Enhancing GNN-based IDS through Open Set Recognition

Thanh-Tung Nguyen, Minho Park

Soongsil Univ.

thanhtungnguyen@soongsil.ac.kr, mhp@ssu.ac.kr

오픈 세트 인식을 통한 GNN 기반 IDS 향상

Thanh-Tung Nguyen, 박민호

숭실대학교

Abstract

Intrusion Detection Systems (IDS) are essential to cybersecurity, providing early alerts for potential threats. However, traditional IDS—especially those using conventional machine learning—often fail to detect novel attacks, particularly in closed-set training environments. To address this, we propose a novel approach that integrates Graph Neural Networks (GNNs) with Open Set Recognition (OSR), enhancing the detection of unknown threats and improving the classification of known attacks.

Ⅰ. Introduction

In today's digitally connected environment, the widespread adoption of internet technologies has significantly expanded and increased the complexity of network infrastructures. While these advancements offer considerable benefits in terms of connectivity and efficiency, they also introduce a broad spectrum of security vulnerabilities that pose serious risks to organizations. To mitigate these threats, Network-based Intrusion Detection Systems (NIDS) have become a fundamental element of modern cybersecurity architecture. These systems operate by continuously analyzing network traffic to identify suspicious or malicious activities, utilizing packet-level inspection and traffic pattern analysis to detect threats such as malware, denial-of-service attacks, and unauthorized access attempts. However, the core detection techniques employed by NIDS—primarily signature-based and anomaly-based methods—face growing limitations.

Signature-based NIDS operate by comparing observed network traffic against a database of known attack patterns or signatures. While this method is highly effective in identifying previously encountered threats with high accuracy and low false positive rates, it struggles to detect new, unknown, or obfuscated attacks, making it less suitable for dynamic and evolving threat landscapes. In contrast, anomaly-based IDS establishes a model of normal network behavior and flag deviations from this baseline as potential threats. Although this approach can identify novel and zero-day attacks, it often suffers from high false positive rates due to the difficulty in precisely defining normal behavior in complex and heterogeneous network environments.

Despite their importance, conventional NIDS like signature-based IDS or anomaly-based IDS limit their effectiveness against previously unseen or evolving cyberattacks. This inherent limitation renders them inadequate in defending against zero-day vulnerabilities and advanced persistent threats. As cyber threats grow more sophisticated and dynamic, there is a critical need for advanced, adaptive solutions. In this context, machine learning-based NIDS (ML-based NIDS) offer a promising path forward. By learning from historical network data, these systems can generalize to new attack vectors and adapt to changing threat environments, thereby

providing more effective, scalable, and intelligent intrusion detection capabilities than traditional approaches.

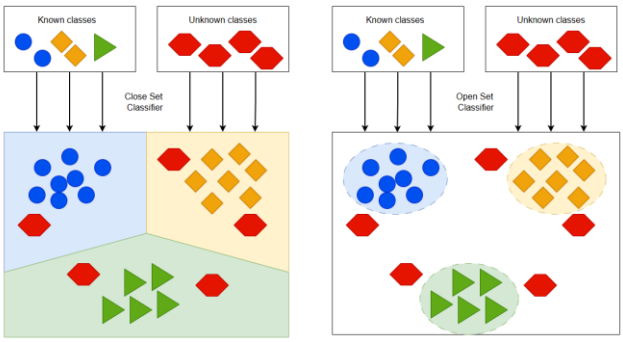

Fig 1. Open Set Recognition example

ML-based NIDS [1] presents a promising solution by leveraging historical network data to generalize to novel attack vectors and adapt to evolving threats, offering more scalable and intelligent detection than traditional methods. Deep learning [2], particularly Graph-based IDS (GNN-based IDS) [3], has shown effectiveness in modeling the complex relationships inherent in network traffic. GNNs exploit node connectivity to capture relational structures, making them well-suited for IDS tasks. However, their performance degrades in open set scenarios, as they are typically trained under closed set assumptions. Consequently, GNNs struggle to identify unseen attack classes, often misclassifying them as known threats, which undermines detection reliability in real-world, dynamic environments.

Open Set Recognition (OSR) [4] addresses the limitations of traditional closed set models by enabling the detection of previously unseen classes—crucial for dynamic domains like cybersecurity. In GNN-based IDS, OSR enhances performance by allowing models to differentiate between known and unknown attacks. While GNNs effectively capture structural patterns in network traffic, their closed set assumption leads to misclassification of novel threats. Integrating OSR into GNN-based IDS improves both accurate classification of known attacks and rejection of unfamiliar or anomalous patterns.

This paper proposes an enhanced GNN-based IDS by incorporating OSR techniques. The integration of OSR equips the system with improved robustness and adaptability, offering a more effective solution for handling real-world cybersecurity challenges. In the remainder of this paper, we present a foundational overview of OSR techniques and introduce a novel enhancement to the existing GNN model. The paper concludes by outlining potential avenues for future research.

Ⅱ. Method

A. Open Set Recognition

OSR is a machine learning technique aimed at addressing scenarios where a model encounters previously unseen or unknown classes during inference. Unlike traditional closed-set classification, where all possible classes are known and well-represented in the training data, OSR assumes the presence of novel instances that do not belong to any of the trained classes. This assumption is critical in dynamic environments where new or rare categories may emerge unpredictably. OSR methods aim to not only correctly classify known classes but also to identify and reject unknown inputs. The challenge lies in achieving a reliable balance between discriminating among known categories and recognizing when an input deviates significantly from the learned class distributions.

B. Proposed methods

In this paper we proposed the integration of a discriminative approach, OpenMax into an existing GNN-based IDS model. OpenMax is a prominent open set recognition technique that modifies the traditional softmax layer to better handle unknown inputs. It operates by estimating the probability that a given sample belongs to an unknown class based on its distance to the known class centroids in the feature

space. OpenMax replaces the softmax layer with a recalibrated output layer using the Weibull distribution to model the tail of the distance distribution for each known class. This statistical modeling enables OpenMax to assign higher confidence scores to unknown samples, effectively separating them from known classes. By augmenting classification with a rejection option, OpenMax significantly improves robustness in scenarios where the test data includes out-of-distribution inputs.

The principles of OpenMax—distance-based modeling and probabilistic rejection—can be adapted to GNNs for open set recognition in graph-structured data. GNNs, widely used in IDS for their ability to capture complex dependencies in network traffic, learn expressive embeddings where distance metrics are effective. Incorporating OpenMax into the GNN pipeline involves computing class centroids in the embedding space and applying a Weibull distribution to sample distances. During inference, samples far from all centroids are probabilistically flagged as unknown intrusions.

This adaptation is well-suited for IDS, where evolving attack strategies introduce novel, unseen threats. Integrating OpenMax-inspired open set capabilities into GNN-based IDS enhances adaptability and resilience by enabling the rejection of inputs that deviate from known attacks or normal patterns, even without prior exposure. This improves generalization and strengthens security, making IDS more effective in real-world, open-world environments.

Ⅲ. Conclusion

In conclusion, this paper proposes a novel approach to enhancing GNN-based IDS by integrating the OpenMax algorithm for open set recognition. Recognizing that traditional IDS models often assume a closed set of known attack types, which limits their ability to detect novel intrusions, the proposed method leverages OpenMax to introduce a rejection mechanism based on statistical distance modeling. This approach represents a promising new direction for enhancing the performance of IDS systems within increasingly complex network environments. However, developing effective models will require the design of robust methods for selecting representative nodes, which remains a key challenge for future research.

ACKNOWLEDGMENT

This work was jointly supported by the National Research Foundation of Korea (NRF) via a grant provided by the Korea government (MSIT) (grant no. NRF-2023R1A2C1005461), and by the MSIT (Ministry of Science and ICT), Korea, under the Convergence security core talent training business support program (IITP-2024-RS-2024-00426853) supervised by the IITP (Institute of Information & Communications Technology Planning & Evaluation).

REFERENCES

[1] L. A. H. Ahmed and Y. A. M. Hamad, “Machine learning techniques for network-based intrusion detection system: A survey paper,” in 2021 National Computing Colleges Conference (NCCC), pp. 1- 7, IEEE, 2021.

[2] S. Gamage and J. Samarabandu, “Deep learning methods in network intrusion detection: A survey and an objective comparison,” Journal of Network and Computer Applications, vol. 169, p. 102767, 2020.

[3] R. Islam, R. U. D. Refat, S. M. Yerram, and H. Malik, “Graph-based intrusion detection system for controller area networks,” IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 3, pp. 1727- 1736, 2020.

[4] C. Geng, S.-j. Huang, and S. Chen, “Recent advances in open set recognition: A survey,” IEEE transactions on pattern analysis and machine intelligence, vol. 43, no. 10, pp. 3614- 3631, 2020.