Reinforcement Learning for Autonomous Resilient Cyber Defence

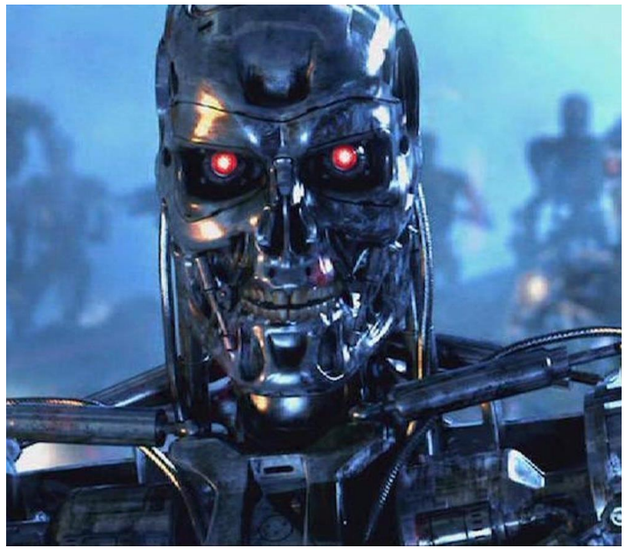
Ian Miles, Sara Farmer

arcd@fnc.co.uk

Frazer-Nash Reference: 016273-146560V

Briefing Contributors

Ian



FRAZER-NASH CONSULTANCY
A KBR COMPANY

Sara

[dstl]

APPLIED DATA SCIENCE PARTNERS

BAE SYSTEMS

BT BMT

cambridge consultants
Part of Capgemini Invent

EXALENS®

QINETIQ

Smith Institute

TRUSTWORTHYAI

Autonomous Resilient Cyber Defence

UK ARCD program

Mission:

- Machine speed cyber response & recovery on military platforms & systems
- Defending IT & OT systems

Goals:

- Understand & demonstrate Autonomous Cyber Defence (ACD)
- Build national skills & knowledge

100+ projects, 4 years

Because

Not enough cyber responders

- Not enough personnel
- No cyber defenders at tactical edge
- Military operator overload

Machine speed attacks

- Volume, velocity, variety

SOAR limitations

- Context awareness, mission awareness

ARCD Ecosystem

Leads

- Defence Science & Technology Laboratory: Customer
- Frazer-Nash Consultancy: ARCD Concepts
- QinetiQ: ARCD Test & Evaluation
- Alan Turing Institute: Fundamental Research

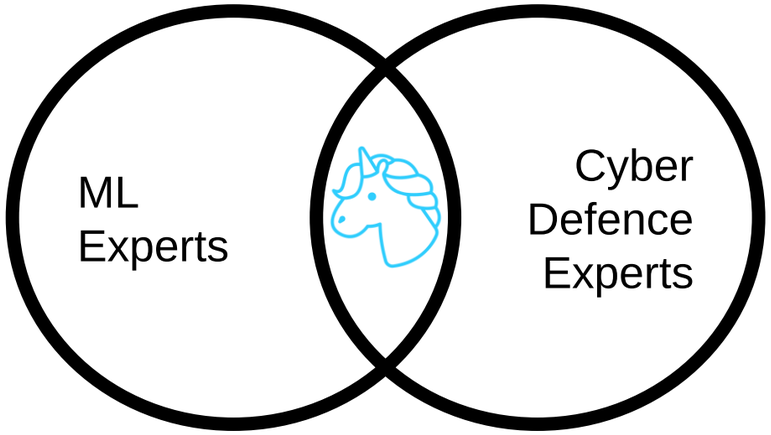
Partnerships

Ministry of Defence

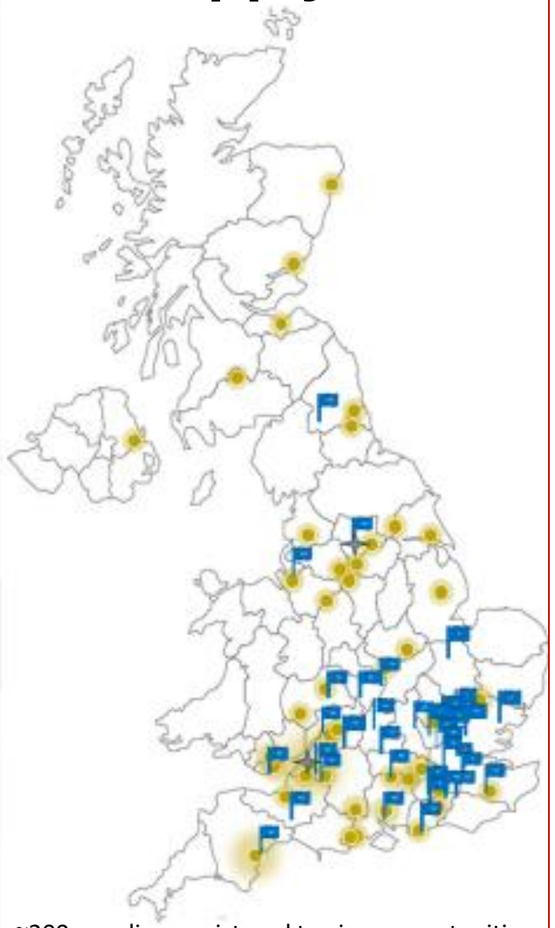
DARPA

National Cyber Security Centre
a part of GCHQ

THE TECHNICAL COOPERATION PROGRAM TTCP



UK Supply Chain



~200 suppliers registered to view opportunities

Unicorn image: www.vexels.com

ARCD Research

Integration

Cyber Threat Detection

Cyber Situational Awareness

Autonomous Machine Speed Response & Recovery

Focus of this Briefing

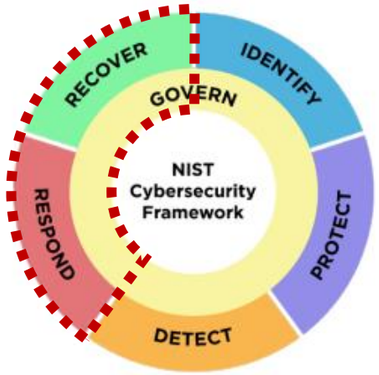


Image: www.nist.gov/cyberframework

Fundamental Research

Governance & Assurance

ACD: Autonomous Cyber Defence

Trains and deploys blue (defense) cyber agents

- Rule-based or probabilistic reasoning

Observing a cyber environment

- Capable of detecting an attack
- Inputs = converted infosec feeds (pcaps etc)

Acting in a cyber environment

- Respond or recover in real time
- Acts, or suggests actions to humans

Autonomous Cyber Operations (ACO) trains both blue and red (attacker) agents

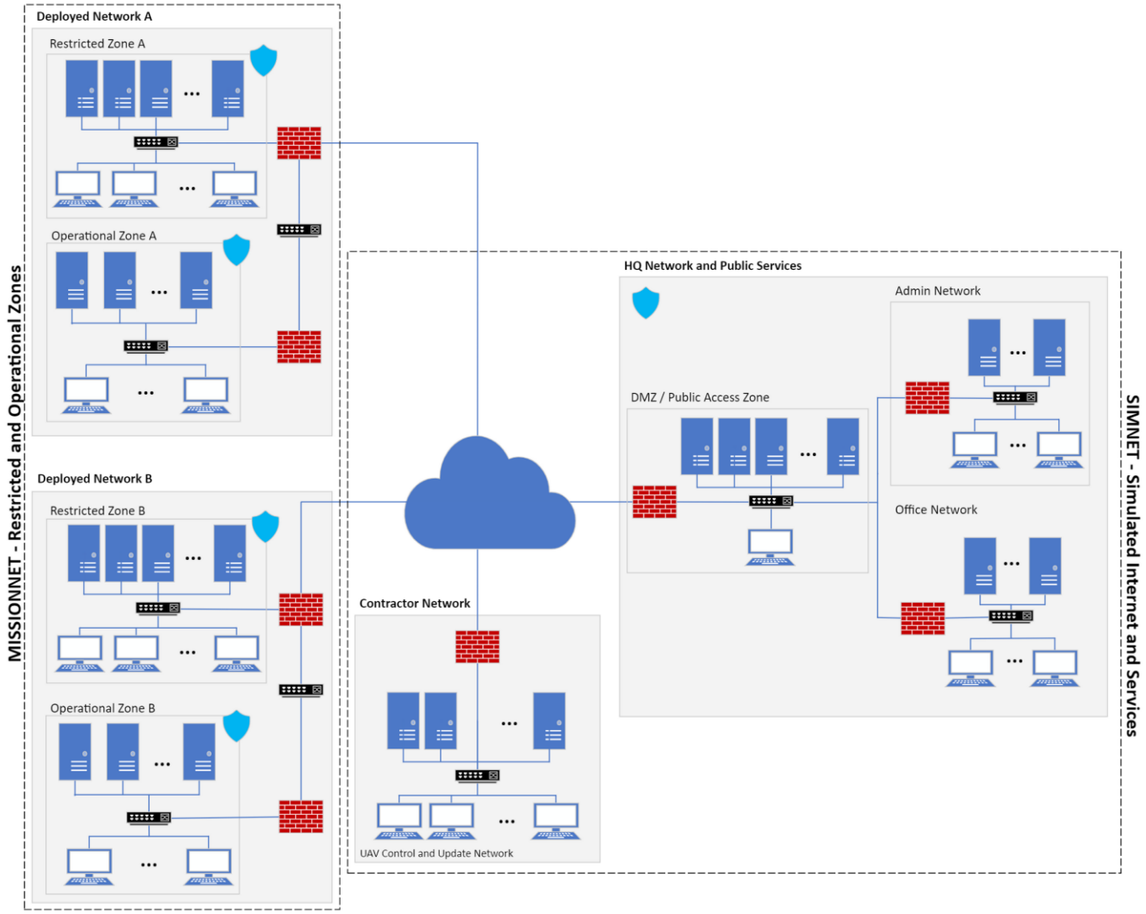


Image: CAGE4 challenge

Training Defence Agents

Learning algorithms

- **RL**: PPO, DQN, DDQN, MARL etc
- **LLMs**
- **Others**: Genetic Algorithms, Graph Neural Networks
- **Combinations:** RL + LLM, GNN, GA, etc.

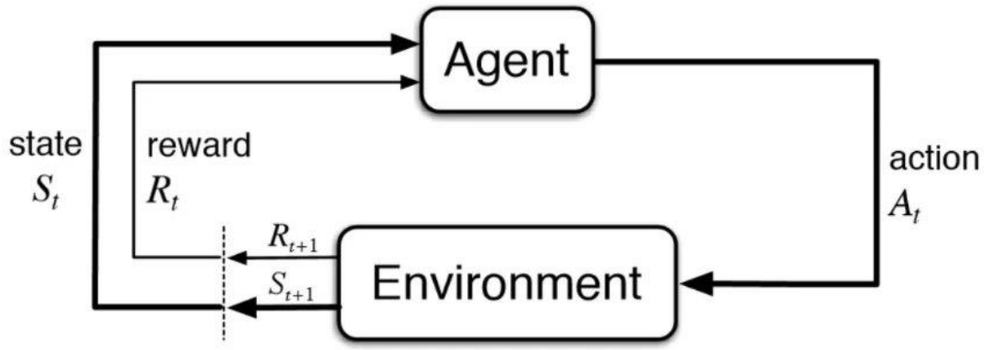


Image: Sutton and Barto

Cyber-specific issues

- **Scale**
- **Partial visibility of state space**
- **Sparse rewards**
- **Needs lots of data**
- **Availability of datasets**
- **Generalisability**
- **Explainability**

PPO = Proximal Policy Optimisation
DQN = Deep Q Networks
DDQN = Double DQN
GA = Genetic Algorithm
GNN = Graph Neural Network
MARL = Multi Agent Reinforcement Learning

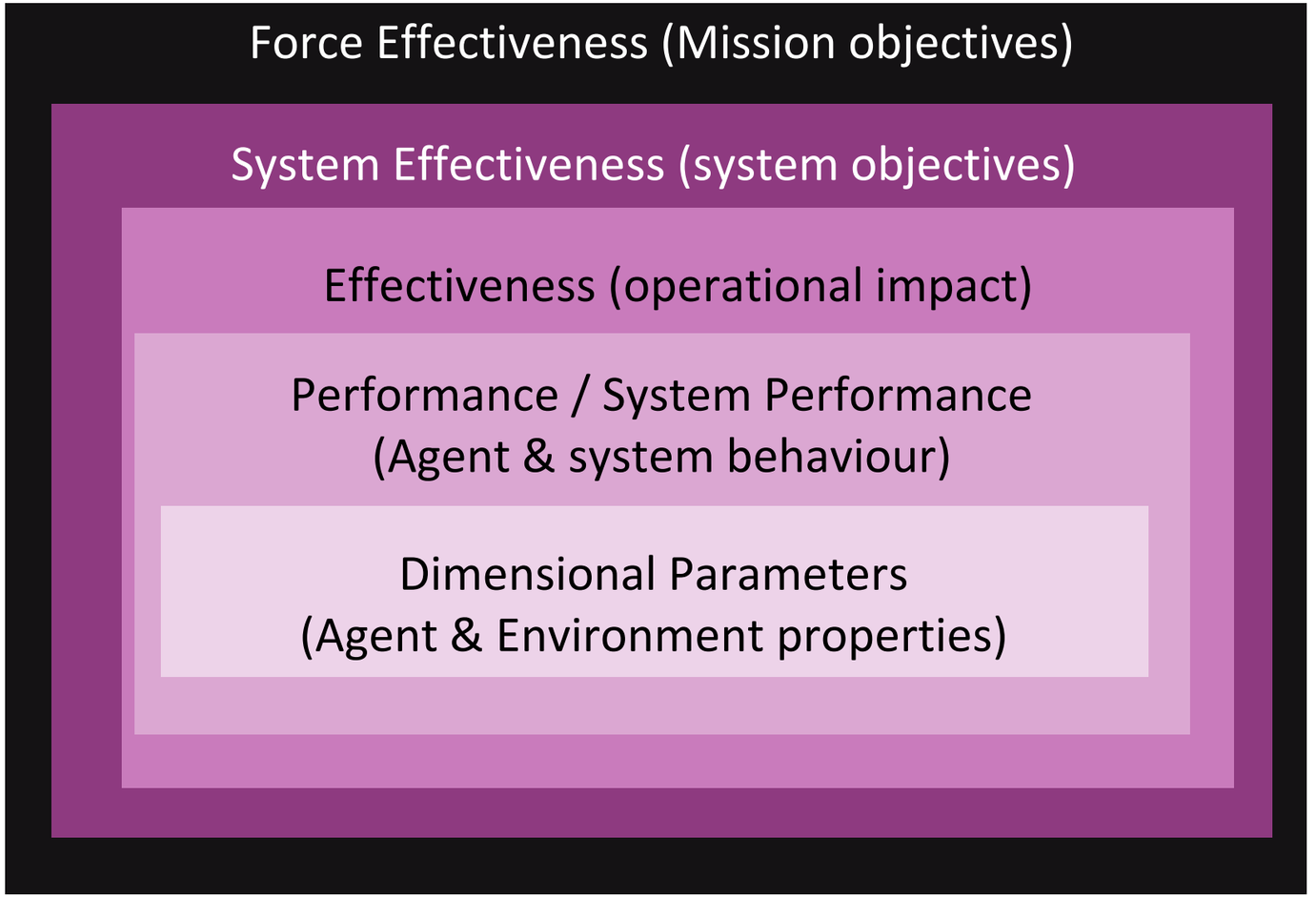
Research question: Meeting "good"

Robustness

- Tractability
- Scalability
- Generalisability

Trust

- Mission-level rewards
- Explainability
- ACD security



Evaluation Lead & Image: QinetiQ

ARCD Environments

ARCD Simulators

- PrimAITE, Yawning Titan, Cyborg (TTCP)

ARCD emulators (cyber ranges)

- Imaginary Yak, PalisAIDE

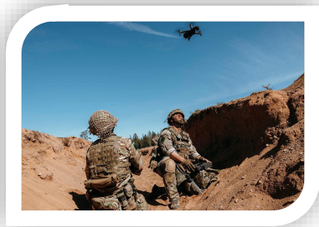
Real world

- IT and OT

Sim2Real: Moving from sim/em to real-world

- Scaling (from 10s to 100s-1000s of nodes)
- Real-world observations
- Real-world actions
- More uncertainty (intrusion detection system etc.)

Low Fidelity → High Fidelity

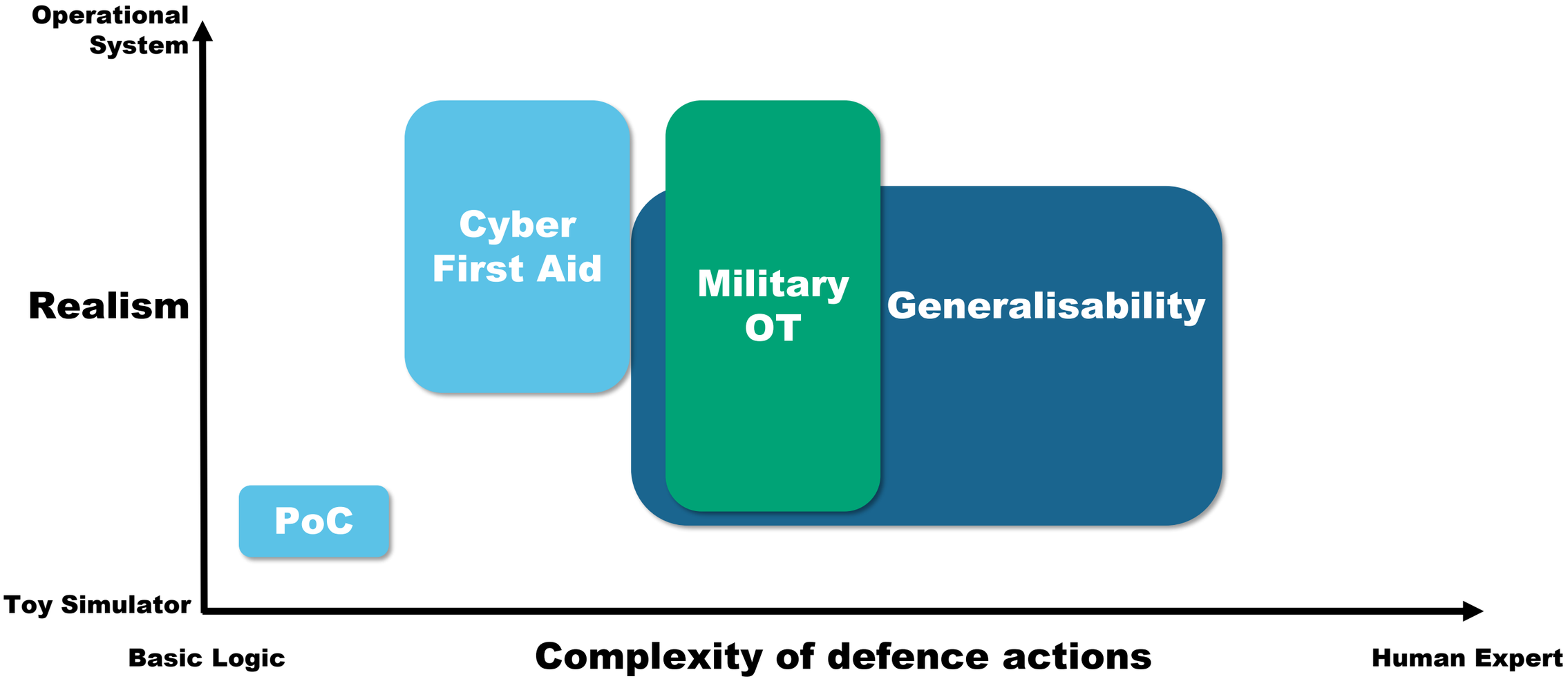


Images: www.husarion.com, www.defenceimagery.mod.uk

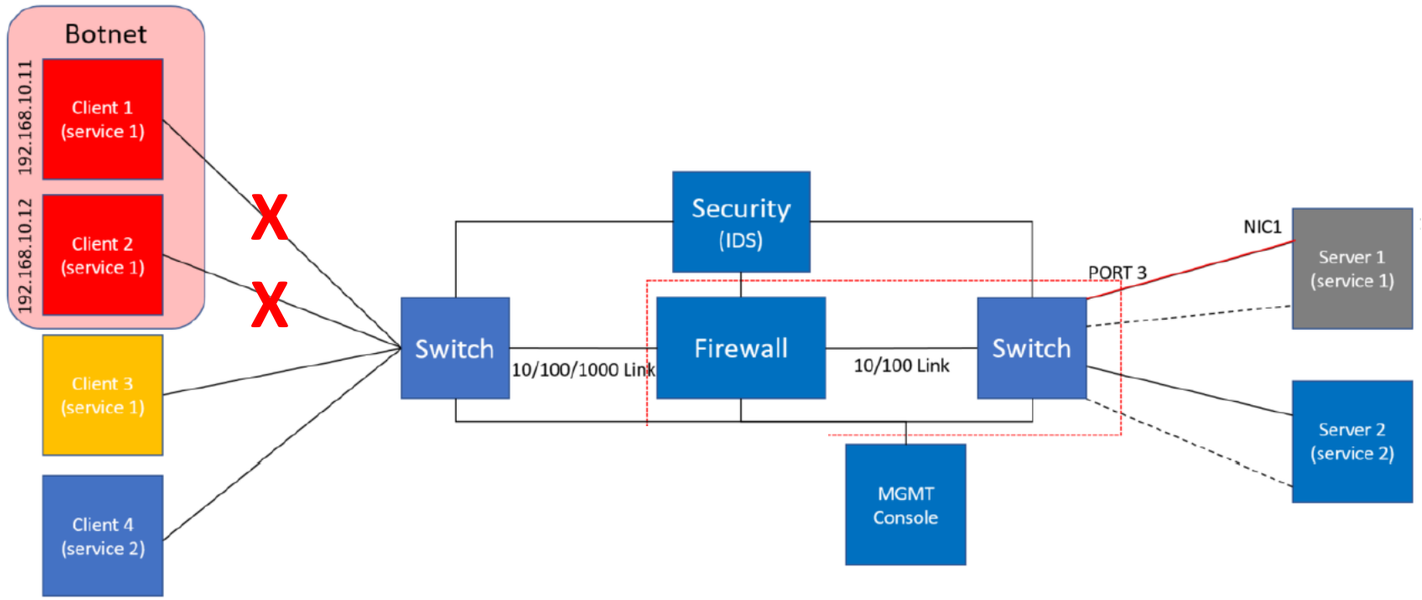
Environments Lead & Logos: QinetiQ

ARCD Demonstrators

Military Demonstrators



Can RL defend a system?



Key Results

- Learnt an appropriate response
- Outscored the rules-based agent (but gamed the scenario)
- Adapted to environment misconfiguration
- Less effort to adapt after environment modifications
- Overfitting – **need more generalised approaches**

Image: QinetiQ, project delivered by Applied Data Science Partners

Can ACD deploy to un-seen networks?

Problem

- RL performs poorly in scenarios not experienced in training.
- Handcrafting large volumes of simulated networks not scalable.

Setup

- GPT4 generated 80 simulated tactical networks (60 for training, 20 for evaluation).
- Deep RL + Graph Neural Networks

Key results & next steps

- PoC - More training networks improved generalisability
- Upgrading to emulated environment with real tooling
  - Red: Cobalt Strike, Blue: Elastic
- Red teaming exercise early 2025

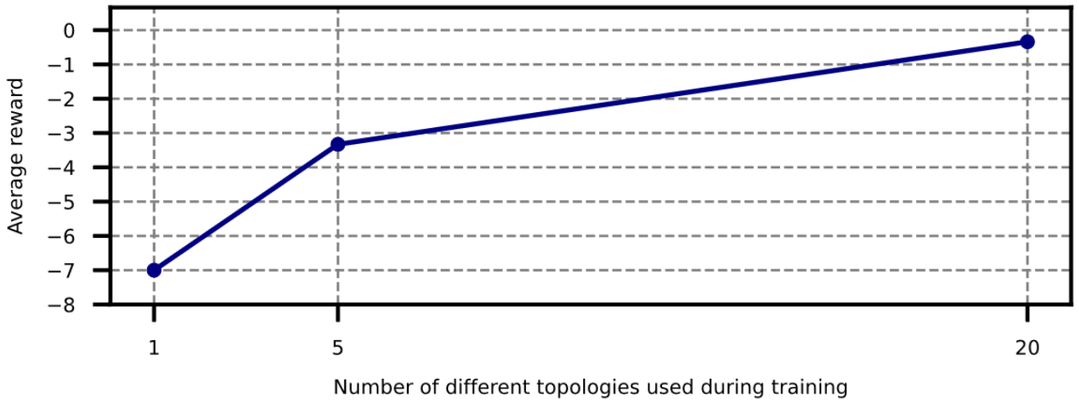


Image & project delivery: Trustworthy AI

Can we build better training adversaries?

Problem

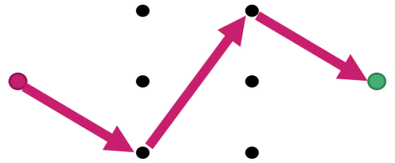
- Poor performance against adversaries not experienced in training.
- Handcrafting large volumes of attack trajectories not scalable or stochastic.

Setup

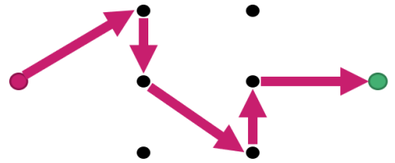
- Create RL-based red agents (to train blue agents)
- Red rewards = stealth, effort, persistence

Key results & next steps

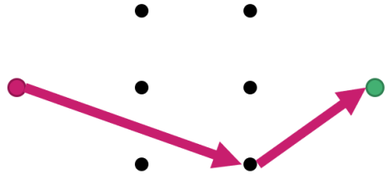
- Training reduced invalid actions and time to target
- Co-evolution to train blue agents



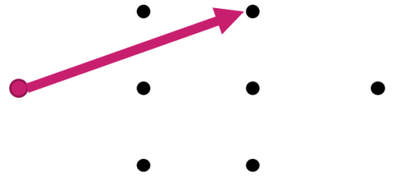
Fastest Time to Target



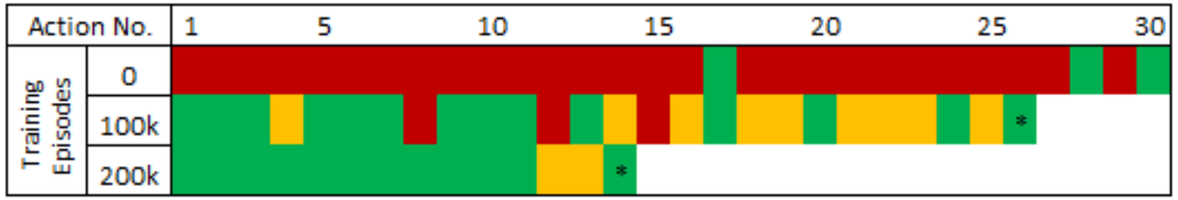
Stealthiest Path to Target



Shortest Path to Target (least exploits)



Stealthy Persistence



Red = invalid action, Orange = duplicate action
Green = valid action, * = reached target

Images & project delivery: BT

Can ACD defend against next gen attackers?

Problem

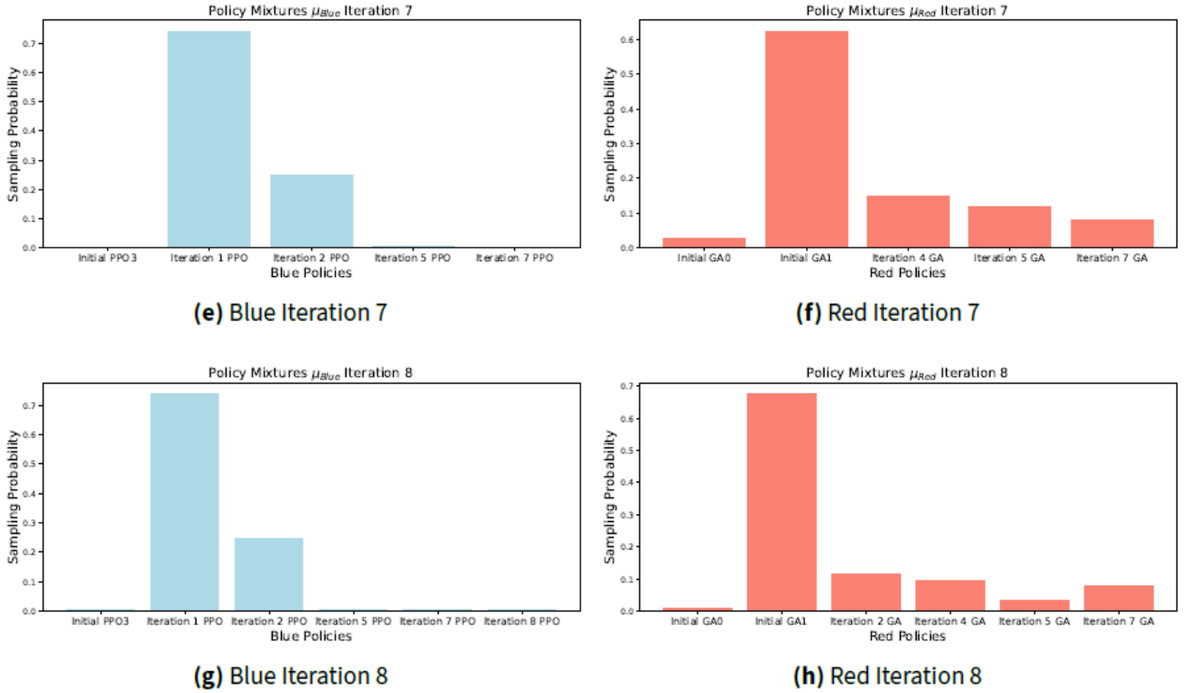
- AI threats may target ACD agents
- Difficult to upgrade ACD agents once deployed

Setup

- Adversarial Learning - Multiple Response Oracles
  - Don’t forget previous adversaries
    (AKA catastrophic forgetting)
  - Defend against novel attacks
  - Risks of underestimating the adversary

Key results & next steps

- Red could not win the game.
- Extending to more complex scenarios (CAGE4)



(e) Blue Iteration 7 **(f)** Red Iteration 7

(g) Blue Iteration 8 **(h)** Red Iteration 8

Images & project delivery: BAE Systems

Does ACD work in a real system?

Problem

- Few cyber experts at the edge

Setup

- Cyber first aid: simple actions, to contain cyber attacks at source & buy time for a human expert
- Train in simulator, deploy to ROSbot

Key results & next steps

- Our first end-to-end demonstration of ACD on a real system (RDP overload DoS). Time to recover <1 second
- Field trials: integration into automated air system

Images: www.husarion.com, www.defenceimagery.mod.uk
Project delivery: Exalens

Will it work for OT?

Problem

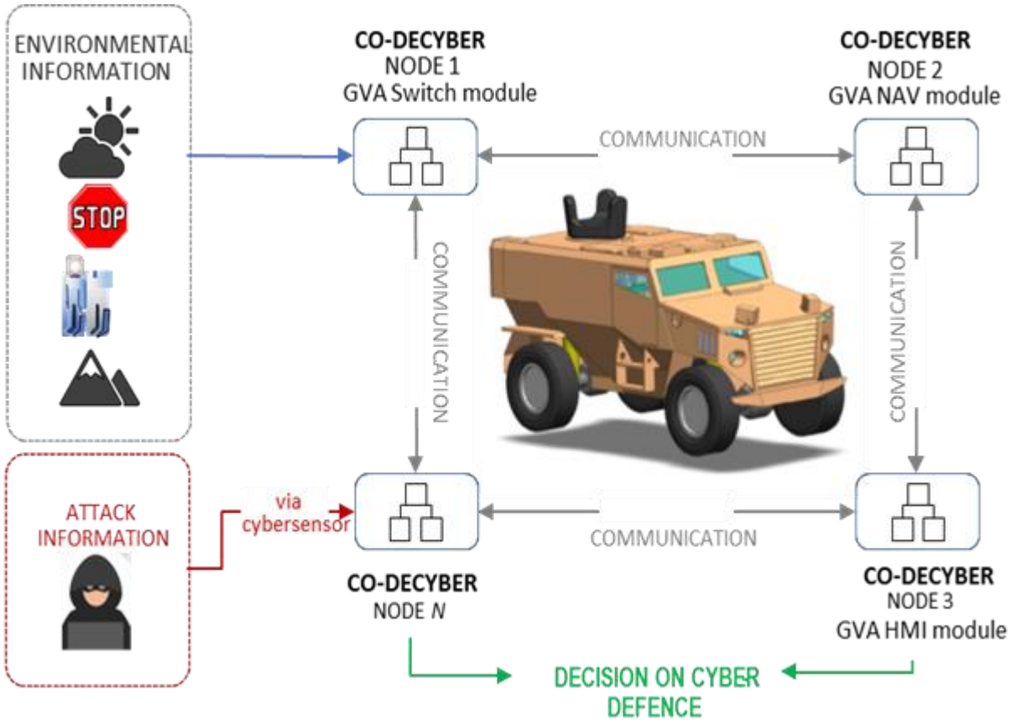
- Semi-autonomous logistics vehicles
  (Manned leader, autonomous follower(s))
- Task-saturated operator with limited cyber expertise

Setup

- Real vehicle architecture (GVA / DDS)
- Multi Agent RL (~30 agents) matching vehicle arch.
- OT action space (power systems, fire alarms, etc.)

Key results & next steps

- Multi-agent RL can defend against simulated false alarms, manipulated GPS messaging and DoS on V2V link.
- Our approach (offline RL) is difficult but supportable
  - MLSecOps processes and flows
- Digital twin opportunity



GVA = Generic Vehicle Architecture
DDS = Data Distribution Service
V2V = Vehicle to Vehicle

Image & project delivery: Cambridge Consultants

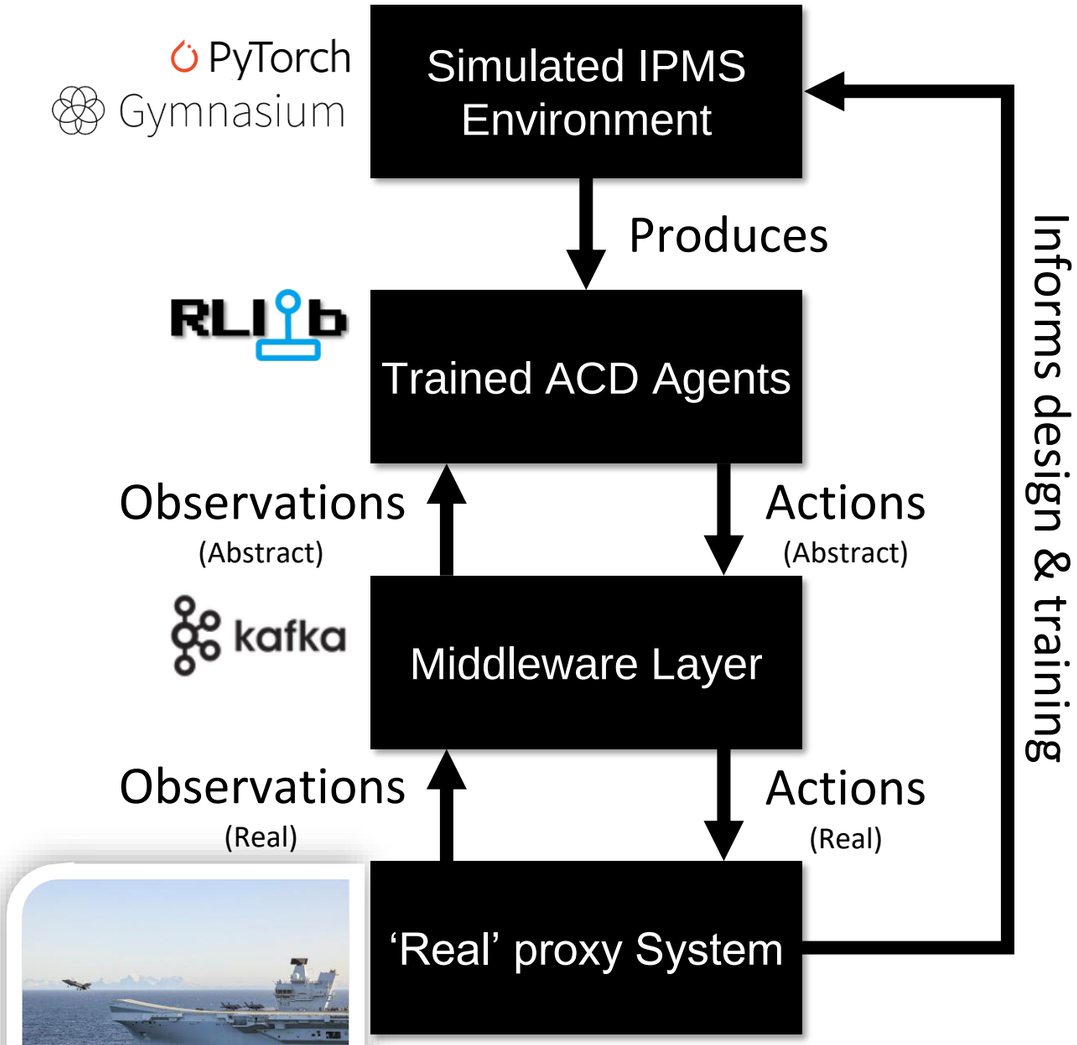
Big OT : Defending Maritime IPMS

Problem

- Integrated Platform Management System (IPMS): Warship's 'brain', ICS using sensor data to control machinery
- Cyber operator overloaded, responds slower
- Uncertain data: false positives, uncertainty of action success

Setup

- IPMS simulator with component interactions
- Varying levels of difficulty
- Multi Agent PPO
- Explainable AI supporting diagnostics
- Deploying to 'real' Proxy system (PLCs, HMIs, software, etc.)



HMI = Human-Machine Interface
ICS = Industrial Control System
PLC = Programmable Logic Controller
PPO = Proximal Policy Optimisation

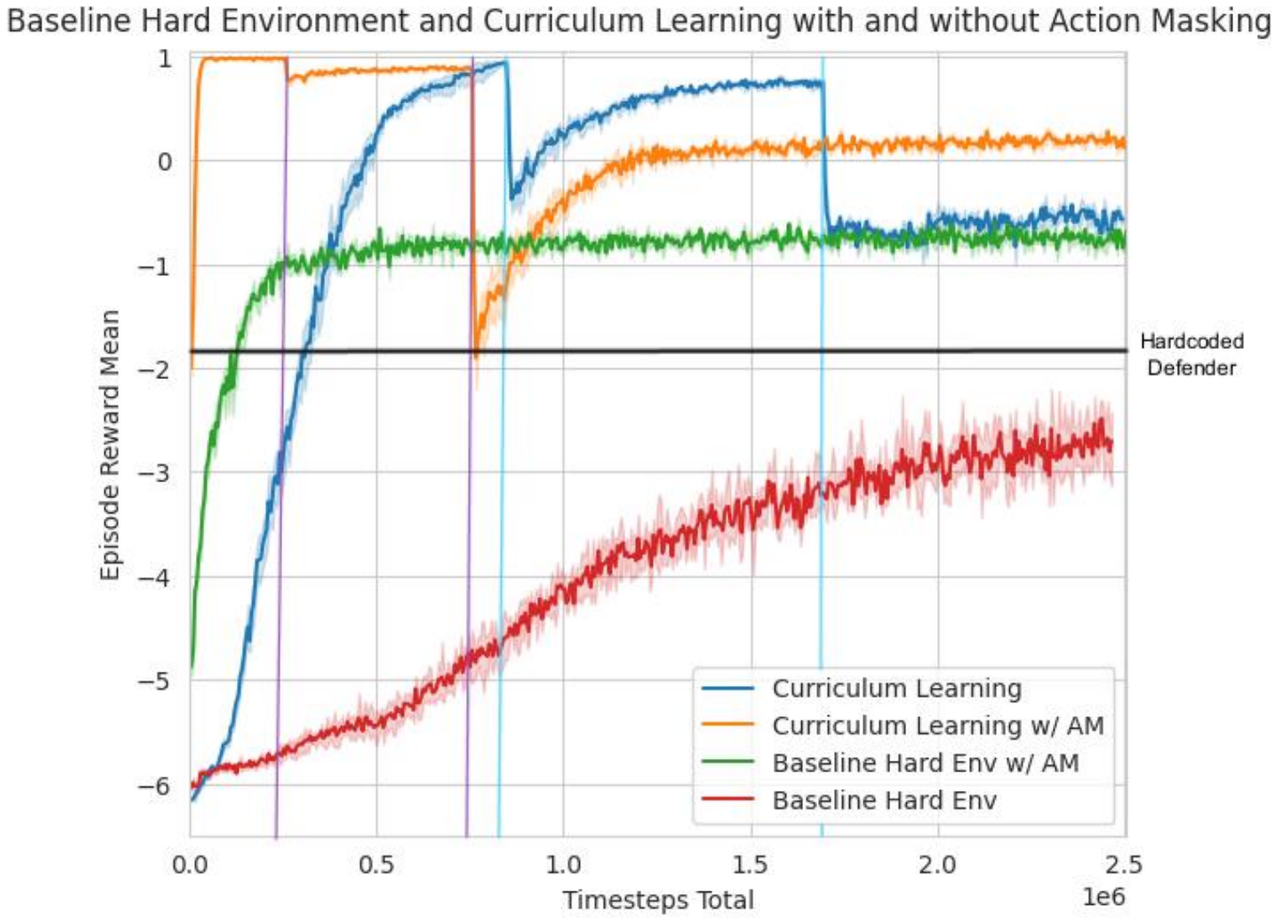
Ship image: www.defenceimagery.mod.uk

Project delivery: BMT & ADSP

Maritime IPMS

Key results & next steps

- Multi-agent defenders out-perform single agents & offer resilience, agents adopted specialist roles
- Struggled to solve ‘hard’ scenarios (red)
  - Alert delays, uncertain false positives/ negatives & action success
- Curriculum learning (blue) & action masking (green) = step change in scalability & exceeds benchmark, combining (orange) compounds benefits
- Distributed architectures - where to put the agents?
- Independent ‘real’ attacks on Proxy



Coloured vertical lines represent switches to a more difficult environment configuration (Easy → Medium → Hard)
Graph shows results for a single agent defender

Image & project delivery: BMT Ltd

Conclusions and Recommendations

Outcomes

Key achievements

- Enhanced UK Cyber/AI and MLSec capability
- Proof of concept – RL works!
- Extended ACD & supporting theory
  - Multiple novel technologies
- End-to-end defence against a 'real' cyber-attack on a 'real' network
  - First reported deployment of ACD to a ‘real’ military OT system

Key results

- RL > rules-based agents, more so complex scenarios
- Multi-agent > single agents & scales
- Generative AI scaling training to enhance robustness
- Consistent requirements for scaling to ‘real’:
  - Action masking
  - Curriculum learning
  - Transfer learning

What's next?

- Increase maturity
  - More realistic & challenging applications
  - Integration with Cyber Situational Awareness tooling
  - Evaluation incl. red teaming & user trials
  - Exploitation routes
- Route to 'Full Auto': Human-Machine Teaming
- Emerging ML approaches (e.g. Foundation Models)
- Open sharing: social good
- International collaboration

Questions for you

- How would you defend against high volume, velocity & variety of cyber attacks?
- Do you have places where human cyber responders aren't available or are limited in capability/capacity?
- If you have an ACD system, have you thought about its vulnerabilities?
- Do you have other use cases for ACD technologies? Training, automated pen test?
- Should you start tracking research on ACD / ACO?
- What did we miss??

Some (ARCD) Light Reading for you

2022 ARCD published papers

- Collyer, "ACD-G: Enhancing Autonomous Cyber Defense Agent Generalization Through Graph Embedded Network Representation", ICML ML4Cyber workshop, 2022
- Andrew, "Developing Optimal Causal Cyber-Defence Agents via Cyber Security Simulation", ICML ML4Cyber workshop, 2022

2023 ARCD published papers

- Kent, "Using a Deep Boltzmann Machine for Reinforcement Learning in Cyber Defence", 7th IMA conference on math in defence and security, 2023. <Talk on quantum RL>
- Little, "Applying machine learning to attribute cyber attacks" ARCD ICD poster, CAMLIS 2023
- Revell, "Can We Trust Autonomous Cyber Defence for Military Systems?" ARCD HRDO poster, CAMLIS 2023
- Gregory, "FNC ARCD Track 1 newsletter", ARCD showcase 2023
- Cheah, "CO-DECYBER: Co-operative Decision Making for Cybersecurity", SECAI 2023 (presentation)
- Wilson, MARL for maritime operational technology security, CAMLIS 2023
- Jeffrey, PrimATE codebase
- Palmer, "Deep reinforcement learning for autonomous cyber operations: a survey", 2023
- Pasteris, "Nearest Neighbour with Bandit Feedback", Neurips 2023

2023 ARCD published papers (continued)

- Hicks, "Canaries and Whistles: Resilient Drone Communication Networks with (or without) Deep Reinforcement Learning", AISEC 2023
- Bates, "Reward Shaping for Happier Autonomous Cyber Security Agents", AISEC 2023
- Pasteris, "A Hierarchical Nearest Neighbour Approach to Contextual Bandits"
- Caron, "Structure Learning with Adaptive Random Neighborhood Informed MCMC", Neurips 2023
- Caron, SBAE, github repo
- Rice, "Digital defenders", Conduit Newsletter, Serapis Framework
- Mavroudis, Adaptive Webpage Fingerprinting from TLS Traces

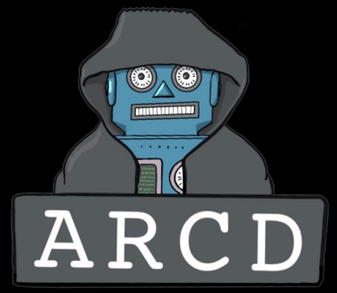
2024 ARCD published papers

- McFadden, Wendigo: Deep Reinforcement Learning for Denial-of-Service Query Discovery in GraphQL, DLSP 2024
- ATI, Mitigating Deep Reinforcement Learning Backdoors in the Neural Activation Space, DLSP 2024
- ATI, Autonomous Cyber Defence: Beyond Games
- **Black Hat USA White Paper [link]**
- More coming!

20+ more research reports exploring evaluations and environments are available by request here: www.qinetiq.com/en/what-we-do/services-and-products/autonomous-resilient-cyber-defence

black hat USA 2024

Thank You



- ARCD Concepts [Frazer-Nash Consultancy]
  www.fnc.co.uk/arcd
  arcd@fnc.co.uk
- ARCD Test & Evaluation [QinetiQ]
  www.qinetiq.com/en/what-we-do/services-and-products/autonomous-resilient-cyber-defence
  ARCD-Track2@qinetiq.com
- AI for Cyber Defence research centre [ATI]
  www.turing.ac.uk/aicd
  aicd@turing.ac.uk
- ARCD GitHub
  https://github.com/Autonomous-Resilient-Cyber-Defence
- CAGE Challenge
  https://github.com/cage-challenge
- DSTL
  arcd@dstl.gov.uk