Predictive Safety Shields for Reinforcement Learning Based Controllers

Stage Master 2, 2024–2025

Key words : Safety Shields
Reinforcement Learning
Control Design
Planning in Dynamic Environments

Laboratory: U2IS, ENSTA Paris

Address: 828 boulevard des maréchaux 91762 Palaiseau Cedex

Advisors: Elena Vanneaux **elena.vanneaux@ensta-paris.fr**
Hanna Krasowski **krasowski@berkeley.edu**

Duration: 6 months

Scholarship: according to the legislation.

Context Reinforcement learning (RL) has been widely adopted in robotics for its ability to learn from interaction with the environment through feedback. It enables robots to adapt to environmental changes and optimize their behavior according to performance criteria not known in advance [6]. However, to use RL-based controllers for safety-critical tasks, one should also ensure that nothing "bad" occurs during the training and deployment of RL agents. Indeed, autonomous vehicles should never drive off the highway, robotic prostheses should never force their users' joints past their range of motion, and drones should never fall out of the sky. The vulnerability of standard RL-based controllers to failures has spurred significant growth in research on safe RL in the past decade [2].

In this internship, we will focus on provably safe RL, that provides hard safety guarantees for both training and operation [7]. Provably safe RL approaches can be categorized into preemptive and post-posed shielding [1]. In the preemptive method, the agent can only choose from actions that have been a priori verified as safe. However, if a preemptive shield is too conservative, i.e., it identifies only a few actions from the action space as safe, the agent's capabilities for exploring the environment are significantly reduced, which can lead to lower overall performance [3]. In post-posed shielding, the safety filter monitors the RL agent behavior. If the agent wants to take an unsafe action, the shield replaces it with a fallback strategy. Post-posed shields are usually more computationally efficient than preemptive. Also, they are often easier to use in dynamic environments, which we want to investigate in this internship. Still, in dangerous scenarios, a shield forces the system to use a predetermined safe but likely sub-optimal policy [1]. Hence, while guaranteeing safety, shielding often contradicts task efficiency. This internship aims to balance safety and performance by developing provably safe RL algorithms with the agent's guaranteed near-optimal behavior.

In our proof-of-concept work [5], we propose a predictive safety shield for model-based reinforcement learning agents in discrete space. The safety shield updates the Q-function locally based on safe predictions, which originate from a safe simulation of the environment model. This shielding approach improves performance while maintaining hard safety guarantees. Our experiments on grid-world environments demonstrate that even short prediction horizons can be sufficient to identify the optimal path. We observe that our approach is robust to distribution shifts, e.g., between simulation and reality, without requiring additional training. This internship aims to extend the proposed approach to dynamically changing environments [4].

Goals The goals of the internship consist of

- exploring the state-of-the-art safety shields for reinforcement learning algorithms
- proposing a shield that ensures safe behavior in dynamically changing environments.
- testing the proposed approach in GridWorld and PacMan environments

Profile of a candidate. For this position, you should meet the following requirements:

- enrollment in a Master's program or equivalent in computer science, applied mathematics science, engineering, or related disciplines;
- rigorous knowledge in formal verification, control design, and reinforcement learning;
- excellent programming skills (Python);
- proficiency in spoken and written English;

The candidate will have to submit the documents following:

- a cover letter;
- a resume;
- a copy of diplomas, bachelor's and master's degree transcripts.

In case of a successful internship, a Ph.D. offer in ENSTA Paris might be proposed.

References

[1] Mohammed Alshiekh, Roderick Bloem, Rüdiger Ehlers, Bettina Könighofer, Scott Niekum, and Ufuk Topcu. Safe reinforcement learning via shielding. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32, 08 2017.

[2] Lukas Brunke, Melissa Greeff, Adam W. Hall, Zhaocong Yuan, Siqi Zhou, Jacopo Panerati, and Angela P. Schoellig. Safe learning in robotics: From learning-based control to safe reinforcement learning. *Annual Review of Control, Robotics, and Autonomous Systems*, 5(1):411–444, 2022.

[3] Kai-Chieh Hsu, Haimin Hu, and Jaime F. Fisac. The safety filter: A unified view of safety-critical control in autonomous systems. *Annual Review of Control, Robotics, and Autonomous Systems*, 7(1):47–72, July 2024.

[4] Nils Jansen, Bettina Könighofer, Sebastian Junges, Alex Serban, and Roderick Bloem. Safe reinforcement learning using probabilistic shields (invited paper). Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2020.

[5] Pin Jin. A safety filter for rl algorithms based on a game-theoretic mpc approach, 2024. PRE - Research Project, ENSTA.

[6] Jens Kober and Jan Peters. *Reinforcement Learning in Robotics: A Survey*, pages 9–67. Springer International Publishing, Cham, 2014.

[7] Hanna Krasowski, Jakob Thumm, Marlon Müller, Lukas Schäfer, Xiao Wang, and Matthias Althoff. Provably safe reinforcement learning: Conceptual analysis, survey, and benchmarking. *Transactions on Machine Learning Research*, 2023. Survey Certification.