Master's Thesis Internship Proposal

Location: U2IS – ENSTA, Institut Polytechnique de Paris, Palaiseau
Duration: 6 months, starting in March/April 2027 **Stipend:** ∼680€/month

Countering Coordinated Drone Swarms with Decentralized MARL

The objective of this internship is to lay the foundations of a learning-based approach to the interception of drone swarms by a swarm of interceptor Unmanned Aerial Vehicles (UAVs). The considered scenario is the defense of a protected area against a group of intruding drones: several interceptors must jointly decide *which intruder to engage* and *how to reach it*, while the intruders coordinate to saturate, split or deceive the defense. Classical counter pipelines decouple these two questions: a target assignment is solved first (Linear Assignment Problem, LAP [1]), then each interceptor applies a guidance law. This decoupling can degrade performance, since intruders react to the defense and the assignment cost does not account for what the guidance can actually achieve.

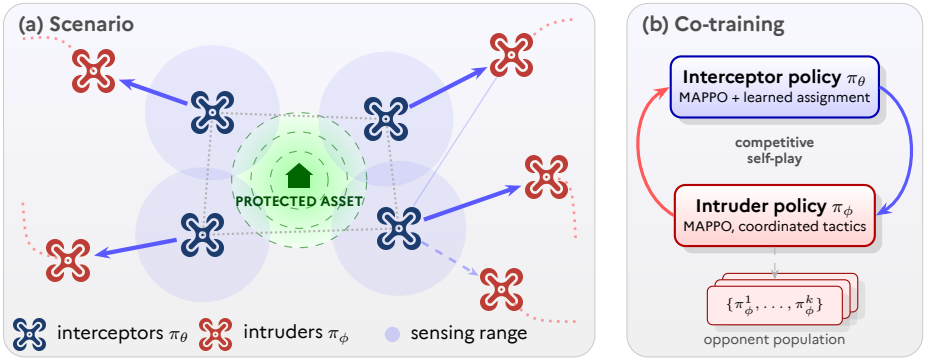


Figure 1: (a) Swarm-vs-swarm scenario: the assignment emerges from the decentralized interceptor policy. (b) Both swarms are trained by Multi-Agent Reinforcement Learning (MARL) against a growing population of opponents.

Learning-based interception has recently attracted strong interest, with promising results. Gavin *et al.* trained an agile interceptor and an agile target against each other with PPO in a GPU-parallel simulator [2], then extended this to a team of net-carrying pursuers against a single evader, using MAPPO and prioritized fictitious self-play [3]. With a single evader, however, there is *nothing to assign*: the question of which interceptor engages which intruder does not arise.

We propose to treat the problem as a **competitive game between two cooperative swarms** (N interceptors versus M intruders, $M > 1$, possibly $N < M$), in which **both teams are trained by MARL**, and where the assignment emerges from a decentralized policy: each interceptor attends over the intruders it observes and over the intentions of its teammates, so that commitment to a target and avoidance of duplicated engagements are learned jointly with the pursuit (Figure 1). On the other side, the intruders form a team that learns coordinated tactics (saturation, splitting, decoys) rather than individual evasion. Each team is trained in the Centralized Training Decentralized Execution (CTDE) paradigm [4], and the two teams are trained against each other [3, 5].

Objectives of the Internship

- **Simulation environment and baselines:** Develop a fast, vectorized (GPU) simulation of N interceptors versus M intruders with simplified dynamics, limited sensing and capture

conditions, in the spirit of existing JAX multi-agent frameworks [6]. Implement reference strategies and scripted behaviors (split, decoy, evasive maneuvers) for the intruders.

- **Decentralized learned assignment:** Design an interceptor policy in which target selection is part of each agent's action, through an attention or pointer mechanism over the set of observed intruders [7], and train it with MAPPO against the scripted intruders. Evaluate it against assignment baselines with learned guidance, in terms of interception rate, time to interception, number of intruders reaching the protected area, and scaling with the number of agents.
- **First steps towards competitive swarms:** Replace the scripted intruders with a learned intruder team and run first experiments where both swarms are trained against each other. Possible extensions: transfer to the PX4 Software-in-the-Loop simulation of the U2IS fleet.

Candidate Profile

We are looking for a student in the final year of a Master's degree (M2) or of an engineering school, in robotics, machine learning, applied mathematics or computer science, with a strong interest in multi-robot systems and research. The following skills are required:

- Solid background in deep/Reinforcement learning (policy-gradient methods such as PPO).
- Good programming skills in Python and experience with a deep learning framework.
- Basic knowledge of robot/vehicle dynamics and control.
- Ability to work autonomously and analyze experimental results critically.

The following skills are considered an advantage but are not mandatory: JAX, Multi-Agent Reinforcement Learning (MARL), drones and robotics softwares (PX4, ROS 2, Gazebo). The internship is intended as a first step towards a PhD, candidates interested in pursuing a PhD will be given priority.

How to Apply

Candidates are invited to send a resume and their academic transcripts for the last two years to Hamza Chakraa (`hamza.chakraa@ensta.fr`) and Elena Vanneaux (`elena.vanneaux@ensta.fr`).

References

[1] H. W. Kuhn. The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1–2):83–97, 1955.

[2] T. Gavin, S. Lacroix, and M. Bronz. Agile interception of a flying target using competitive reinforcement learning. arXiv:2603.16279, 2026.

[3] T. Gavin and M. Bronz. Intercepting an agile target with net-carrying drones using competitive multi-agent reinforcement learning. In *International Conference on Unmanned Aircraft Systems (ICUAS)*, pp. 759–766, 2026.

[4] C. Amato. An introduction to centralized training for decentralized execution in cooperative multi-agent reinforcement learning. arXiv:2409.03052, 2024.

[5] M. Lanctot, V. Zambaldi, A. Gruslys, A. Lazaridou, K. Tuyls, J. Pérolat, D. Silver, and T. Graepel. A unified game-theoretic approach to multiagent reinforcement learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2017.

[6] A. Rutherford et al. JaxMARL: Multi-agent RL environments and algorithms in JAX. In *NeurIPS Datasets and Benchmarks Track*, 2024. arXiv:2311.10090.

[7] W. Kool, H. van Hoof, and M. Welling. Attention, learn to solve routing problems! In *International Conference on Learning Representations (ICLR)*, 2019.