Research on Graph AI Based Anomaly Detection for Endpoint Detection and Response

Hyun-woo Lee , and Tae-jin Lee*

Hoseo University, Republic of Korea
{yi054543, kinjecs0}@gmail.com

Abstract

Anomaly detection tasks are a crucial security operation in the ever-evolving cyberspace, where numerous threats occur on a daily basis, aimed at safeguarding the assets of individuals and businesses. Most of the existing AI technologies used in security have primarily performed AI tasks based on the value of individual data. However, to carry out an attack, a series of attack processes become evident, and this topology information can be valuable for anomaly detection operations. Therefore, we aim to enhance effective anomaly detection in EDR by utilizing Graph AI that takes into account the topology information of the data. We preprocess the log data collected in EDR and apply an Autoencoder model.

Keyword : GCN, EDR, Anomaly Detection, Neighborhood Aggregation Based preprocessing

1 Introduction

In the cybersecurity environment, it is crucial to detect attacks before they penetrate the system to minimize damage. To achieve this, Endpoint Detection and Response (EDR) has emerged, which analyzes data collected from all endpoints to detect suspicious threats in real-time and automatically respond to them.As mentioned in previous studies[1], in EDR environments, rapid response is crucial, so Anomaly Detection Algorithm must be lightweight. Furthermore, according to Wei[2], it is mentioned that several procedures become evident in order to carry out an attack. Therefore, we propose a lightweight anomaly detection algorithm by incorporating a GCN approach into data preprocessing to reflect the inherent attack topology information in the log data collected from EDR.

2 Proposed Framework and Experiment

2.1 Dataset and Log Featuring Method

We perform anomaly detection based on the log data collected when using HWP, which is commonly used as a malware entry point in South Korea. These logs are generated when a new process is created, and they contain essential information related to the processes in the collected logs. Through the Process Guid information among this data, we were able to

The 7th International Conference on Mobile Internet Security (MobiSec'23), Dec. 19-21, 2023, Okinawa, JAPAN, Article No. P-33

* Corresponding author: Hoseo University, Seoul, 06724, Republic of Korea, Email: yi054543@gmail.com

uncover the contextual information of each process. Therefore, we extracted edges to enable GCN operation using this information, and we conducted NLP processing based on n-gram hashing for the string fields present in the original logs.

2.2 GCN Based Neighborhood Aggregation

In this section, we perform Graph AI Based Anomaly Detection as shown in Figure 1 using the output of the previous step, which is the featured log data and edge information from section 2.1. Since attack data tends to be associated with other attack data, as evidenced by the clustering of feature values before and after aggregation, we can observe that the attack patterns become more pronounced during neighborhood aggregation

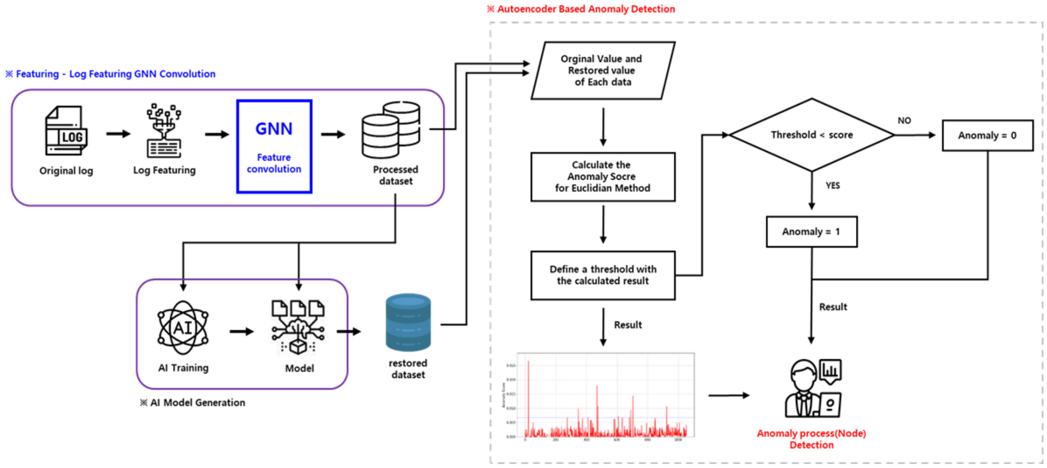
※ Autoencoder Based Anomaly Detection
※ Featuring - Log Featuring GNN Convolution
LOG
Original log
Log Featuring
GNN
Feature convolution
Processed dataset
Orginal Value and Restored value of Each data
Calculate the Anomaly Socre for Euclidian Method
Define a threshold with the calculated result
Threshold < score
NO
Anomaly = 0
YES
Anomaly = 1
Result
Result
AI
AI Training
Model
restored dataset
※ AI Model Generation
Anomaly process(Node) Detection

Figure 1: Proposed Framework : Graph AI Based Anomaly Detection

2.3 Comparison Anomaly Detection Results

Total Data : 2120 Total Attack : 76		Basic	GNN Based Preprocessing	
			Aggregation Mean	Aggregation Sum
Thres Rate	1% (21)	9 (11.8%)	7 (9.2%)	9 (11.8%)
	5% (106)	**18 (23.6%)**	26 (34.2%)	**54 (71%)**
	10% (212)	**26 (34.2%)**	35 (46%)	**66 (85.8%)**

Table 1: Result of Anomaly Detection through each Approach

The anomaly detection performance for each approach is as shown in Table 1. We can identify that the Sum-based Aggregation technique in our proposed GNN-based preprocessing method has improved anomaly detection performance by approximately threefold compared to the previous approach.

3 Conclusion

We conducted anomaly detection by preprocessing data using Graph AI to account for anomalous topologies that are only present in attack data, as the approach based on single data did not capture these. The experimental results confirmed that the approach incorporating topology yielded higher detection rates. In the future, we plan to conduct experiments introducing Graph Aggregation methods with heterogeneous data to perform Anomaly Detection from various perspectives.

Acknowledgement : This work was supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government(MSIT) (No.2022-0-00089, Development of clustering and analysis technology to identify cyber attack groups based on life-cycle)

References

[1] Zhao, Jun, et al. "Cyber threat intelligence modeling based on heterogeneous graph convolutional network." 23rd international symposium on research in attacks, intrusions and defenses (RAID 2020). 2020.

[2] Gao, Yali, et al. "Hincti: A cyber threat intelligence modeling and identification system based on heterogeneous information network." IEEE Transactions on Knowledge and Data Engineering 34.2 (2020): 708-722.