Adaptive Strategy Learning for Offensive Cyber Agents

Enhancing Cyber Attack Autonomy Through Multi-Agent Reinforcement Learning (MARL)

Graduate

Christoph Robert Landolt

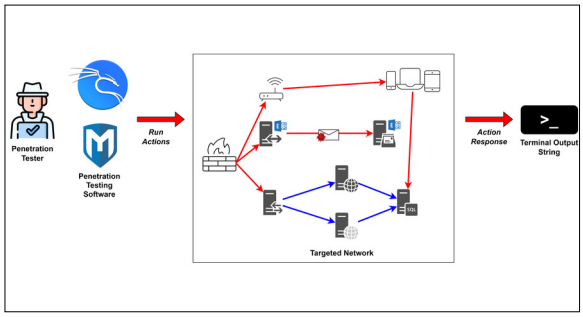
Introduction: Cybercrime is projected to cost the global economy $10.5 billion in 2025, driven by increasingly sophisticated cyber threats that challenge traditional defense mechanisms. Conventional IT security mechanisms are often reactive and static and struggle to counter evolving attacks effectively. To address this, adaptive and proactive defense strategies have earned much attention in the cybersecurity community. Initiatives like DARPA's Cyber Grand Challenge have demonstrated autonomous defense capabilities, but reacting to threats in an adaptive manner remains challenging. Reinforcement Learning (RL) offers a promising solution by enabling agents to learn dynamically and adapt to evolving attack strategies. This thesis focuses on developing Autonomous Intelligent Cyber Agents (AICA) capable of executing advanced cyber-attacks and identifying vulnerabilities in complex networks. Leveraging Multiple RL Agents in a Multi-Agent Reinforcement Learning (MARL) framework aims to develop adaptive agents with strategic decision-making capabilities in high-dimensional, dynamic environments.

Problem: This thesis investigates whether RL can train an attack agent to execute sequential actions and gain control over a targeted system. Different reinforcement policy networks were trained and evaluated on a defined network topology to assess the capability of this agent to perform such attacks. The agent must exploit various systems to reach its target, selecting optimal paths based on its capabilities to minimize costs and maximize rewards. While previous research explored single-agent reinforcement learning for penetration testing, this thesis improves training methodologies, focusing on the agent's ability to generalize across different network scenarios. An existing CyberGym training environment for cyber agents was used to systematically evaluate the agents' adaptability in complex network structures under uncertainty and partial observability. The agent is trained using the Proximal Policy Optimization (PPO) algorithm. The action space includes Scanning, Exploitation, and Privilege Escalation, enabling agents to gather information, exploit vulnerabilities, and escalate privileges for complete system control.

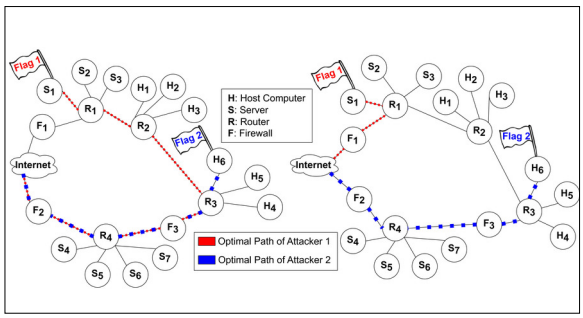
Result: The effectiveness of RL for training autonomous cyber agents by comparing different policy and value network architectures was analyzed in this thesis. This comparison revealed insights into the policy network's action diversity, distribution characteristics, and adaptability. The Self-Attention Policy displayed the highest entropy, indicating greater diversity in decision-making. At the same time, other policy networks heavily relied on specific actions and achieved a low reward in the test scenario.

The thesis underlines the importance of three key components for optimal cyber agent policies: (1) an encoder for effective observation representation, (2) a memory module prioritizing local information, and (3) a controller for action selection. Despite improvements, existing CyberGym environments face limitations, including the sim-to-real gap, lack of modular training scenarios, and insufficient support for adversarial multi-agent interactions. Addressing these limitations through improved CyberGyms and integrating generative models will be essential for advancing robust MARL-based cyber agents.

The attacker uses software tools as an action interface for network attacks, with results displayed as terminal strings.
Own presentment



Left: Attackers share a set of exploits but have different attack targets. Right: Both exploit sets and targets differ.
Own presentment



Heatmap of Self-Attention Policy Network weights shows early exploration and later exploitation in training.
Own presentment

Advisor
Prof. Dr. Christoph Würsch

Co-Examiner
Dr. Julian Jang-Jaccard , Cyber-Defence Campus - CYD, Lausanne

Subject Area
Data Science, Computer Science

Project Partner
Cyber-Defence Campus, armasuisse, Wissenschaft und Technologie, Thun