Enhancing GNN-Based Network Intrusion Detection Systems through Memory-Replay Approach

Dinh-Hau Tran[1], Minho Park[2]
[1]Department of Information and Telecommunication Engineering,
Soongsil University, Seoul 06978, Republic of Korea
[2]School of Electronic Engineering, Soongsil University, Seoul 06978, Republic of Korea
hautransns@soongsil.ac.kr, mhp@ssu.ac.kr

***Abstract*—Many recent studies have applied deep learning techniques to network intrusion detection systems (NIDS) to detect increasingly sophisticated cyberattacks. However, several limitations still exist when deploying these models in real-world network environments. Specifically, conventional deep learning models are often trained using a single dataset, typically assuming a static environment with unchanging data distribution. In reality, the data of the network system always change over time. This change causes deployed models to become obsolete, leading to performance degradation. Additionally, retraining these models with new datasets faces the challenge of catastrophic forgetting - the loss of previous knowledge when learning new information. In this study, we introduce a novel framework that leverages the capabilities of the FN-GNN [1] model and memory-replay continual learning techniques to improve the performance and adaptability of NIDS. Continual learning enables the system to continuously learn and adapt to emerging attack scenarios while retaining previously acquired knowledge. Experimental results on benchmark datasets, CIC-IDS2017 and UNSW-NB15, show that the proposed method helps improve continuous learning ability while mitigating the problem of catastrophic forgetting.**

***Index Terms*—Network Intrusion Detection System (NIDS), Graph Neural Networks (GNNs), Continual Learning, Catastrophic Forgetting**

I. Introduction

Recently, deep learning (DL) techniques have been widely applied in various fields such as data processing [2], [3] and cybersecurity [4]. IDS systems are also being integrated with DL to tackle the challenges of detecting unknown and sophisticated attacks. Despite the significant results achieved, these approaches also have limitations. Specifically, most research works focus on the performance of models in a static environment, where the data distribution remains the same. In contrast, the data domain in real-world network systems constantly changes over time due to alterations in network structure, types of services, and the emergence of new traffic patterns. Therefore, the data received by the models in practical systems often have a different distribution from the data on which the models were previously trained. Furthermore, the nature of benign traffic also evolves, leading to the creation of many new variants. All of these distinct properties of real-world environments contribute to a gradual decline in model performance over time. To maintain the performance of IDS, it is essential to retrain existing models with newer data. This research aims to develop a single model capable of learning and maintaining its performance in multiple datasets.

Most previous studies have focused on transfer learning, where models leverage existing knowledge to support the acquisition of new information. However, some research [5] only uses new data for retraining, causing the model's parameters to gradually shift towards fitting the new data. This shift leads to catastrophic forgetting [6], where the model severely loses previously learned information when learning new data. Meanwhile, other studies employ a method to retrain the model directly with data that includes both new datasets and all old data. However, as unseen data continuously appear in the system over time, frequent model retraining is required. Consequently, the size of the dataset continuously increases and becomes enormous over multiple updates. This significantly increases both computational complexity and data storage costs. Furthermore, using all old data for retraining can easily lead to overfitting, as the model learns repeatedly on the same old datasets. Therefore, this method is not particularly feasible, or at least not ideal in practice.

The problem of catastrophic forgetting remains a major challenge in the process of retraining models. To address this, the continual learning (CL) method is considered a promising technique and has received significant attention in recent studies. The two main objectives of CL are to help models improve their adaptability to new data while retaining the knowledge learned from the past data. CL techniques are divided into three categories: regularization-based, memory-based, and architecture-based techniques. CL is widely applied in the field of computer vision and has shown considerable effectiveness, as demonstrated in [7]–[9]. However, this technique has not been thoroughly explored in network security. Especially, to the best of our knowledge, there has been no specific research applying CL to GNN model-based NIDS. GNNs are considered the most suitable and effective for the data that IDS monitors. The ability of GNNs to learn from neighboring nodes allows them to exploit complex relationships between flows generated within the system effectively. In this study, we propose a novel framework that applies memory-replay techniques to the FN-GNN model, aiming to improve the performance and adaptability of NIDS. Among the various CL techniques, the memory-replay method is the most suitable for application in GNN-based detection systems. Memory-replay continual learning methods involve saving part of the input samples in a memory buffer during training. The idea is

979-8-3315-0694-0/25/$31.00 ©2025 IEEE

to use these memory samples for model training along with new data to prevent catastrophic forgetting. The integration of this technique into the FN-GNN model, which we previously proposed and achieved impressive performance in classifying malicious network data, significantly enhances the system's capability for continuous learning in real-world environments.

We summarize our contributions as follows.

- We proposed a novel framework that integrates memory-replay techniques into the FN-GNN model to address the catastrophic forgetting problem.
- The proposed method was applied to two benchmark datasets, and we provided simulation results demonstrating its effectiveness.

The remainder of this paper is organized as follows. Section II presents our proposed method. The experimental results are discussed in Section III. Finally, we provide our conclusions in Section IV.

II. Proposed method

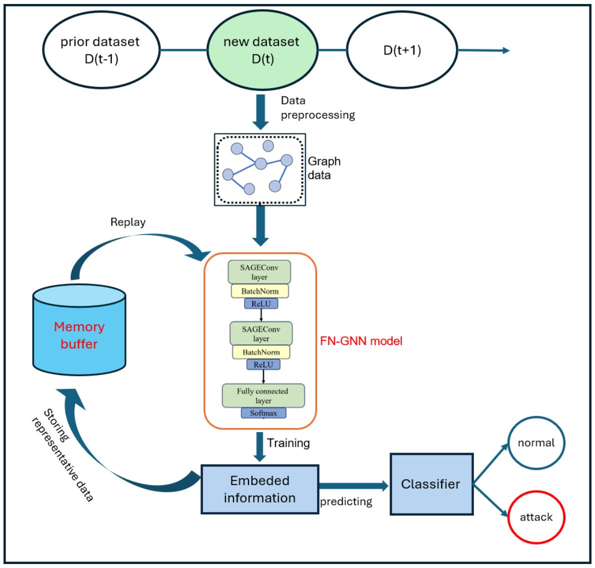

Fig. 1: Diagram of the proposed model.

Our proposed model is illustrated in Fig. 1. The framework combines the FN-GNN model with a memory-replay technique for network flow classification. The FN-GNN framework converts network flow data into a graph-structured model, using a modified graph convolutional network (GCN) to classify nodes as normal or attack flows. After training, the representative key nodes are preserved in a memory buffer for future analysis. When the model is retrained with new data, these stored nodes are used for the new task. During retraining, the data from the memory buffer are replayed and used concurrently with the new graph data to learn the graph representation for classification. At the end of the retraining process, representative nodes of this graph are selected again and sent to the memory buffer, as in the previous task. This process is repeated for each task that the model performs.

The effectiveness of the memory-replay method is governed by the approach used to store sample nodes in the memory buffer and replay them during the retraining phase. The selected samples must be representative of the old dataset while not creating a burden on storage costs. Similarly, the retraining method must be chosen appropriately to ensure that the model's parameters are not updated in a way that overly favors the new data. This is crucial for the model to achieve balanced performance on both datasets, minimizing catastrophic forgetting of previous data. In this study, we employ the mean of feature method to sample data after training. Consequently, an average feature vector is calculated for each label based on the embedded information. Subsequently, nodes are selected such that these nodes have embedded feature vectors closest to the calculated average feature vector. An equal number of nodes for label 0 and label 1 are chosen to avoid class imbalance in the buffer. To update the model parameters during retraining, we utilize the method proposed in the ER-GNN model [10]. The FN-GNN model uses the Adam optimizer and Cross-entropy loss function for training. Loss values are calculated separately for the new dataset and the data from the buffer. Then, a final loss value is synthesized from these component loss values according to the formula presented in formula (1) below.

$$L_{final}(f_\theta, D_T, D_M) = \alpha L(f_\theta, D_T) + (1-\alpha)L(f_\theta, D_M) \quad (1)$$

Where L_{final} is the final loss value calculated from the training set D_T of new data and the replayed data D_M from the buffer using our classifier f_θ. α is a balancing weight that adjusts the proportion of nodes from D_T and D_M. This weight is dynamically updated according to the following formula:

$$\alpha = D_M/(D_M + D_T) \quad (2)$$

The number of nodes in new data is typically much larger than in buffer data, so the α coefficient helps prevent catastrophic forgetting of old data during retraining. Using stored data from previous tasks effectively in the retraining process, the proposed method allows the FN-GNN model to improve its adaptability to new data while preserving existing knowledge.

III. Results and Discussion

We conducted experiments on two benchmark datasets: CIC-IDS2017 and UNSW-NB15. The number of data points used for these datasets was 26,554 and 56,000 flows, respectively. The experimental data was split into 80% for training and 20% for testing.

To clarify the phenomenon of catastrophic forgetting when a model learns continuously on two datasets, we trained the model sequentially on the two datasets in the order: CIC-IDS2017 followed by UNSW-NB15. After each training phase, the evaluation results are presented in Figs. 2 and 3. Fig. 2 shows that the model trained on the CIC-IDS2017 dataset achieved high performance (approximately 99%) on this data but performed poorly on the untrained UNSW-NB15 data. After continuing to train the model on the UNSW-NB15 dataset, its prediction results improved significantly (approximately 94% in Fig. 3). However, the model obtained after the second training phase suffered a severe performance drop on the test set of the first dataset. The prediction results on the CIC-IDS2017 dataset at this point were only 44%. Thus, the catastrophic forgetting phenomenon occurred when the model learned with new data, with a forgetting rate of about 45%.

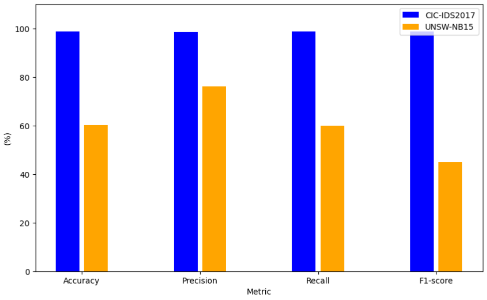

Fig. 2: Model trained on CIC-IDS2017 dataset perform poorly on UNSW-NB15 dataset.

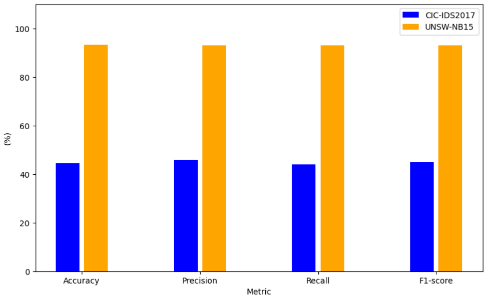

Fig. 3: Model first trained on CIC-IDS2017 dataset and then on UNSW-NB15 only perform well on the later dataset.

The performance of the proposed model is shown in Fig. 4. After retraining with the UNSW-NB15 dataset and buffer data, the prediction results on the test set of the CIC-IDS2017 were maintained at 88%. Consequently, the forgetting rate of previous data decreased from 45% to 11%, representing a 34% improvement. Moreover, performance in new data also improved significantly compared to the pre-retraining performance shown in Fig. 2. These results demonstrate that the proposed framework enhances the ability of the FN-GNN model to learn new data while substantially mitigating the catastrophic forgetting phenomenon.

IV. Conclusion

In this study, we proposed a novel framework that applies memory-replay techniques to enhance the continuous learning capability of the FN-GNN model while significantly mitigating the problem of catastrophic forgetting. Experimental results on benchmark datasets demonstrate the effectiveness of the proposed method. In future work, we plan to optimize the model's performance with improved sample selection methods and conduct more comprehensive experiments on various datasets.

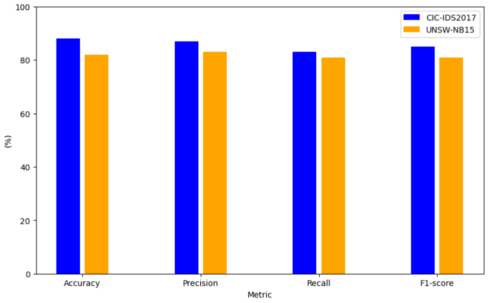

Fig. 4: Performance of model first trained on CIC-IDS2017 dataset and then on UNSW-NB15 dataset.

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] D.H. Tran, and M. Park, “FN-GNN: A Novel Graph Embedding Approach for Enhancing Graph Neural Networks in Network Intrusion Detection Systems,” Applied Sciences, vol. 14, no. 16, pp. 6932, 2024.

[2] X.T. Dang, H.V. Nguyen, and O.S. Shin, “Optimization of IRS-NOMA-assisted cell-free massive MIMO systems using deep reinforcement learning,” IEEE Access, vol. 11, pp. 94402-94414, 2023.

[3] T. A. Nguyen, and J. Lee, “Using Long Short-Term Memory to Estimate the 2-D Interference of Bit-Patterned Media Recording Systems,” in IEEE Transactions on Magnetics, vol. 60, no. 9, pp. 1-5, Sept. 2024.

[4] D. -H. Tran, and M. Park, “Graph Embedding for Graph Neural Network in Intrusion Detection System,” In 2024 International Conference on Information Networking (ICOIN), pp. 395-397, Jan. 2024.

[5] K. Sethi, E. Sai Rupesh, R. Kumar, P. Bera, and Y. Venu Madhav, “A context-aware robust intrusion detection system: a reinforcement learning-based approach,” International Journal of Information Security, vol. 19, pp. 657-678, 2020.

[6] R. Kemker, M. McClure, A. Abitino, T. Hayes, and C. Kanan, “Measuring Catastrophic Forgetting in Neural Networks,” AAAI, vol. 32, no. 1, Apr. 2018.

[7] K. Doshi and Y. Yilmaz, “Continual learning for anomaly detection in surveillance videos,” In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops, pp. 254-255, 2020.

[8] S. Wang, Z. Laskar, , I. Melekhov, X. Li, and J. Kannala, “Continual learning for image-based camera localization,” In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3252-3262, 2021.

[9] W. Zhang, D. Li, C. Ma, G. Zhai, X. Yang, and K. Ma, “Continual learning for blind image quality assessment,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 3, pp.2864-2878, 2022.

[10] Zhou, Fan, and Cao. Chengtai, “Overcoming Catastrophic Forgetting in Graph Neural Networks with Experience Replay,” Proceedings of the AAAI Conference on Artificial Intelligence, vol. 35, no.5, pp. 4714-4722, 2021.