AI-driven Intrusion Detection Systems in SDN-Based IoT Networks: Evaluating and Defending Against Adversarial Attacks

Manlaibaatar Tserenkhuu

Cyber Resilience Laboratory

Professor Youki Kadobayashi

Abstract

The Internet of Things (IoT) has been continuing to expand alongside digital devices to comprise physical and virtual things with networking capabilities. The IoT provides new possibilities across various fields, from homes to industry, by enabling new services, increasing productivity, and supporting daily life. Behind such growth, managing networks within conventional architectures is becoming increasingly challenging and inefficient because a large number of heterogeneous devices and things need to connect to the network and transfer a massive volume of data through the network infrastructure. Therefore, software-defined networking (SDN) is considered an appropriate solution for IoT network management, and together these technologies are shaping today's Internet.

However, cybersecurity attacks, including both existing and emerging ones, regularly threaten SDN-based IoT networks, underscoring the necessity for effective countermeasures. Recently, artificial intelligence (AI)-driven intrusion detection system (IDS)s have been applied to such networks to identify malicious traffic by analyzing network traffic patterns. Despite their high potential to distinguish between normal and attack classes, these systems encounter limitations, such as the complexity of the model architecture and the quality of the training dataset, and they become an attack surface themselves due to adversarial attacks, raising issues—increased computational overhead and degraded detection capability over time.

This thesis presents an efficient and adversarial-resilient IDS framework to detect existing and previously unseen cyberattacks for SDN-based IoT networks. The proposed IDS framework leverages three well-known deep learning algorithms with benchmark datasets and our own created dataset. During the development of the IDS framework, the first work focused on creating efficient IDSs by optimizing the model architecture through hyperparameter tuning and reducing computational

overhead via feature reduction using explainable AI (XAI)-based feature selection techniques with domain-constrained features while maintaining high performance (a top performance with an accuracy of 99.9% and 98% on InSDN and X-IIoTID dataset, respectively). The following two works address the vulnerability of the IDS framework to adversarial attacks, which basically represent new variants of attacks or sophisticated versions of existing ones in real-world networks. We first evaluate the robustness of AI-driven models using a proposed adversarial attack approach that integrates hybrid-space transformation. Afterwards, to address the identified vulnerabilities, an innovative adversarial defense method that incorporates a novel data reduction module is developed and integrated into the IDS framework, improving the model robustness against adversarial attacks while retaining minimal computational overhead. The experimental results demonstrated that the proposed defense method achieves a 53.5% memory savings contrasted with the conventional method, and retraining time was accelerated by an average of 60.64% across IDS models. Designing and developing the AI-driven IDS framework in terms of architectural and dataset optimization and adversarial resilience offers significant advancement on the security of SDN-based IoT networks.