JAIST Repository
https://dspace.jaist.ac.jp/

Title	Automated Cyber Defense Based on Reinforcement Learning Techniques
Author(s)	Nguyen, Thanh Cong
Citation	
Issue Date	2025-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/20030
Rights	
Description	Supervisor: BEURAN, Razvan Florin, 先端科学技術研究科, 修士 (情報科学)

JAIST
JAPAN
ADVANCED INSTITUTE OF
SCIENCE AND TECHNOLOGY

Japan Advanced Institute of Science and Technology

Abstract

In light of increasingly sophisticated and complex cybersecurity threats, developing autonomous cyber defense agents has become a critical and urgent area of research. Traditional human-based defense systems can no longer cope with the speed and scale of modern attacks. Recent studies show that intelligent agents using artificial intelligence can provide flexible defense capabilities by handling key tasks like monitoring, detection, and threat response. However, current defensive agents are trained in nearly perfect environments that do not mirror real-world conditions. Intrusion detection information can be imperfect and often includes specific error rates. In particular, AI-based anomaly detection models frequently misclassify normal user actions as anomalous.

This thesis focuses on developing cyber defense agents that operate in these imperfect environments through a multi-agent reinforcement learning (MARL) approach. Current research often concentrates on training cyber agents for only one side, either attack or defense, causing agents to focus too heavily on single strategies and restricting their adaptability. To address this limitation, we propose competitive training involving two adversarial agents trained simultaneously, allowing them to learn from and counteract each other in dynamic scenarios.

The research contributes a specialized simulation environment extending the Network Attack Simulator (NASim) with competitive agents and realistic IDS integration. We implement and compare Multi-Agent Proximal Policy Optimization (MAPPO) with centralized training and Independent Proximal Policy Optimization (IPPO) with decentralized training across clean baseline and operational noise scenarios simulating real-world sensor uncertainty.

Our evaluation employs exploitability metrics against worst-case adversaries using specialized cybersecurity metrics: True Block Rate (TBR), Hosts Compromised (HC), and Decoy Interaction Ratio (DIR). Testing encompasses simulation and realistic implementations using Snort IDS, iptables, Docker-based decoys, and Metasploit frameworks.

The results demonstrate significant defensive improvements over unpro-

tected baselines. In clean conditions, MAPPO and IPPO achieved approximately 79% reduction in compromised hosts against random attackers (from 3.8 to 0.8 hosts) while maintaining perfect 100% true block rates. Against sophisticated trained adversaries, both approaches limited compromise to a single host, showing robust defensive capabilities. The deception strategy proved effective with decoy interaction ratios reaching 26-40% against random attackers, successfully redirecting attack attempts from critical assets.

Important algorithmic differences emerged under operational noise conditions simulating realistic IDS false positives. When facing worst-case trained attackers, IPPO demonstrated superior robustness by limiting compromises to approximately two hosts compared to MAPPO's three hosts. IPPO maintained perfect detection accuracy (100% TBR) while MAPPO experienced slight degradation to 96.1% in realistic environments. These findings suggest that IPPO's decentralized training approach provides greater resilience to sensor uncertainty, particularly against sophisticated adversaries trained to exploit observation noise.

Critical validation came through successful simulation-to-real transfer, with performance metrics remaining within 5% between settings, confirming our simulation captures essential cyber defense dynamics. This enables practical training of defensive agents that maintain capabilities when deployed operationally, addressing the significant training-deployment gap challenge in cybersecurity applications.

The comparative analysis reveals that while both approaches perform similarly under ideal conditions, IPPO demonstrates greater robustness under uncertainty with superior stability against sophisticated attacks in noisy conditions. In contrast, MAPPO achieves better coordination through centralized training. These findings demonstrate that competitive training between adversarial agents produces robust defensive capabilities that transfer effectively to realistic environments, providing practical insights for deploying autonomous cyber defense systems that handle real-world complexities while maintaining operational effectiveness against evolving threats.

Keywords: autonomous agents, cyber defense, competitive training.