Multi-Agent Multi-Objective Deep Reinforcement Learning for Efficient and Effective Pilot Training

FT2019

Johan Källström, Saab AB & LiU, Fredrik Heintz, LiU

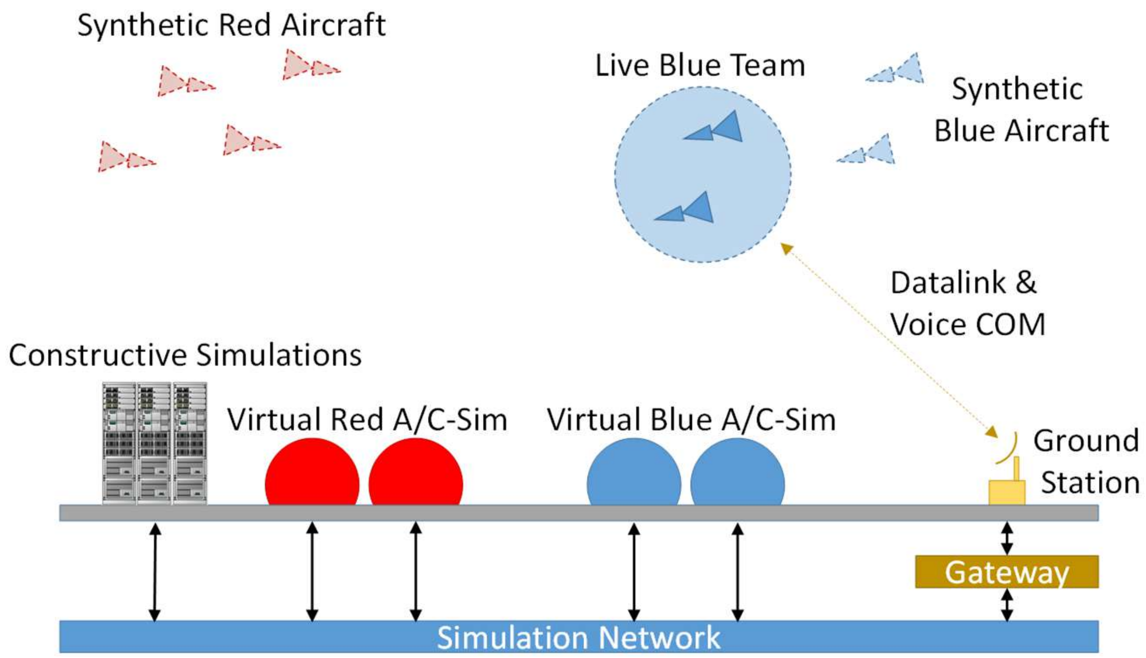
Johan.Kallstrom@saabgroup.com

Background

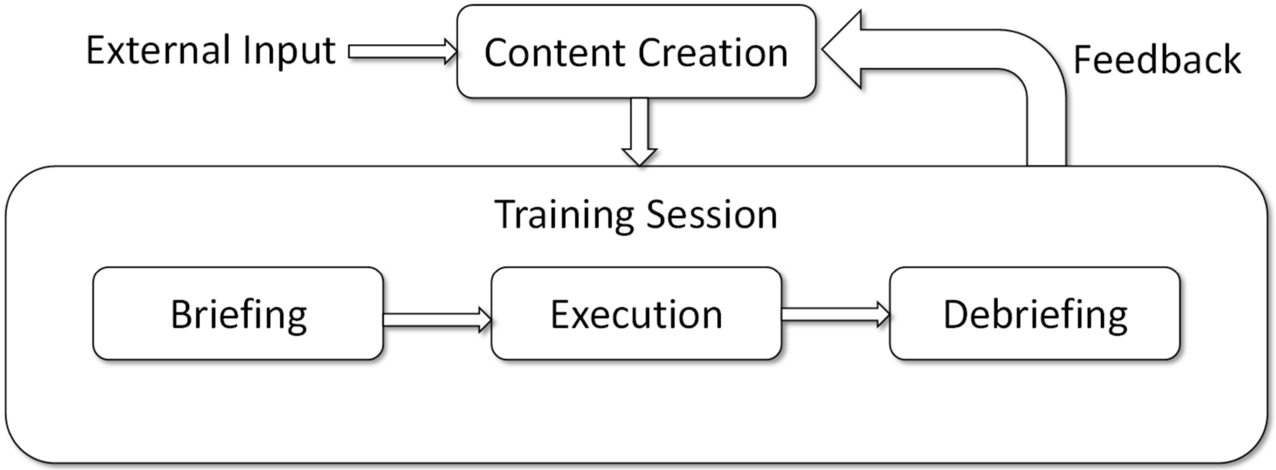
- Challenges in air combat training
  - High costs
  - Limited air space
  - Safety
  - May reveal tactics during live training
  - Difficult to realize relevant training scenarios
- Must use simulation to a higher degree
  - Ground-based simulators
  - Embedded simulation capabilities in aircraft
  - Simulation networks - Live, Virtual and Constructive (LVC) simulation

NFFP7 Project Overview

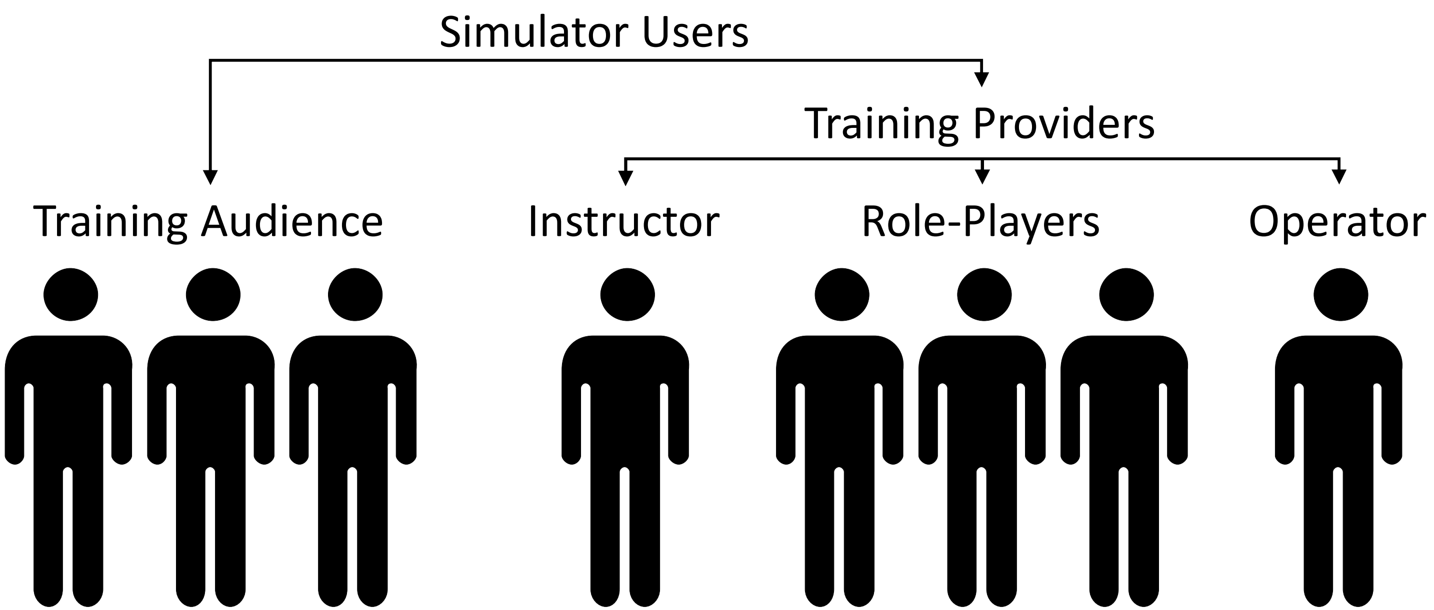
- Find efficient and effective pilot training solutions for fighter aircraft:
  - Lower costs
  - Improve availability
  - Realize more complex scenarios
  - Higher training value
- Research focus:
  - Use machine learning to contruct synthetic allies and adversaries
  - Use machine learning to provide competency-based training



Training Process

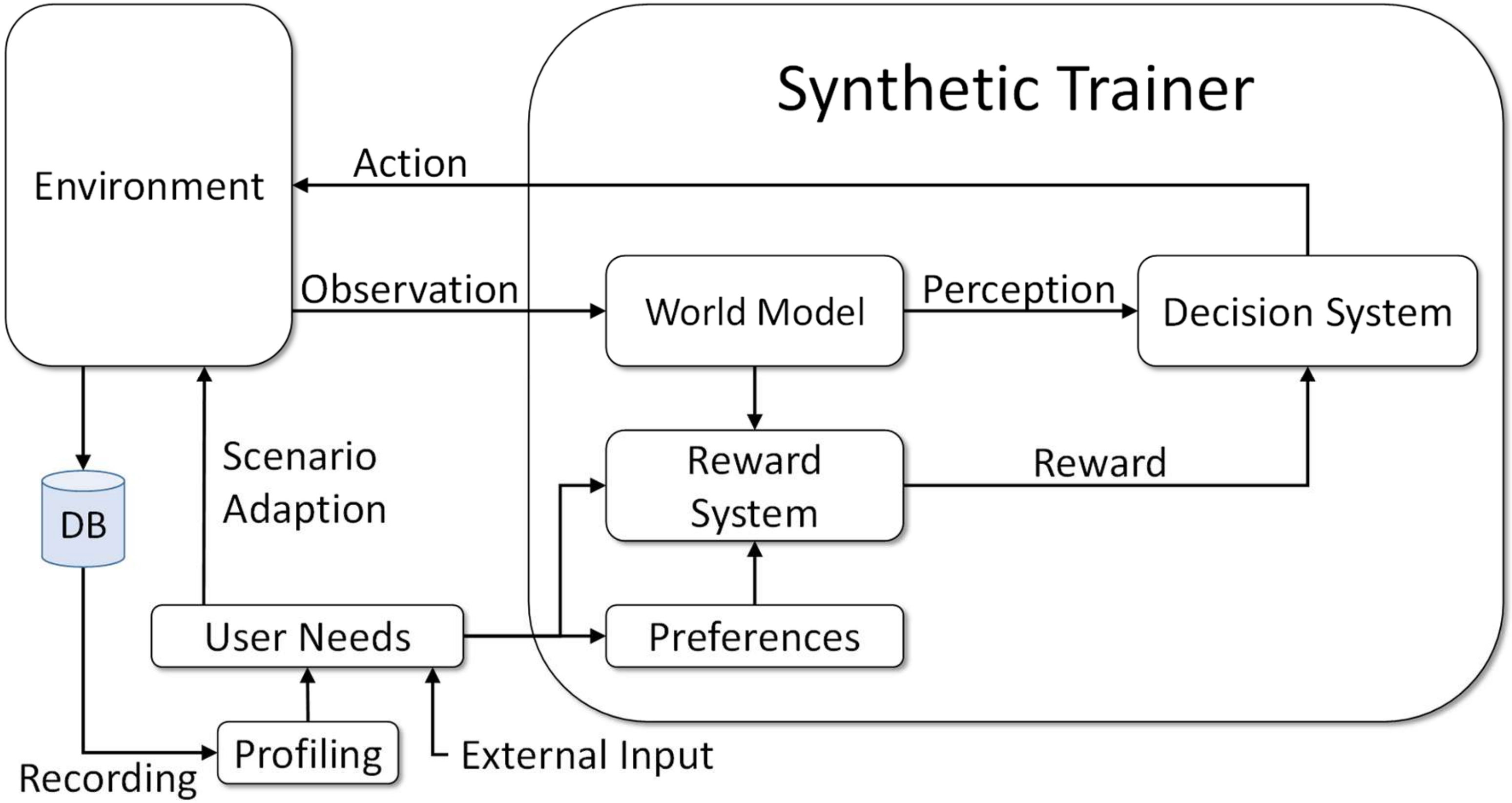


User Roles



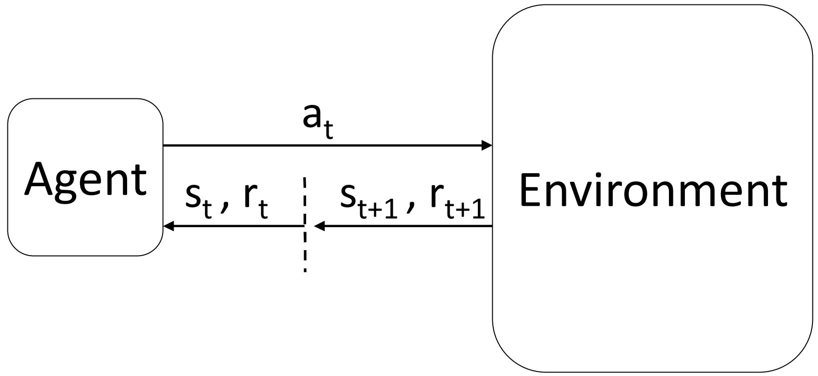
Simulator Users
Training Providers
Training Audience
Instructor
Role-Players
Operator

Adaptive Training System



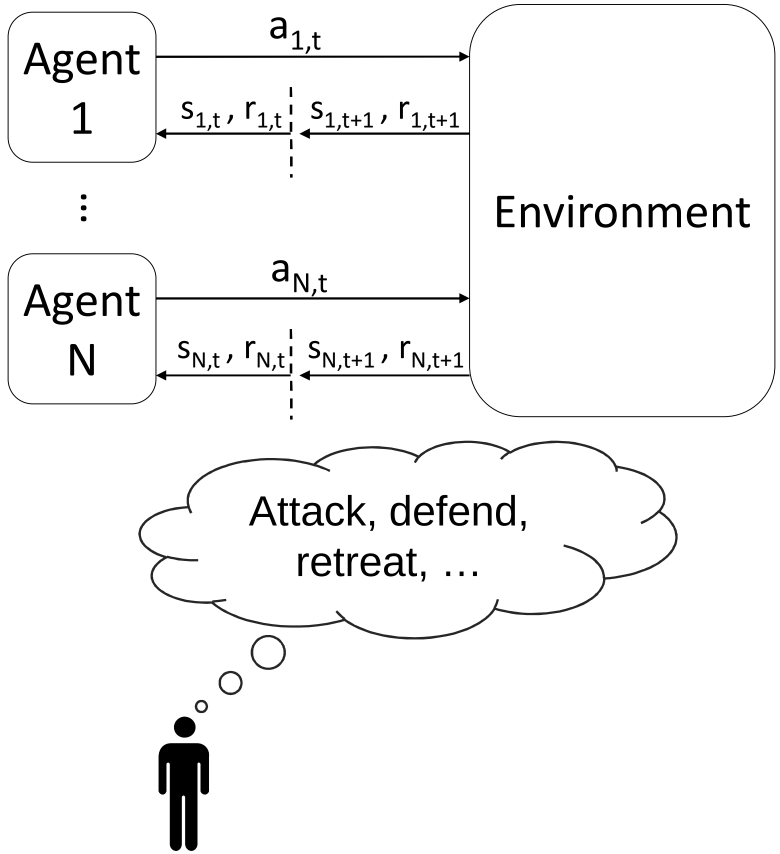
Reinforcement Learning

- **Reinforcement Learning**
  - Learning by interaction with an environment
  - Learning is guided by a reward function
  - The goal is to maximize future return
  - Must balance between exploration and exploitation
- **Deep Reinforcement Learning**
  - Use neural network to represent the decision making policy



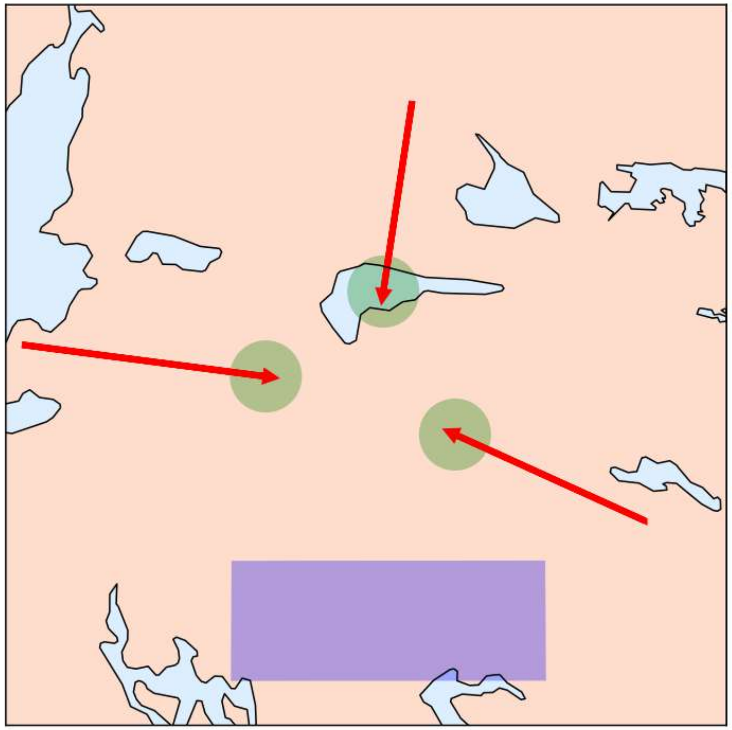
Reinforcement Learning

- **Multi-Agent Reinforcement Learning**
  - Train teams of competing agents
    - Multi-agent exploration
    - Multi-agent credit-assignment
- **Multi-Objective Reinforcement Learning**
  - Prioritize among conflicting objectives
    - Build diverse agents
    - Build adaptive agents
  - Approaches
    - Learn sets of policies
    - Learn single policy that is conditioned on objective preferences



Multi-Agent Reinforcement Learning

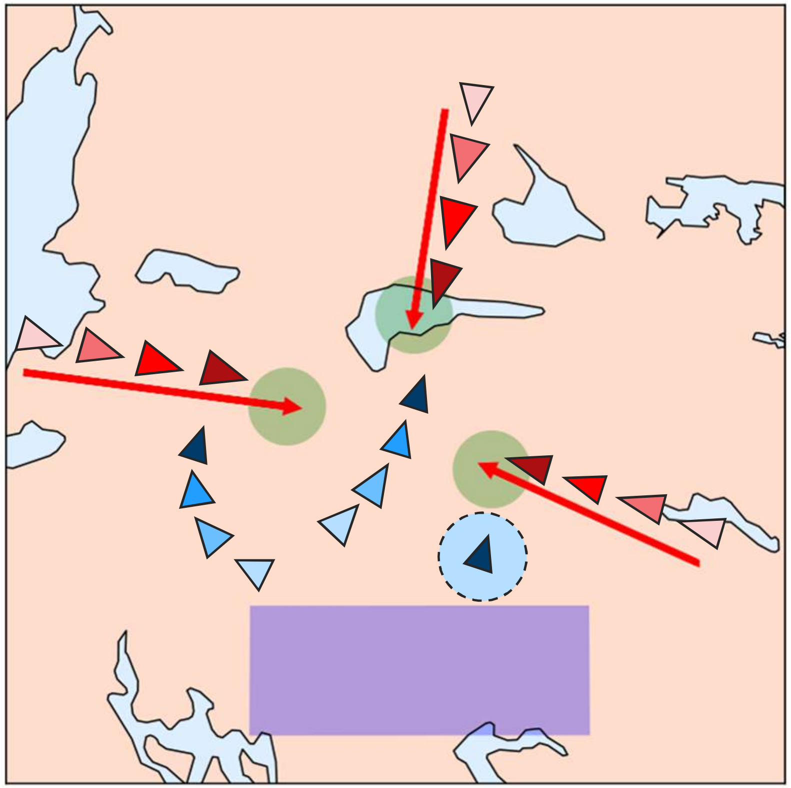
- Cooperative defense of high-value assets, using MADDPG algorithm
- Attackers are controlled by handcrafted behavior trees (”if not threatened attack, else return to base”)
- Defenders try to optimize shared reward by minimizing distance between each attacker and closest defender



$$r_t = -\sum_{i=1}^{3} \min(\|p_{a_i} - p_{d_1}\|, \|p_{a_i} - p_{d_2}\|, \|p_{a_i} - p_{d_3}\|)$$

Multi-Agent Reinforcement Learning

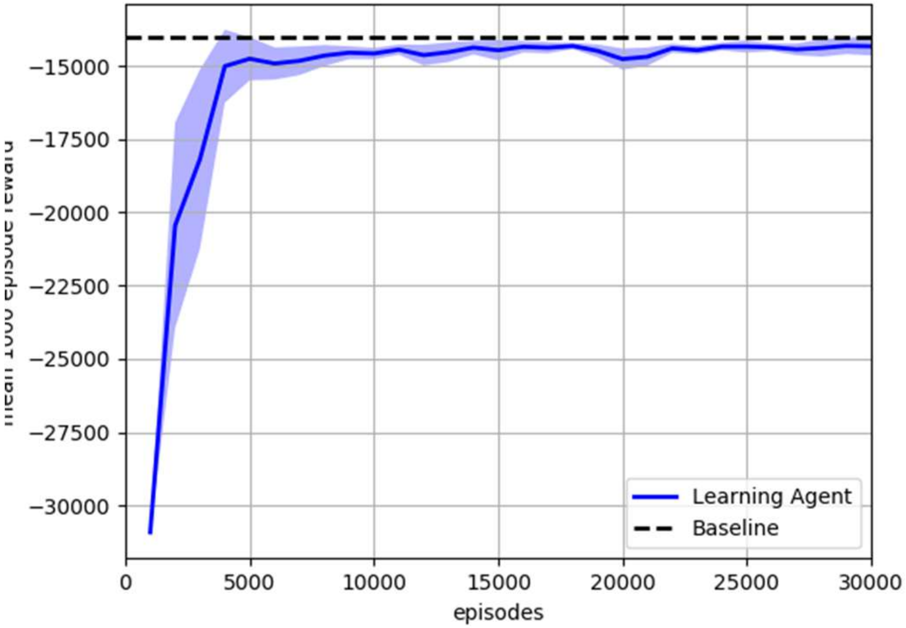
- Observation space
  - Other agents' positions in last 4 time steps
- Action spaces
  - High-level & discrete: Selected enemy to pursue (given as input to low-level controller)
  - Low-level & continuous: Left/right turns with various load factor (2-4g)
    - Silent and communicating agents



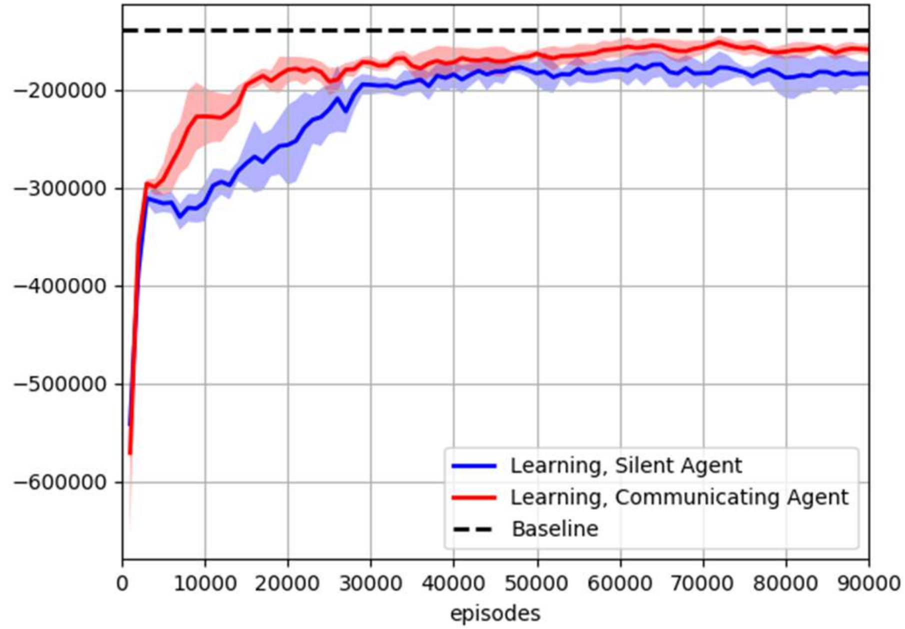
Multi-Agent Reinforcement Learning

Learning progress

Using high-level discrete action space

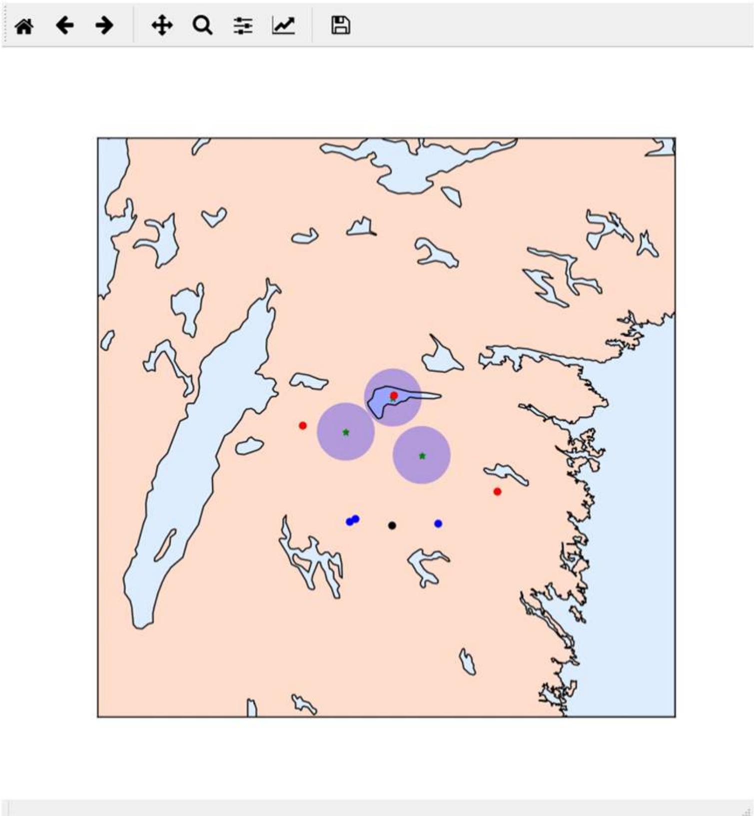


Using low-level continuous action space

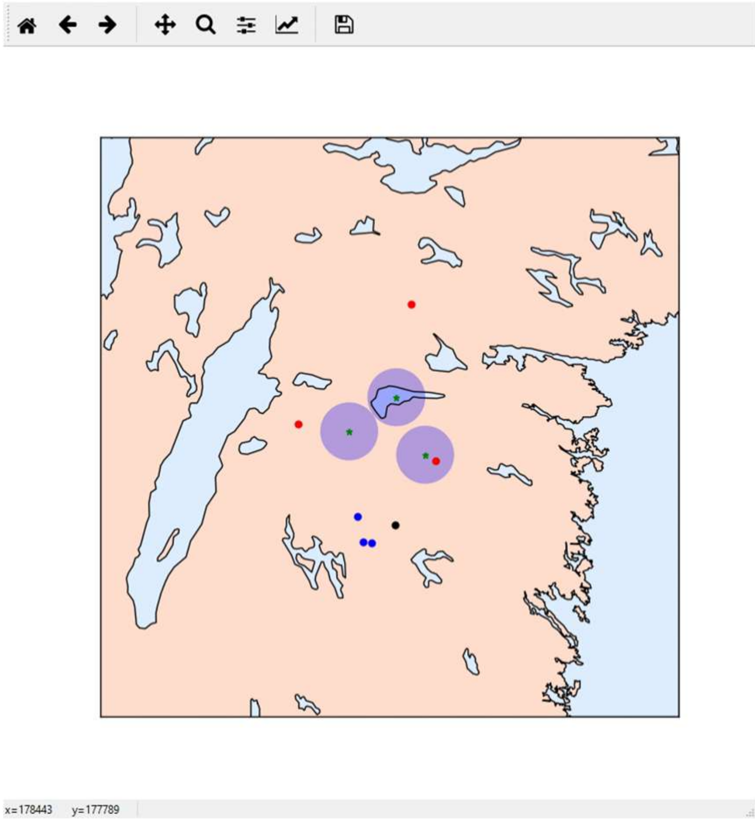


Multi-Agent Reinforcement Learning

High-level discrete action space

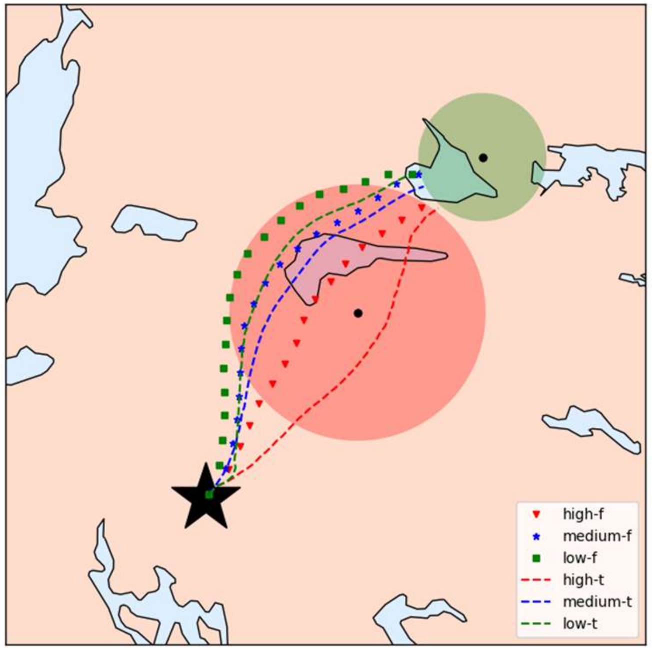


Low-level continuous action space



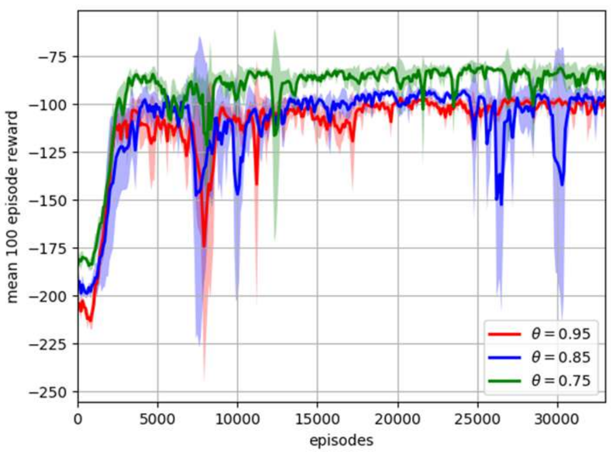
Multi-Objective Reinforcement Learning

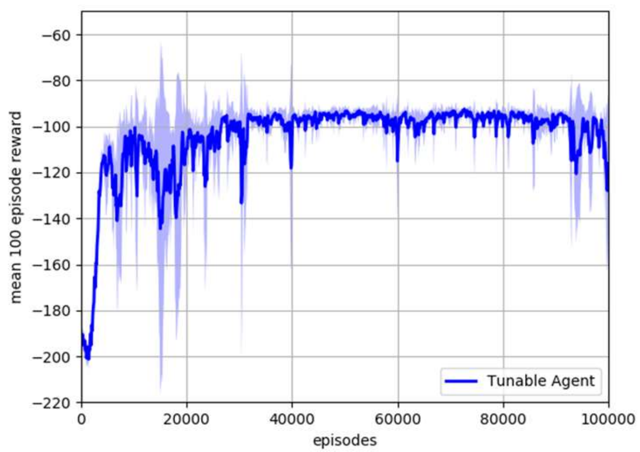
- Example using ***Time*** and ***Safety*** objectives and the DQN algorithm
  - Negative rewards for time and proximity to air defense system
- Observation space
  - Relative distance and direction of air defense and target in last 8 time steps
- Action space
  - Left/right turns with various load factor (2-4g), in discrete steps
- Trained policies
  - Fixed policies for various objective preferences
  - Single, tunable policy

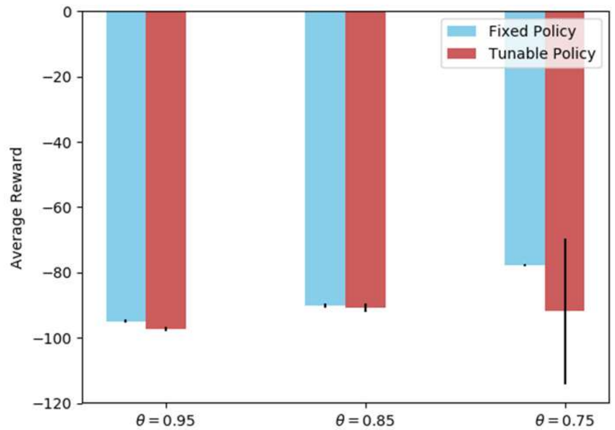


Multi-Objective Reinforcement Learning

Learning progress and relative performance of policies







Conclusions

- Reinforcement learning in simple air combat scenarios
  - Allows agents to learn cooperation
  - Allows agents to learn prioritization among objectives
  - May require many simulations to find good policies
- Directions for future work
  - Study more complex scenarios
  - Study combinations of multi-agent & multi-objective learning
  - Evaluate training value in experiments with manned simulators
- For more information:
  - Read our paper!

Thank you!

Questions?

This work was partially supported by the Swedish Governmental Agency for Innovation Systems (NFFP7/2017-04885), and the Wallenberg Artificial Intelligence, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation