Multi-Agent Reinforcement Learning for Maritime Operational Technology Cyber Security

Alec Wilson[1], Ryan Menzies[1], David Foster[2], Marco Casassa Mont[1], Neela Morarji[1], Esin Turkbeyler[1], Lisa Gralewski[1]

CAMLIS 2023 | 20th October 2023

(BMT[1], ADSP[2])

BMT

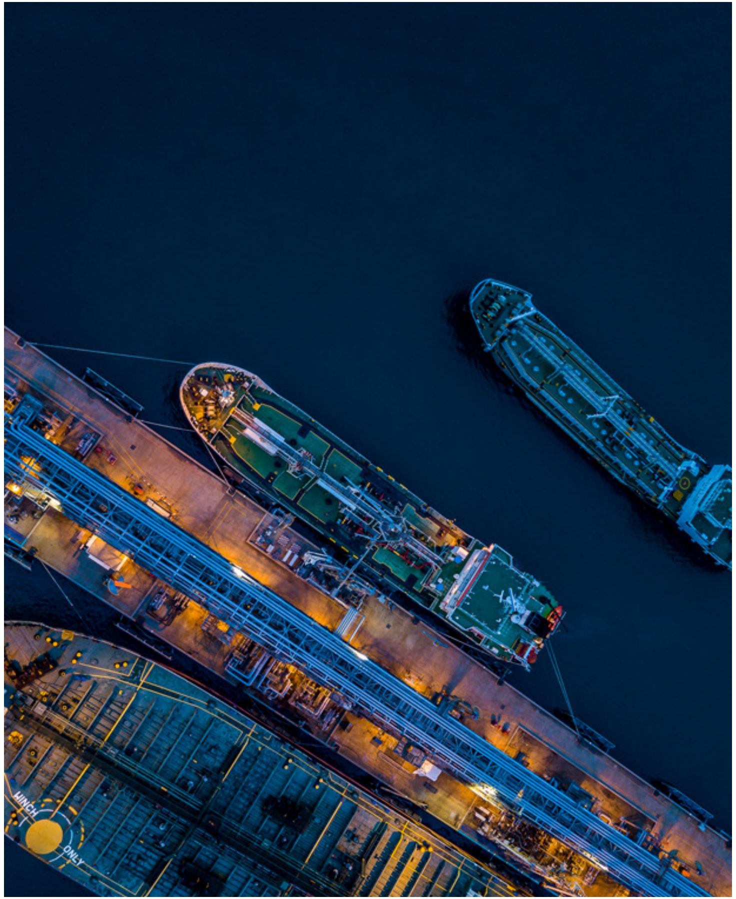
APPLIED DATA SCIENCE

Overview

- Context Setting
- IPMSRL
- IPPO vs MAPPO
- Hyperparameters
- Reward Shaping
- Impact of Partial Observability

Context Setting: Maritime, Vessels and OT

- Vessels are complex systems-of-systems inclusive of Information Technology (IT) and Operational Technology (OT) infrastructures.
- OT systems are vulnerable to cyber-attacks as traditional IT cyber security controls may either not be available or may not be able to prevent attacks.
- OT cyber defensive actions are less mature than for Enterprise IT.
- Cyber security skills and SMEs might not be readily available during vessels' missions and operational activities.



Context Setting: IPMS

- Integrated Platform Management System (IPMS) representation consists of:
  - An abstraction of a bridge (a set of HMIs).
  - A Chilled Water Plant system.
  - A representation of a ships Propulsion system.

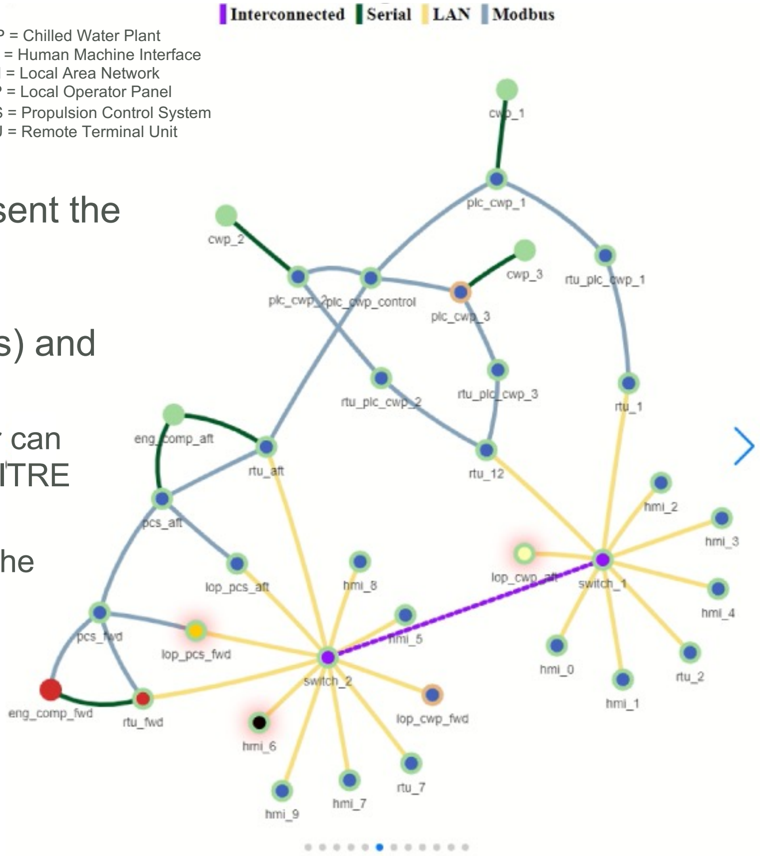
- While the scenario is a high-level abstraction of a real scenario, its design is grounded in reality and exhibits several features which are intended to reflect real challenges when developing agents capable of autonomous cyber responses.

IPMSRL Environment

- IPMSRL – network-based environment where nodes represent the different components within an IPMS.

- The nodes are sub-divided into infectable nodes (e.g. RTUs) and critical nodes (e.g. CWPs). :
  - **Infectable nodes** are nodes in the network that the attacker can spread through and have 12 infection levels based on the MITRE ATT&CK framework.
  - **Critical nodes** are the nodes in the network that represent the critical infrastructure.

Abbreviations: CWP = Chilled Water Plant
HMI = Human Machine Interface
LAN = Local Area Network
LOP = Local Operator Panel
PCS = Propulsion Control System
RTU = Remote Terminal Unit



Key

Intrusion detection alert (red glow)

MITRE ATT&CK

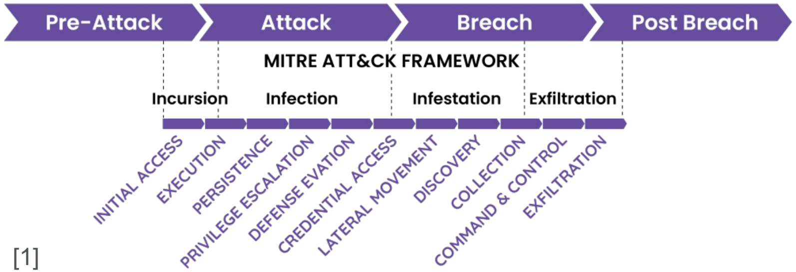
Infection level

not infected 1 2 3 4 5 6 7 8 9 10 11 12

IPMSRL Attacker

- An attacker's virality can be configured on a sliding scale from fully targeted to fully viral:
  - **Fully targeted** - Attackers will always seek to move directly towards critical infrastructure.
  - **Fully viral** - Attackers will move randomly to any adjacent node.

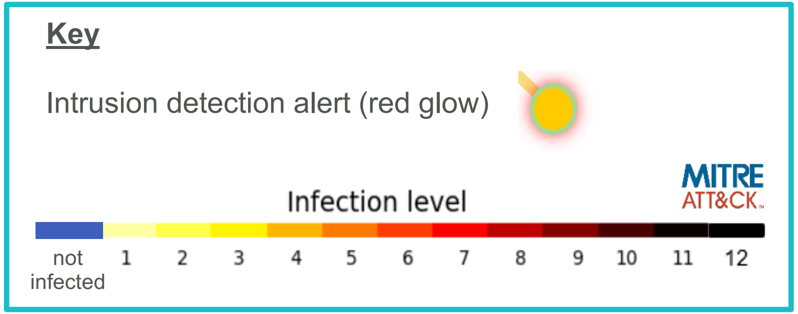
- An attacker's behaviour is also informed by the following parameters:
  - **Lateral Movement Probability -** The probability that a lateral movement is successful.
  - **Infection Progress Probability -** The probability that the infection on any given infected node will progress to a later stage in the MITRE ATT&CK framework.

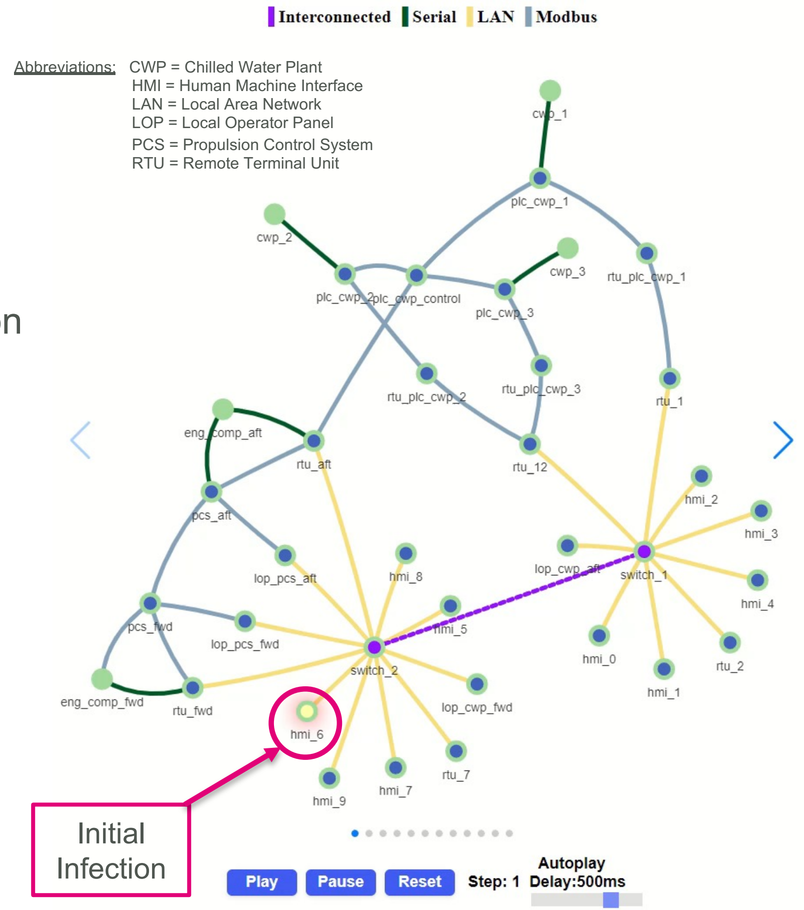


[1]

IPMSRL Attacker Visualisation

- The IPMS quickly becomes overrun, and the propulsion system is taken offline, resulting in a mission failure.





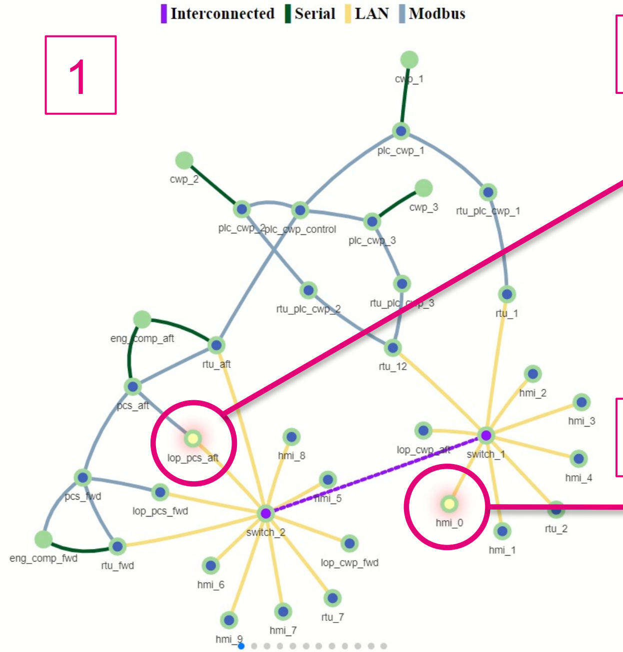
IPMSRL Defender

- Defenders have a restricted view and initially can only see alerts from nodes.
- Defenders collect infection progress information for each infectable node they interact with. This knowledge cannot be shared and is static in nature.
- The actions available to a defender are:
  - **Contain** - Prevents the infection moving laterally from the node;
  - **Eradicate** - Removing an infection from the node;
  - **Recover** - Puts the node back into operational mode;
  - **Wait** - This is a null action that does not result in any change to the environment.

Trained Demo - Walkthrough

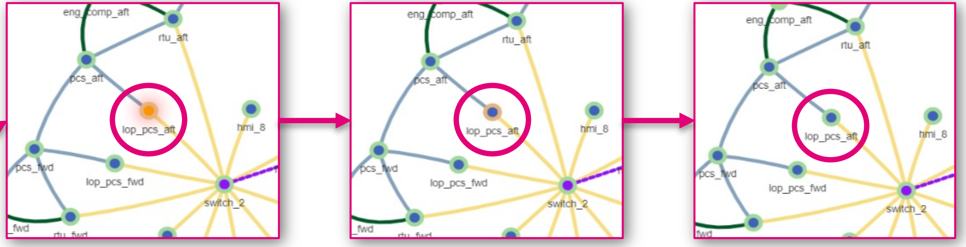
Infection level
not infected 1 2 3 4 5 6 7 8 9 10 11 12

1



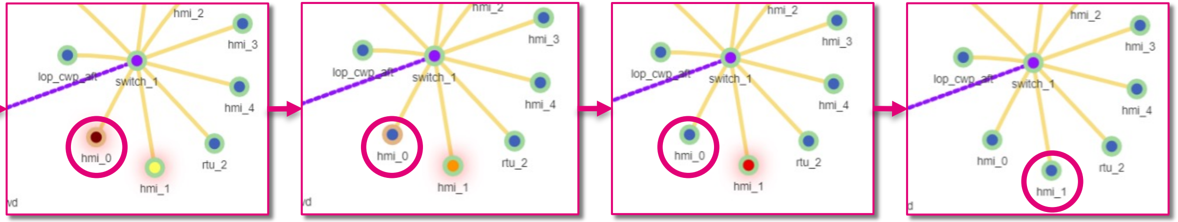
Initial Infection + video of defence

2



Contain, eradicate and recover from infection at propulsion control system operator panel

3



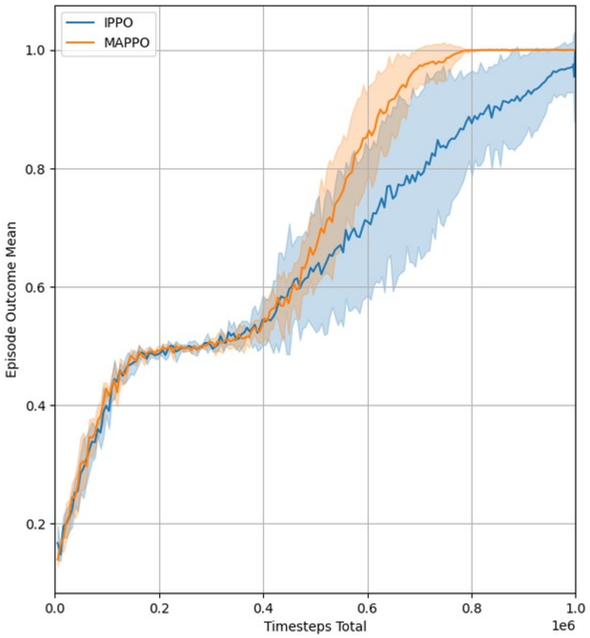
Contain, eradicate and recover from infection at Human Machine Interface (HMI) 0; lateral movement of infection to HMI_1 recovered via same process.

Cyber-attack defeated and full capability restored

Abbreviations: CWP = Chilled Water Plant
HMI = Human Machine Interface
LAN = Local Area Network
LOP = Local Operator Panel
PCS = Propulsion Control System
RTU = Remote Terminal Unit

IPPO vs MAPPO

- We tested Independent PPO (IPPO - independent critics) against multi-agent PPO (MAPPO – single centralised critic).
- MAPPO outperformed IPPO, with all hyperparameters kept constant, showing the benefit of a centralised critic in this instance.
- A centralised critic allows all agents to value the current environment in the same way, leading to faster collaborative efforts.
- Without the centralised critic, each agent must learn a suitable value function independently which results in a slower training process.

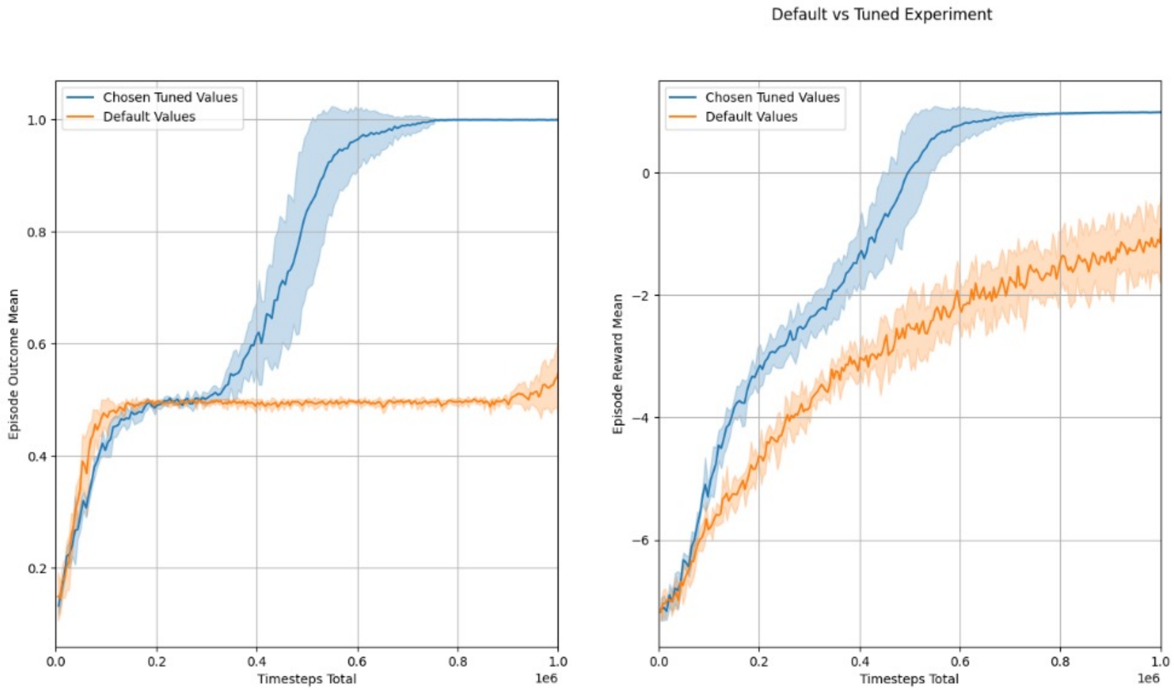


Default vs Tuned Hyperparameters

Tuning hyperparameters was found to be highly important - tuned parameters (blue) vs default PPO parameters (orange)

We tuned 11 hyperparameters in total:

- 3 general RL parameters (train batch size, learning rate, gamma (discount factor).
- 8 PPO parameters (lambda (GAE), KL coefficient, VF clip parameter, SGD minibatch size, num SGD iterations, VF loss coeff, entropy coeff and clip parameter (epsilon).

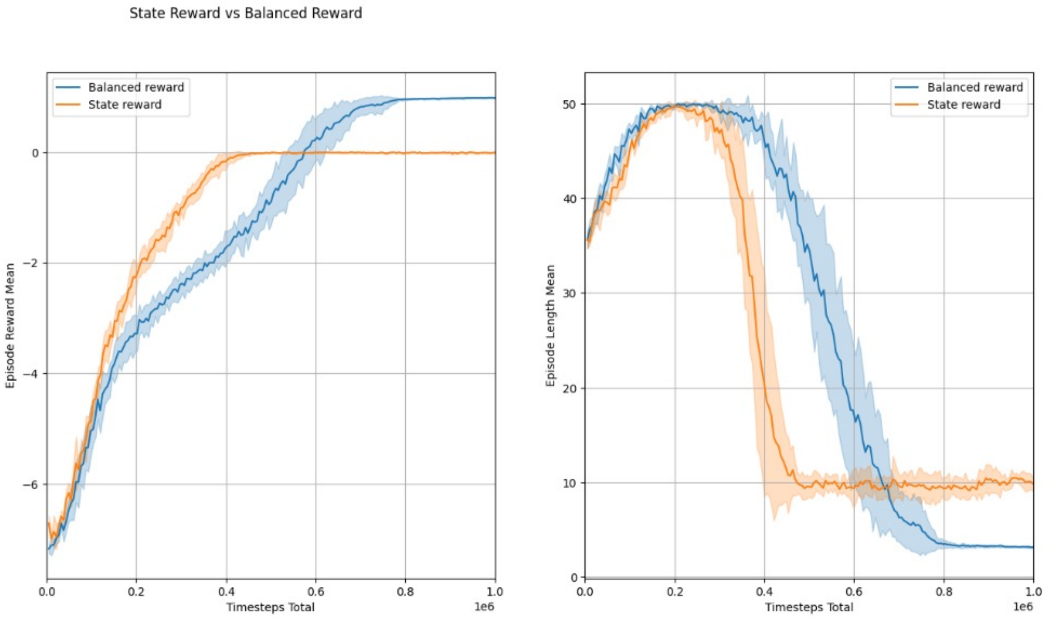


Reward Shaping

- We found that ‘shaping’ the reward function had an impact on the agent finding optimal policies.

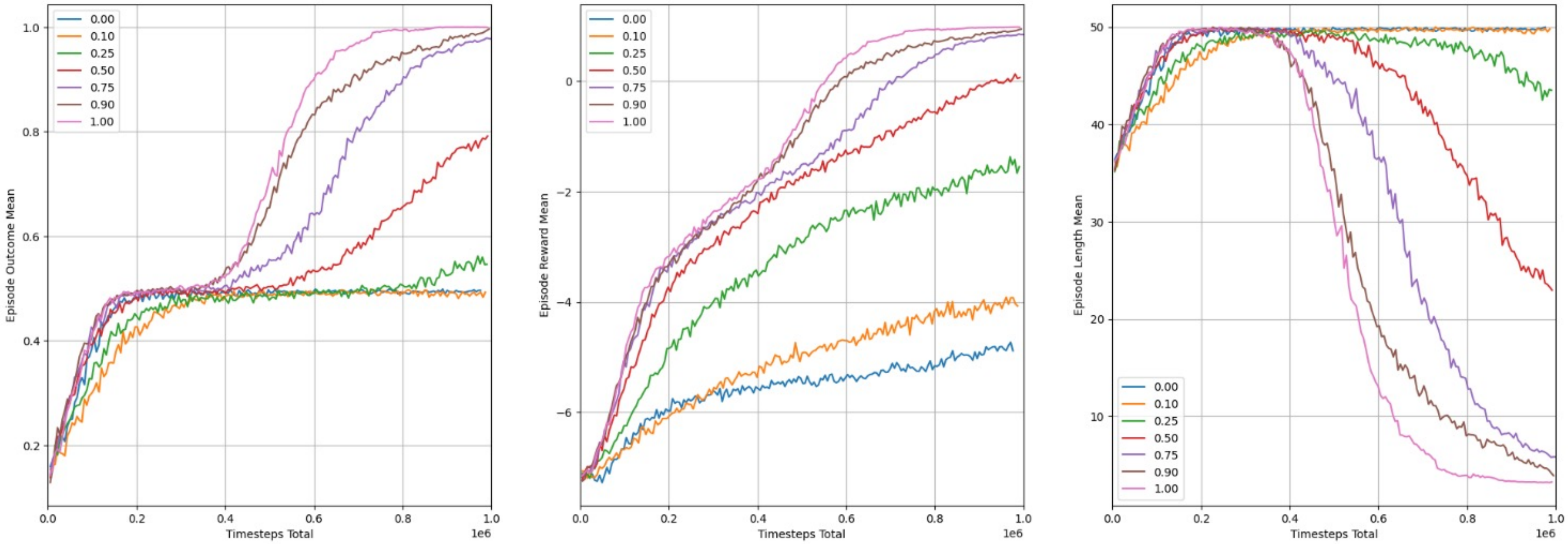
- If the agent was only rewarded on the state of the environment (state reward, orange), there was no incentive to complete the episode quickly, so it was inefficient (~10 timesteps).

- By providing a small negative reward for action taking (balanced reward, blue), the agent was able to find a more efficient policy (~4 timesteps).



Impact of Partial Observability

- We experimented with making the environment partially observable by adjusting the alert success probability.
- We found that after 1m training steps, an agent with only 75% observability could still almost perfectly solve the environment, but that performance dramatically decreased when observability was reduced to 50% and 25%.



Thank You.

Questions?