Shielding Atari Games with Bounded Prescience

Mirco Giacobbe, Mohammadhosein Hasanbeig, Daniel Kroening, Hjalmar Wijk
Computer Science Department, University of Oxford, UK

Introduction

We present the first explicit-state method for analysing and ensuring safety of DRL agents for Atari games.

- We propose 42 safety properties for 31 games.
- We evaluate the safety of available Deep Reinforcement Learning (DRL)[1] agents.
- We improve safety through shielding [2] using bounded explicit-state exploration.

Background

- We consider 31 Atari games with unique dynamics given by a black-box emulator.
- Each game is a deterministic MDP $(\mathcal{S}, \mathcal{A}, T, R)$.
- "no-op" non-determinism added: no inputs for the first $k \in \{0, \ldots, 30\}$ frames.

- Learn deterministic policy $\pi : \mathcal{S} \to \mathcal{A}$ through SOTA DRL methods.

Safety Properties

- Safety property $\phi \subseteq \mathcal{S}$ is set of safe states.
- Satisfied if for all reachable states $s \in \phi$.
- Labelling handcrafted from graphical output.

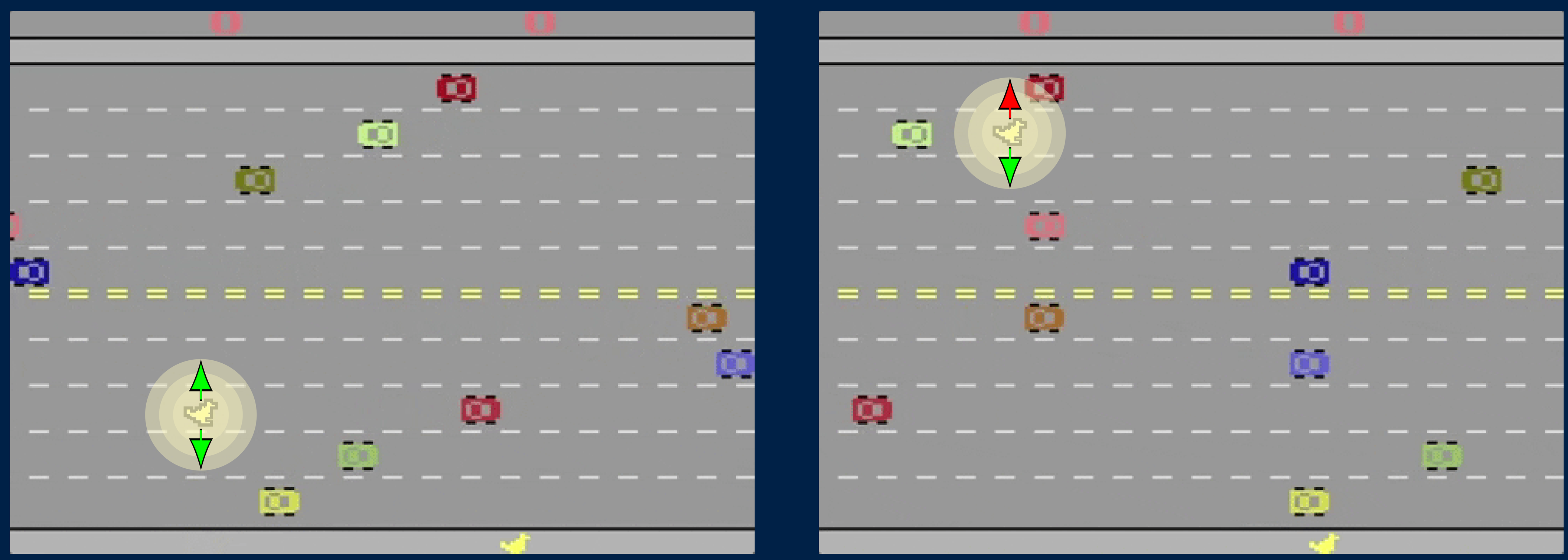
Example Properties	
Name	Description
Assault-Overheat	Die from overheating
Bowling-NoHit	Miss all pins
Freeway-Hit	Get hit by a car

- ▶ To verify ϕ for policy π run games with π for all values of k.
- ▶ This will visit every reachable state, ϕ true iff. for all states visited $s \in \phi$.

Bounded Prescience Shield (BPS)

- ▶ Modifies policy π at runtime by changing the action when $\pi(s)$ unavoidably leads to unsafe state within n steps.
- ▶ Computed by using the ability to roll back the emulator state, with no knowledge of the MDP.

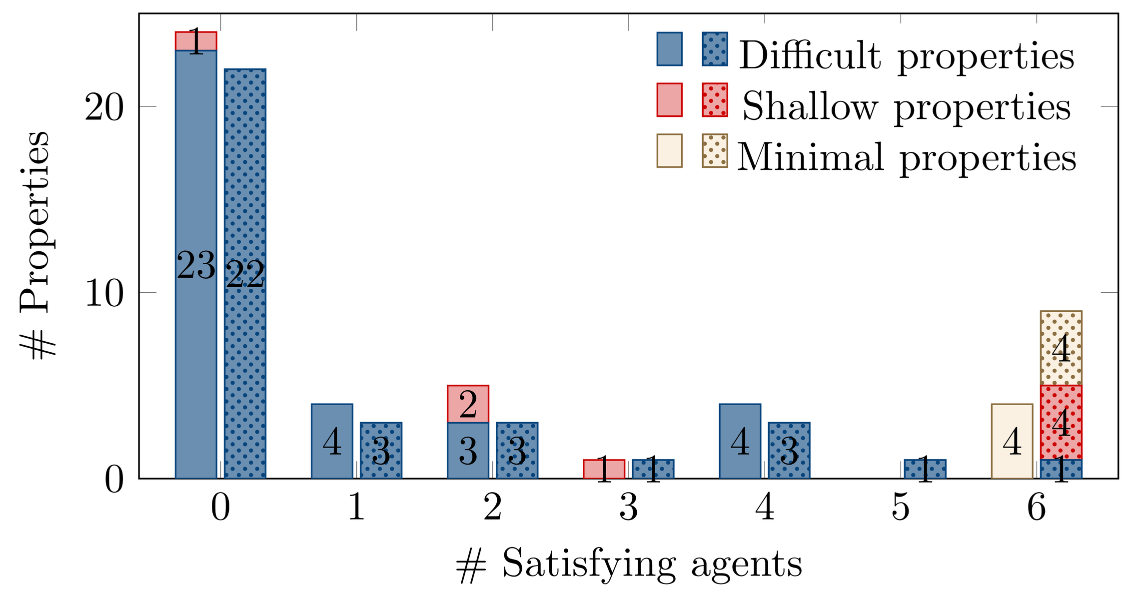
"**Explicit-state** verification demonstrates that **DRL** algorithms do not learn to satisfy even simple **Safety Properties**".



(a) (b)

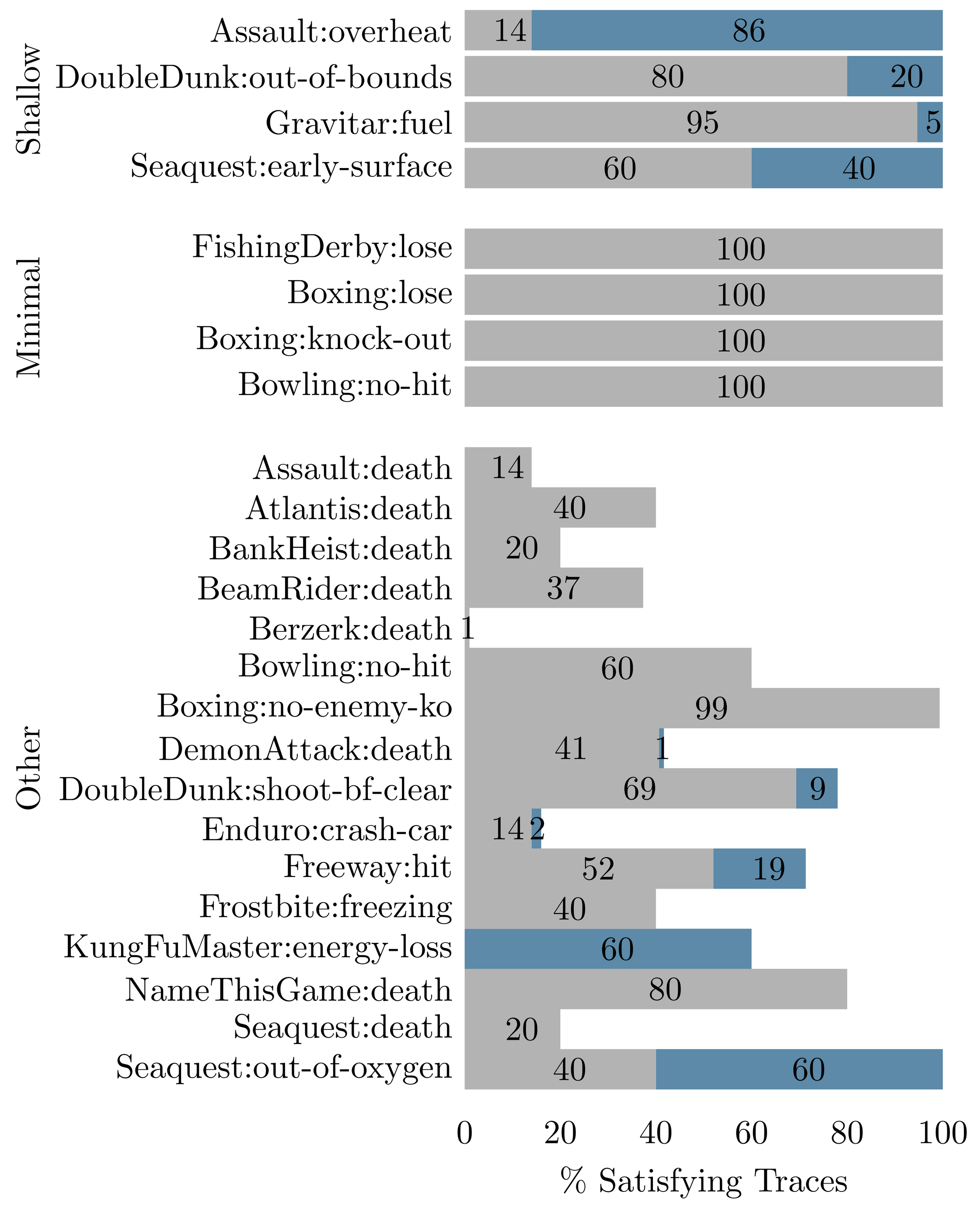
Take a Picture to Download the Full Paper

Results



Properties grouped by number of satisfying agents before (w/o dots) and after BPS (with dots).

- ▶ Minimal properties are satisfied by random agents, shallow properties require no planning.
- ▶ No non-minimal property is satisfied by more than 4 agents.



Effect of shielding on the average safety achieved.

- ▶ With BPS all agents satisfy all shallow properties.

References

[1] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[2] M. Alshiekh, R. Bloem, R. Ehlers, B. Könighofer, S. Niekum, and U. Topcu, "Safe reinforcement learning via shielding," in *AAAI*. AAAI Press, 2018, pp. 2669–2678.