LLM-Based Controller for Multi-Objective Network Defense Task

Noah Weaver
MBZUAI
Noah.Weaver@mbzuai.ac.ae

Loic Martins
MBZUAI
Loic.Martins@mbzuai.ac.ae

Kadri Mufti
MBZUAI
Kadri.Mufti@mbzuai.ac.ae

Abstract

The rapid advancement of Artificial Intelligence (AI) and its growing ability to perform autonomous tasks have led to widespread adoption across numerous domains, particularly in cybersecurity. In fact, the term "autonomous" has become increasingly prevalent, as current research aims to demonstrate that intelligent agents can independently manage system defense. However, in the field of cybersecurity—and especially in network security—the decisions made by professionals in specific situations often have cascading effects across multiple levels of an organization. Making such decisions requires not only an understanding of the current state of the environment but also a clear awareness of how each action may influence that state. Therefore, the challenge is not merely to design agents capable of stopping attacks, but to develop autonomous systems that comprehend the broader implications and interconnected stakes involved at various organizational levels when defending against threats.

1 Introduction

In the project proposal, we emphasized that network administrators must navigate an inherent trade-off: adopting aggressive defense strategies can effectively block attacks but risk disrupting legitimate traffic, whereas more cautious approaches preserve service quality yet may fail to detect sophisticated threats. Consequently, we formulated the following research question:

How do security-focused versus availability-focused reinforcement learning agents compare in their ability to defend networks against diverse attack patterns while maintaining different operational priorities?

This question was particularly significant for us, as it underscores the challenge of designing and training an agent that not only counters attacks but also for maintaining the stability of the broader ecosystem.

From this research question, we proposed a specific approach based on a comparison between two distinct RL defense philosophies (Security-Focused Agent and Availability-Focused Agent) within a controlled network simulation environment using Mininet Faris Keti [2015].

After careful consideration, we decided to conduct a more in-depth review of the existing literature to refine our problem statement and research approach. Although we retained our original research focus, we expanded our investigation to address the lack of theoretical grounding and the absence of a clear baseline framework.

Preprint. Under review.

2 Related work

3 Literature review

The project seeks to explore how an agent navigates the fundamental tension between security-oriented and availability-oriented objectives. Accordingly, this topic is structured around two key research questions that must be addressed and understood:

- In the context of cyber defense, what are the specific characteristics of this environment, and what constraints do they impose on the agent's behavior and decision-making processes?
- What exactly represents this trade-off?

The body of literature on artificial intelligence for cyber defense—particularly concerning autonomous agents—is rapidly expanding. To better understand the landscape, we reviewed several foundational surveys. Vyas et al. [2023] define Automated Cyber Defense as a domain centered on "automated decision-making agents", while Oesch et al. [2024] highlight the critical importance of agent adaptability. Specifically, such agents must be capable of adjusting to diverse network environments, evolving adversarial strategies, and, crucially, varying operational objectives.

From these surveys, it can be inferred that cyber defense represents a distinct environment with unique constraints, particularly concerning networking. To effectively train and assess autonomous agents, it is essential to accurately simulate this environment. Simulation platforms serve this purpose by recreating, under controlled conditions, the complexity and dynamics of real-world operational contexts. Among the most frequently cited are CybORG Baillie et al. [2020], Emerson et al. [2024], an advanced toolkit for reinforcement learning research in network defense; Mininet, a mature tool for constructing realistic virtual networks; CyGym, developed within the OpenAI Gym framework Lanier and Vorobeychik [2025]; and CyberGym Wang et al. [2025], which emphasizes the replication of real-world scenarios and vulnerabilities.

With a clearer understanding of what cyber defense entails and the ability to simulate such environments, it becomes essential to deepen our understanding of the agent itself—particularly and the fundamental tension between security-oriented and availability-oriented objectives.

This issue has been widely recognized in the literature. For instance, Gu et al. [2022] describe it as a *key challenge that lies in balancing multiple objectives while simultaneously meeting all stringent safety constraints.* Numerous approaches have been proposed to tackle this challenge. In particular, several studies have approached the problem through the perspective of Multi-Objective Reinforcement Learning (MORL) Gu et al. [2022], which enables an agent to simultaneously optimize multiple, potentially conflicting, objectives. Each of these objectives contributes its own gradient. The literature is rich, and researchers continue to advance and refine this approach. In the specific context of cyber defense, Gu et al. [2022] propose a novel natural policy gradient manipulation technique that simultaneously optimizes multiple reinforcement learning objectives and mitigates conflicts among their gradients. Complementarily, O'Driscoll et al. [2025] provide a comparative analysis of two approaches — Multi-Objective Proximal Policy Optimization (MOPPO) and Pareto Conditioned Networks (PCN). In contrast, Molina-Markham et al. [2025] emphasize that "network defense is not characterized by a single task with a fixed set of rules." Consequently, they explore the potential of Open-Ended Learning (OEL) as a means of enabling autonomous agents to acquire a diverse repertoire of behaviors.

3.1 Limitations Identified in the Literature

The challenge of enabling autonomous agents to pursue multiple objectives has been widely investigated in the literature, with most approaches grounded in Reinforcement Learning, and more specifically, in Multi-Objective Reinforcement Learning (MORL). However, O'Driscoll et al. [2025] emphasize that a fundamental challenge in multi-objective reinforcement learning (MORL) lies in its generation of multiple potential policies rather than a single one. Each policy embodies a distinct balance among conflicting objectives, a structure that inherently entails two specific limitations:

- Multi-objective learning produces many policies that are expensive to train and impractical to use simultaneously.
- Fixed policies can't adapt to changing environments.

4 Problem Statement

Our initial research question served as the starting point for our reflection on the trade-off between availability and security. We soon realized that this issue often corresponds to a multi-objective network defense problem. Building on this understanding and acknowledging the associated limitations, we were then able to clearly define the problem statement:

In environments where task priorities and threat levels vary, autonomous agents trained through multi-objective reinforcement learning struggle to identify which policy best satisfies current conditions, leading to suboptimal performance when objectives shift over time.

Therefore, the project aims to address the following research question:

How can an autonomous agent dynamically select the most appropriate policy when operating under uncertain and changing environmental conditions?

5 Proposed Approach

5.1 Overview of the Approach

After identifying several limitations of the MORL approach, O'Driscoll et al. [2025] propose introducing a controller—a higher-level mechanism responsible for selecting the most appropriate policy based on the current context. For instance, if an attack is detected, depending of the state of the system (e.g., intensity of the attack, spread of the attack)the controller would activate a policy that emphasizes security. They thus advocate for a hierarchical framework in which decision-making occurs at multiple levels:

- **Lower level:** The agent (or network) executes a given policy — e.g., a neural network trained by MOPPO.
- **Higher level (controller):** Monitors the environment and decides which policy the agent should follow right now.

The idea is to use a high-level controller — a high-level decision unit — that can:

- Observe the context, interpret the situation (e.g., system state, threat level, mission priorities)
- Decide which policy should be active.
- Issue commands or weightings to the lower-level agent(s).

So, instead of training or running all policies simultaneously, the controller just picks one that fits the current needs.

5.2 LLM-Based Controller

As the controller, we employ a Large Language Model (LLM) due to its ability to interpret complex and ambiguous conditions, generalize beyond its training data, and comprehend and execute mission directives expressed in natural language. In the literature, Large Language Models (LLMs) have already been employed as controllers in various contexts, including control engineering (Control Theory) Zahedifar et al. [2025].

More precisely, we have:

- Input: Network metrics and information about the attack (packet rate, cpu usage, etc.).
- Output: Recommended policies to apply.

5.3 Theoretical Foundations

Our approach is grounded in three key theoretical foundations:

- Network theory and security, focusing on structural properties, interconnectivity, and attack modeling to understand and enhance the resilience of complex systems.

- Controller design, which governs the behavior of agentic systems by managing actions, feedback loops, and decision policies to ensure stability, goal alignment, and safe operation within dynamic environments.
- Large Language Models (LLMs), advanced AI architectures trained on large-scale textual data that enable reasoning, contextual understanding, and adaptive interaction, serving as cognitive components in intelligent control and decision-making frameworks.

5.4 Scope of the project

The range of policies an agent can implement and attacks it can defend against in the context of network security is quite broad. For this project, we will focus on:

- **A specific policy:** Ensuring the availability of the network.
- **A specific attack:** Introducing a DDoS-like attack with controllable parameters such as intensity (packet rate, duration) and spatial distribution (targeted nodes or subnetworks).
- **The controller:** Rather than an autonomous defense agent.

We will have:

- **Input:** Network metrics and information about the attack (packet rate, cpu usage, etc.).
- **Output:** Recommended availability level to set (Continuous percentage) based on observed attack conditions.

5.5 Baseline Methods

At a higher level, the controller's task is closely related to a prediction problem, as it operates with continuous percentage values. Based on this observation, we employ four baseline methods:

- **Linear Regression**
- **Random Forest Regressor**
- **XGBoost Regressor**
- **k-NN Regressor**
- **MLP Regressor (Neural Network)**
- **5 different LLMs without fine tuning**

5.6 Evaluation Metrics

In relation to the experiments and the prediction task, specific evaluation metrics will be employed:

- **Mean Absolute Error:** Average absolute difference. Easy to interpret (in percentage points).
- **Root Mean Squared Error:** Penalizes large errors more. Standard regression metric.
- **R-squared:** Proportion of variance explained.
- **Mean Absolute Percentage Error:** Relative error as a percentage.

6 Experiment

This first experiment is conducted in three main stages:

- **Data Generation**
- **Implementation of three baseline methods**
- **Results / Evaluation**

6.1 Data Generation

The first experiment will utilize the dataset introduced in the CIC-DDoS2019 study Sharafaldin et al. [2019]. This dataset is labeled and includes approximately 80 network flow features extracted using CICFlowMeter. However, it lacks labels indicating the availability status that can be inferred from the network metrics, which poses a limitation for this study.

To determine the recommended availability level (expressed as a continuous percentage) based on observed attack conditions, we employed a heuristic, rule-based scoring approach. This method was chosen for its speed and interpretability by domain experts. The process involves computing a normalized attack severity score from relevant features, which is then smoothly mapped to an availability percentage. Here is the process:

- Separate the data using the corresponding label present in the dataset: "BENIGN" or "ATTACK".
- Compute a "severity score" inside the attack group.
- Normalize the severity score (0 to 1).
- Convert to availability percentage (the target label).

The critical step is calculating the severity score, which quantifies how different a given network flow is from typical (benign) traffic. The reasoning behind this is straightforward:

- Each sample is represented by 80 features that correspond to coordinates in an 80-dimensional space.
- Thus, every data row is a single point in that space.
- Points representing benign traffic tend to cluster closely around a central region — the "normal" zone.
- In contrast, attack traffic points are generally located farther from this region.
- To measure how far each point lies from the benign cluster, we use the Mahalanobis distance, which is similar to Euclidean distance but accounts for correlations between features Hou et al. [2020].

6.2 Implementation of four baseline methods

In this initial experiment, we implement four baseline models that represent fundamental algorithms in Machine Learning:

- Linear Regression
- Random Forest Regressor
- XGBoost Regressor
- k-NN Regressor

After implementing these models, the results obtained are as follows:

????????

7 Future Work and Planning

This project aims to investigate the role of an LLM-based controller in a context of cyber defense which can be use to handle the important number of policies in this domain.

Regarding the project's progression, we have defined several objectives for the coming weeks, focusing on theoretical research and improvements to the experimental setup:

- Continue the literature review, with a particular emphasis on the concept of a Controller within agentic systems and the potential role of large language models (LLMs) in fulfilling this function.

- As for the dataset, although the data are real, the labels have been generated manually. We plan to explore the use of a network simulation environment—such as CyBORG—to obtain additional types of data and to simulate various scenarios.
- Develop a comprehensive strategy for selecting appropriate LLMs and designing the fine-tuning process.
- Depending on the time constraints and network availability, an alternative policy may be employed. An interesting aspect to investigate is whether the Controller can accurately predict labels when operating under two distinct policies.

References

Callum Baillie, Maxwell Standen, Jonathon Schwartz, Michael Docking, David Bowman, and Junae Kim. Cyborg: An autonomous cyber operations research gym, 2020. URL https://arxiv.org/abs/2002.10667.

Harry Emerson, Liz Bates, Chris Hicks, and Vasilios Mavroudis. Cyborg++: An enhanced gym for the development of autonomous cyber agents, 2024. URL https://arxiv.org/abs/2410.16324.

Shavan Askar Faris Keti. Emulation of software defined networks using mininet in different simulation environments. *2015 6th International Conference on Intelligent Systems, Modelling and Simulation*, 2015.

Shangding Gu, Bilgehan Sel, Yuhao Ding, Lu Wang, Qingwei Lin, Alois Knoll, and Ming Jin. Safe and balanced: A framework for constrained multi-objective reinforcement learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47, 2022. ISSN 1573-7454.

Yubo Hou, Zhenghua Chen, Min Wu, Chuan-Sheng Foo, Xiaoli Li, and Raed M. Shubair. Mahalanobis distance based adversarial network for anomaly detection. *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

Michael Lanier and Yevgeniy Vorobeychik. Cygym: A simulation-based game-theoretic analysis framework for cybersecurity, 2025. URL https://arxiv.org/abs/2506.21688.

Andres Molina-Markham, Luis Robaina, Sean Steinle, Akash Trivedi, Derek Tsui, Nicholas Potteiger, Lauren Brandt, Ransom Winder, and Ahmad Ridley. Training rl agents for multi-objective network defense tasks, 2025. URL https://arxiv.org/abs/2505.22531.

Ross O'Driscoll, Claudia Hagen, Joe Bater, and James M. Adams. Multi-objective reinforcement learning for automated resilient cyber defence, 2025. URL https://arxiv.org/abs/2411.17585.

Sean Oesch, Phillipe Austria, Amul Chaulagain, Brian Weber, Cory Watson, Matthew Dixson, and Amir Sadovnik. The path to autonomous cyber defense. *IEEE Security Privacy, 2024*, 23:38–46, 2024.

Iman Sharafaldin, Arash Habibi Lashkari, Saqib Hakak, and Ali A. Ghorbani. Developing realistic distributed denial of service (ddos) attack dataset and taxonomy. *2019 International Carnahan Conference on Security Technology (ICCST)*, pages 1–8, 2019.

Sanyam Vyas, John Hannay, Andrew Bolton, and Professor Pete Burnap. Automated cyber defence: A review, 2023. URL https://arxiv.org/abs/2303.04926.

Zhun Wang, Tianneng Shi, Jingxuan He, Matthew Cai, Jialin Zhang, and Dawn Song. Cybergym: Evaluating ai agents' cybersecurity capabilities with real-world vulnerabilities at scale, 2025. URL https://arxiv.org/abs/2506.02548.

Rasoul Zahedifar, Sayyed Ali Mirghasemi, Mahdieh Soleymani Baghshah, and Alireza Taheri. Llm-agent-controller: A universal multi-agent large language model system as a control engineer, 2025. URL https://arxiv.org/abs/2505.19567.

A Appendix / supplemental material

Optionally