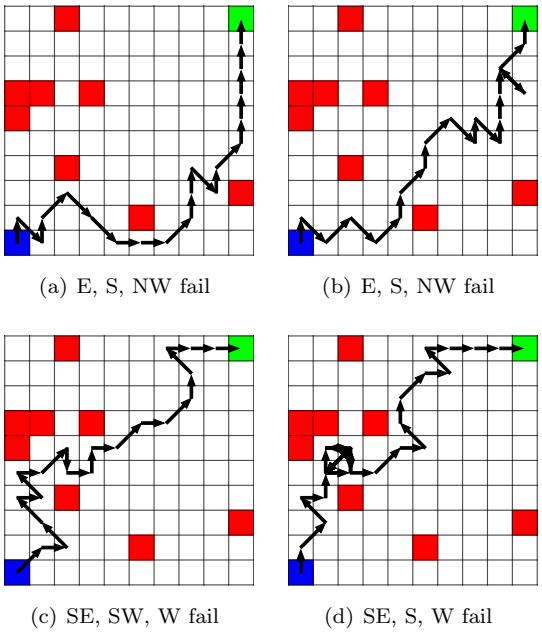


Fig. 4. A navigation environment where the blue cell, green cell and red cells are the starting point, destination and obstacles, respectively.

environment is shown in Fig. 4. The blue and green cells are the starting point and the destination, respectively. The 8 red cells representing obstacles are randomly placed in the environment.

Suppose a mobile agent navigates from the starting point to the destination in the environment, and if it enters the red regions (hitting the obstacles), then it cannot escape and stays there forever. At each state (grid cell), there are a total of 8 actions available to the agent, i.e., {*north (N), northeast (NE), east (E), southeast (SE), south (S), southwest (SW), west (W), northwest (NW)*}. Except at the boundaries, by taking each action, the agent has a high probability of going in the direction and small probabilities of slipping to the neighboring positions. For example, when heading north, the agent has a 0.9 probability of landing on the cell directly to the north of the current position, and it has a 0.05 probability of landing on the cells to the northeast and northwest directions each. On the boundary, the agent stays at the current location with the probability that an action takes it outside of the environment.

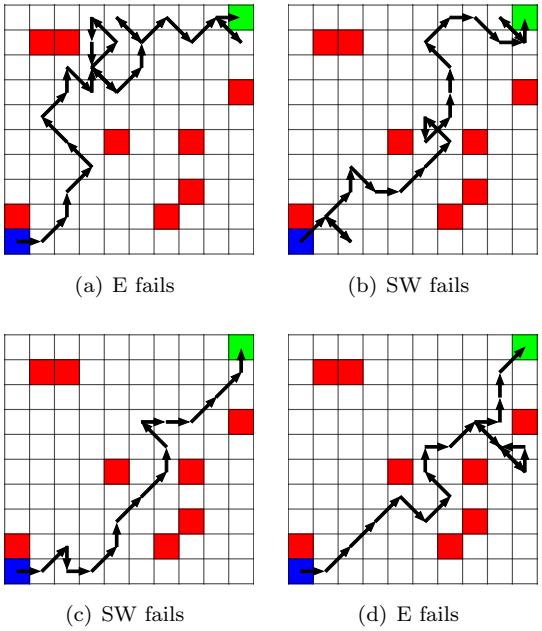
We study the resilience of the agent navigation problem to action failures described above. We first randomly generate two environments as shown in Fig. 5 and 6, respectively, where the obstacles are randomly placed. The resilience degree p, computed via Algorithm 1, are 4 and 2, respectively. We then randomly remove $p - 1$ actions for the agent and show the sample paths generated by the optimal reaching policy. Notably, in the second environment (Fig. 6), since there is an obstacle right above the starting point, the actions N, NE, and NW that result in a positive probability of landing on the obstacle must not be used. On the other hand, since the agent is on the boundary, it stays at the current location when taking actions W and SW. Therefore, only the actions E and



(a) E, S, NW fail (b) E, S, NW fail

(c) SE, SW, W fail (d) SE, S, W fail

Fig. 5. The first environment with resilience degree 4.



(a) E fails (b) SW fails

(c) SW fails (d) E fails

Fig. 6. The second environment with resilience degree 2.

SE can be part of the optimal policy (both of these two actions have strictly positive probabilities of heading east). It turns out that the agent can make it to the goal when one of the two actions fails, but certainly not when both of them do.

5. CONCLUSION

We proposed and studied a notion of resilience to action failures for MDPs under almost-sure reachability objectives. The number of actions that must be disabled so that almost-sure reachability property no longer holds is adopted as a quantitative measure for the resilience degree. We started with a preprocessing step that removes

irrelevant states and actions for the computation of the resilience degree, which reduces the size of the problem. We then showed that the computation problem is NP-hard, and we proposed an exact solution method based on the mixed-integer linear programming. A natural generalization of the current work is to consider other objective functions and develop a more unified analysis and computation framework.

Appendix A. PROOF OF LEMMA 8

We first give the definition for the end components (ECs) of an MDP (Baier and Katoen, 2008, Section 10.6.3).

Definition 9. (EC). An end component EC of an MDP $\mathcal{M} = (S, \mathcal{A}, P, s_{\text{init}})$ is a tuple $EC = (\mathcal{X}, \mathcal{U})$ where

(i) the set of states $\emptyset \neq \mathcal{X} \subset S$;
(ii) the set of actions $\mathcal{U} = \cup_{s \in S} U(s)$ with $U(s) \subset A(s)$ for all $s \in \mathcal{X}$;
(iii) for all $s \in \mathcal{X}$ and $u \in U(s)$, $\text{Supp}(P(\cdot \mid s, u)) \subset \mathcal{X}$;
(iv) for every pair of states $s, s' \in \mathcal{X}$ and $s \neq s'$, there exists a sequence of states and actions $(s_0, u_0 \cdots, s_t)$ with $t \geq 1$ such that $s_0 = s$, $s_t = s'$, and for all $0 \leq \tau \leq t-1$, $u_\tau \in U(s_\tau)$ and $P(s_{\tau+1} \mid s_\tau, u_\tau) > 0$.

Intuitively, an EC is a part of the MDP that is closed in the sense the once the state enters the EC, it can stay there afterwards. A sink $EC = (\mathcal{X}, \mathcal{U})$ is an EC such that for all $s \in \mathcal{X}$, we have $U(s) = A(s)$.

Proof. Since $\Pr^{\max}_{\Lambda'}(\text{Reach}(\mathcal{T})) < 1$, then by (Forejt et al., 2011, Algorithm 4), we know that there must exist at least one state $s \in S$ that does not have a path to $\mathcal{T}$ in the transition diagram (otherwise, all states will have the maximum probability 1 of reaching $\mathcal{T}$). Therefore, the maximum probability of reaching $\mathcal{T}$ from s is 0. By the Bellman optimality criterion, for all actions $a \in A(s)$, the set of states in $\text{Supp}(P(\cdot \mid s, a)$ must also have 0 probability of reaching $\mathcal{T}$. The same argument applies to the states in $\text{Supp}(P(\cdot \mid s, a)$. By including all these states and actions, we have found a sink EC.

REFERENCES

Antonyshyn, L., Silveira, J., Givigi, S., and Marshall, J. (2022). Multiple mobile robot task and motion planning: A survey. *ACM Computing Surveys*. doi:10.1145/3564696.

Baier, C. and Katoen, J. (2008). *Principles of model checking*. MIT Press.

Bäuerle, N. and Rieder, U. (2011). *Markov Decision Processes with Applications to Finance*. Springer Berlin, Heidelberg.

Bouvier, J. and Ornik, M. (2022). Designing resilient linear systems. *IEEE Transactions on Automatic Control*, 67(9), 4832–4837. doi:10.1109/TAC.2022.3163242.

Bucić, M., Ornik, M., and Topcu, U. (2018). Graph-based controller synthesis for safety-constrained, resilient systems. In *Annual Allerton Conference on Communication, Control, and Computing*, 297–304. Monticello, IL, USA.

Chatterjee, K. and Henzinger, T.A. (2012). A survey of stochastic ω-regular games. *Journal of Computer and System Sciences*, 78(2), 394–413. doi:10.1016/j.jcss.2011.05.002.

Coudert, D., Datta, P., Perennes, S., Rivano, H., and Voge, M.E. (2007). Shared risk resource group complexity and approximability issues. *Parallel Processing Letters*, 17(02), 169–184. doi:10.1142/S0129626407002958.

Coudert, D., Pérennes, S., Rivano, H., and Voge, M.E. (2016). Combinatorial optimization in networks with Shared Risk Link Groups. *Discrete Mathematics and Theoretical Computer Science*, 18(3), 25. doi:10.46298/dmtcs.1297.

Courcoubetis, C. and Yannakakis, M. (1998). Markov decision processes and regular events. *IEEE Transactions on Automatic Control*, 43(10), 1399–1418. doi:10.1109/9.720497.

de Alfaro, L. (1999). Computing minimum and maximum reachability times in probabilistic systems. In *International Conference on Concurrency Theory*, 66–81. Eindhoven, The Netherlands.

Forejt, V., Kwiatkowska, M., Norman, G., and Parker, D. (2011). Automated verification techniques for probabilistic systems. In M. Bernardo and V. Issarny (eds.), *Formal Methods for Eternal Networked Software Systems*, volume 6659 of *Lecture Notes in Computer Science*, 53–113. Springer. doi:10.1007/978-3-642-21455-4_3.

Garey, M.R. and Johnson, D.S. (1979). *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1 edition.

Jha, S., Sheyner, O., and Wing, J. (2002). Two formal analyses of attack graphs. In *IEEE Computer Security Foundations Workshop*, 49–63. Cape Breton, NS, Canada.

Kuipers, F.A. (2012). An overview of algorithms for network survivability. *International Scholarly Research Notices*, 2012, e932456. doi:10.5402/2012/932456.

Löding, C. and Rohde, P. (2003). Solving the sabotage game is PSPACE-hard. In *Mathematical Foundations of Computer Science*, 531–540. Berlin, Germany.

Nilim, A. and Ghaoui, L.E. (2005). Robust control of Markov decision processes with uncertain transition matrices. *Operations Research*, 53(5), 780–798. doi:10.1287/opre.1050.0216.

Puterman, M.L. (2014). *Markov decision processes: Discrete stochastic dynamic programming*. John Wiley & Sons.

van Benthem, J. (2005). An essay on sabotage and obstruction. In D. Hutter and W. Stephan (eds.), *Mechanizing Mathematical Reasoning*, volume 2605 of *Lecture Notes in Computer Science*, 268–276. Springer. doi:10.1007/978-3-540-32254-2_16.

Wiesemann, W., Kuhn, D., and Rustem, B. (2013). Robust Markov decision processes. *Mathematics of Operations Research*, 38(1), 153–183. doi:10.1287/moor.1120.0566.

Xu, H. and Mannor, S. (2009). Parametric regret in uncertain Markov decision processes. In *IEEE Conference on Decision and Control & Chinese Control Conference*, 3606–3613. Shanghai, China.

Zhang, P., Cai, Y., Tang, L., and Zhao, W. (2011). Approximation and hardness results for label cut and related problems. *Journal of Combinatorial Optimization*, 21(2), 192–208. doi:10.1007/s10878-009-9222-0.