

Learning to Exploit a Backdoored RL Controller: Coordinated MARL Attacks in Dynamic 6G Cyber-Physical Systems

Supervisors: Alex Pierron, Joaquin Garcia-Alfaro

Contact: alex.pierron@telecom-sudparis.eu, joaquin.garcia_alfaro@telecom-sudparis.eu

Laboratory: SAMOVAR, Telecom SudParis, Institut Polytechnique de Paris

Keywords: Critical Cyber-Physical Systems, 6G, Reconfigurable Intelligent Surfaces, Multi-Agent Reinforcement Learning, Adversarial Reinforcement Learning, Backdoor Attacks, Physical-Layer Security, Trustworthy AI

1. Context

Reconfigurable Intelligent Surfaces (RISs) are programmable metasurfaces used to shape radio propagation in Beyond-5G/6G systems. Because software decisions directly affect the physical channel, their security is inherently cyber-physical: compromising the control logic can change what legitimate users and eavesdroppers receive. Recent work has studied attacks on programmable metasurfaces, learning-based RIS control, and RIS-assisted secure communication [1–3].

This project starts from an existing Telecom SudParis simulator and attack scenario [4]. The RIS backdoor, its physical-layer implementation, and the attack model are provided; the student will not redesign them. The research problem lies one layer above: several adversarial agents have access to the backdoor and must learn how to exploit it jointly. The study will use PPO and HAPPO [5, 6], together with standard Multi-Agent Reinforcement Learning (MARL) concepts [7] and related work on cooperative strategies for secure wireless systems [8, 9].

2. Objective

The goal is to determine which PPO/HAPPO training choices lead to effective and repeatable coordination, and whether the same choices remain relevant once the radio geometry becomes dynamic. The project will focus on three questions:

- Which training parameters have the strongest effect on attack effectiveness, convergence and stability?
- Are the best configurations robust across random seeds and initial conditions?
- What changes when legitimate and/or adversarial nodes move during an episode?

3. Proposed work

The work is organized in four steps:

1. **Establish the baseline.** Review the core literature, reproduce the existing static scenario, and validate the state/action spaces, reward design, training loop, and evaluation metrics.
2. **Characterize attacker coordination.** Consolidate the PPO/HAPPO setup and identify how coordinated behaviour emerges, including the metrics needed to distinguish individual performance from joint attack effectiveness.
3. **Identify robust training configurations.** Run a targeted sensitivity study on a small set of influential parameters (e.g., learning rate, clipping, entropy regularisation, discount/GAE settings, rollout length), with repeated seeds to separate systematic effects from training noise.
4. **Introduce dynamics.** Add one controlled mobility model while keeping the physical-layer backdoor fixed. Compare representative configurations in static and dynamic settings to identify when coordination transfers, degrades, or fails.

4. Expected outcomes

The expected contributions are a reproducible PPO/HAPPO baseline for coordinated attackers, an experimentally supported characterization of the training parameters that matter most, and a quantitative analysis of the impact of mobility on learned coordination. The resulting code, configurations, seeds, and analysis scripts will accompany a concise research report and make the experiments directly reproducible.

References

- [1] Yunsong Huang et al. Pilot backdoor attack against deep reinforcement learning empowered intelligent reflection surface for smart radio. *IEEE Transactions on Wireless Communications*, 2025.
- [2] H. Chen, H. Saeidi, S. Venkatesh, K. Sengupta, and Y. Ghasempour. Wavefront manipulation attack via programmable mmwave metasurfaces: From theory to experiments. In *Proceedings of the 16th ACM Conference on Security and Privacy in Wireless and Mobile Networks*, pages 317–328, May 2023.
- [3] Yang, Xiong, Zhao, Niyato, Xiao, and Wu. Deep reinforcement learning-based intelligent reflecting surface for secure wireless communications. *IEEE Transactions on Wireless Communications*, 20(1):375–388, September 2020.
- [4] Pierron, Barbeau, Rubio-Hernan, De Cicco, and Garcia-Alfaro. A fairness-aware strategy for b5g physical-layer security leveraging reconfigurable intelligent surfaces. In *18th International Symposium on Foundations & Practice of Security*, LNCS. Springer Nature, 2025. 25–27 November 2025.
- [5] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*, 2017. doi: 10.48550/arXiv.1707.06347.
- [6] Jakub Grudzien Kuba, Ruiqing Chen, Muning Wen, Ying Wen, Fanglei Sun, Jun Wang, and Yaodong Yang. Trust region policy optimisation in multi-agent reinforcement learning. In *International Conference on Learning Representations (ICLR)*, 2022.
- [7] Stefano V. Albrecht, Filippos Christianos, and Lukas Schäfer. *Multi-Agent Reinforcement Learning: Foundations and Modern Approaches*. MIT Press, 2024.
- [8] Peng, Zhang, Kong, Pan, Li, and Wang. Deep reinforcement learning for ris-aided multiuser full-duplex secure communications with hardware impairments. *IEEE Internet of Things Journal*, 9(21):21121–21135, May 2022.
- [9] Si, Cheng, Li, Cheng, Wang, and Al-Dhahir. Cooperative jamming for secure transmission with both active and passive eavesdroppers. *IEEE Transactions on Communications*, 68(9):5764–5777, June 2020.