Deep Graph Learning for DDoS Detection and Multi-Class Classification IDS

Braden J. Saunders[1], Robson E. De Grande[1], Glaucio H.S. Carvalho[2], Isaac Woungang[3],
[1]Department of Computer Science, Brock University, St. Catharines, ON, Canada.
[2]Department of Computer Science and Engineering, Brock University, St. Catharines, ON, Canada.
[3]Department of Computer Science, Toronto Metropolitan University, Toronto, ON, Canada.
Email: bs17jq@brocku.ca, rdegrande@brocku.ca, gdecarvalho@brocku.ca
iwoungan@torontomu.ca

***Abstract*—Critical infrastructure systems have been preyed on by cyber criminals that target to disrupt their operations and national security. Among the most nefarious attacks, the Distributed Denial of Service (DDoS) attack is wreaking havoc on the Telecommunications sector. This paper invests in the vision that Artificial Intelligence (AI) plays an important role in shoring up the cybersecurity of critical infrastructure providers by detecting and classifying malicious engagements. In this respect, we propose an efficient and dependable DDoS specialized intrusion section system (IDS). The proposed system is empowered by Graph Convolutional Networks (GCN), a deep learning technique, which is capable of capturing the topological and statistical information between the attack network and the victim network. The results show that the proposed GCN IDS can detect and classify multiple variations of DoS with a high confidence level.**

***Index Terms*—Distributed Denial of Service, Deep Learning, Graph Neural Networks, Intrusion Detection Systems**

I. Introduction

Critical infrastructure providers across different sectors such as communications, healthcare and public health sector, financial, and energy, to name a few, have been victimized by cybercriminal organizations systematically. Among the most nefarious cyber attacks, is the Distributed Denial of Service (DDoS) that wreaks havoc on the availability of the targeted systems and networks precluding legitimate access from taking place. Recent statistics on DDoS attacks have highlighted a massive increase of 387% in the number of attacks when contrasting the second quarter (Q2) and the first one (Q1) of 2023 where the telecommunication sector alone accounts for approximately 50% of these malicious campaigns [1].

Such statistics of DDoS attacks are alarming and demand innovative defense approaches to defeat the threat actors. In this respect, the use of Artificial intelligence (AI) and, more specifically, Machine Learning (ML) as tools to augment detection capabilities is among the most promising approaches. Remarkably, the ability to identify and learn from complex hidden partners with static and spatial-temporal data and generalize, i.e., to adapt to unseen data such as zero-day attacks, has attracted a great deal of attention within the Intrusion Detection System (IDS) community. To a large extent, Deep Learning (DL) methods such as Deep Neural Network (DNN), Convolutional Neural Network (CNN), Long Short-Term Memory Network (LSTM) to name a few are behind several recent breakthroughs [2].

Undoubtedly, a DDoS attack campaign recruits a massive amount of end-to-end connections (E2E) between a diverse set of compromised Internet of Things (IoT) devices and the target infrastructure which are featured by rich structural relationships. This graph-structured big data provides valuable information that can be successfully leveraged by a graph-based IDS to identify anomalous nodes (i.e., malicious IoT devices), anomalous edge (i.e., abnormal/attack relations), and anomalous sub-graph-level (i.e., malicious IoT groups). When combined with DL techniques, graph anomaly detection with deep learning (GADL) technology has the potential to become a front-runner in the development of advanced cyber attack detection tools [3]. This is the case of Graph Neural Networks (GNN) which is a class of DL models designed to perform inference on graph-structured data and Graph Convolutional Networks (GCN) which is a class of GNN that learns a graph representation and aggregate or update node information from the neighborhoods in a convolutional manner [4]. When applied in the context of DoS attack detection,

978-1-7281-7315-3/20/$31.00 © 2020 IEEE

they have demonstrated excellent results [5]–[8].

In this paper, we abstract the problem of DDoS detection and multi-class classification as an edge detection problem and leverage GADL technology for graph mining in the DDoS dataset. Particularly, the proposed GCN-empowered IDS is specialized in DDoS detection and multi-class classification, which is crucial for the following reasons:

1) DDoS is a destructive attack with devastating consequences to critical infrastructure providers such as telecom operators. Interruptions to telecommunication systems would produce a cascade effect that impacts the targeted operator and other critical infrastructures that depend upon it such as the ones in the healthcare and public health sector, transportation sector, financial sector, and energy sector, to name a few. As a result, a successful DDoS attack might collapse the entire nation;
2) The high level of sophistication of threat actors has resulted in a rich set of DDoS attack variations that might call for AI solutions that are tailored to these attacks. For instance, DDoS can be classified as volumetric DDoS attacks that quickly and abruptly exhaust the targeted capacity whereas slow DDoS is a stealthy attack that sends an ordinary number of service requests to hijack the server resources for extended periods making them unavailable to legitimate service requests. Moreover, DDoS campaigns might leverage different protocols HTTP, UDP, TCP, ICMP, DNS to flood the network bandwidth or compromise server resources such as sockets, CPU, and memory.

These aspects exacerbate the criticality of DDoS detection and the challenges of designing, training, updating, and operating a single DL IDS that is capable of generalizing well and effectively identifying different types of DDoS attacks in conjunction with other sophisticated attacks (well-known attacks and zero-day attacks) due to challenges such as the availability of representative and current datasets in addition to model overfitting caused by class imbalance, hyperparameter tuning, noisy datasets, to name a few.

The current paper addresses the problem of detection and multi-class classification considering a diverse set of DDoS attacks. In this respect, the contributions of this paper are directly related to the previous GNN-empowered IDS papers such as [5]–[8]. Compared to them, our proposal is equally innovative but carries the benefit of being simple and similarly accurate. Considering the widely adopted IDS dataset from the Canadian Institute of Cyber Security known as "CIC-IDS2017" [11], we found that the model achieved an overall $Accuracy$ of over 99% and $Precision$, $Recall$, and $F1{-}Score$ values ranging from 95% for the Slowhttptest attack with few attack samples to more than 99% to well-represented attacks such as GoldenEye. These results demonstrate the effectiveness in terms of high detection and multi-class classification capabilities of the proposed model.

The remainder of this paper is organized as follows. Next Section presented an overview of the literature. Section III deals with the threat modeling including the system under analysis, the adversary capabilities, and the defense system while Section IV describes the proposed GCN-empowered IDS model for DDoS detection and multi-class classification. Section V describes the performance metrics used to evaluate the model. Next Section introduces the dataset and the implementation aspects of the model. Results are illustrated in Section VII. Finally, Section VIII concludes the paper and sheds light on future research avenues.

II. Related Work

AI and, more specifically, Deep Learning (DL) have been important players within the IDS research community. Thanks to its capacity to uproot inherent and meaningful relationships from complex data, DL models have significantly contributed to advancing the state-of-the-art of IDS. One of the most prominent DL detection systems for DDoS is LUCID which is based on Convolutional Neural Networks (CNN). This IDS matches the state-of-the-art of detection systems in terms of classification of benign and malicious activities while presenting a 40x reduction in processing time [12]. DeepSecure, a Long-Short Term Memory (LSTM) IDS, is designed to protect the 5G system which classifies the user traffic as DDoS attack or normal traffic and assigns an appropriate slice to a legitimate user request [13]. Activity Event Network (AEN) and GNN are leveraged by [14] to detect DDoS attacks. Notably, AEN which boosts detection of volumetric and stealthy attacks, builds a graph model that is used by the GNN. Still in the GNN IDS landscape, [5] considers in-flow and between-flow information to improve detection while [6] combines traffic-level and the flow-level to boost a hierarchical graph representation and a GNN model that will be used for DDoS detection. The DDoS multi-class classification problem is addressed by [7] where a graph of graph methodology that uses GNN is employed. The work in [8] leverages a GCN

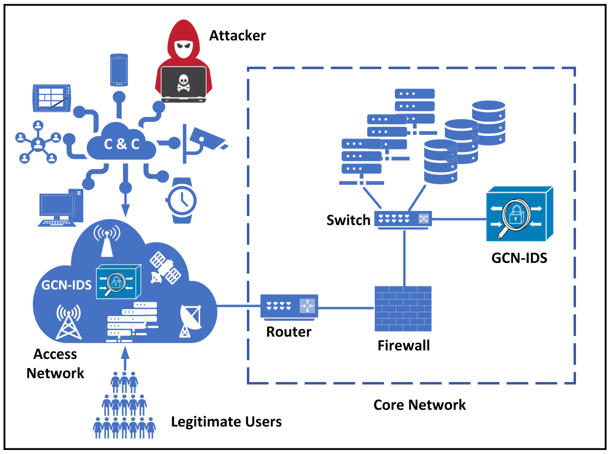

Fig. 1. System under Analysis

to deal with the binary classification and the multi-class classification and achieves excellent results. However, the multi-class classification differentiates DDoS and DoS from other attacks without examining the intrinsic characteristics of the attack engagement such as Hulk, GoldenEye, Slowloris, and SlowHTTPTest. Using an innovative approach, the work in [9] leverages GNN to enforce source rate limits and so mitigate the risk of DDoS in Software-Defined Networking (SDN) settings. Still in the SDN domain, the work in [10] puts forward a GCN DDoS detection system model that focuses on spatial-temporal data characteristics to identify and contain DDoS flows. In this paper, we abstract the DDoS detection problem as a GCN edge classification problem that combines topological and statistical information to improve DDoS detection and multi-class classifications. Despite its simplicity, the proposed model matches the state-of-the-art of advanced DDoS IDS systems considering Hulk, GoldenEye, Slowloris, and SlowHTTPTest attacks.

III. Threat Model

Fig. 1 depicts the threat model. As shown, the attacker can orchestrate the DDoS attack by a command and control (C&C) infrastructure where a massive number of IoT devices is recruited to send requests to the target critical infrastructure. Fig. 1 assumes a telecommunications service provider and the focus of the malicious campaign can be the access network, the core network, or the entire system. Regardless of the focus, the attacker aims to cause a loss of availability, which will lead to financial loss and a reputational loss that might be exacerbated by indirect consequences such as casualties and fatalities that the system unavailability might result.

To shore up the infrastructure cybersecurity, the proposed GCN-empowered IDS can be deployed in the access network and core network to inspect, store, manipulate, and block communications. In parsing the networking flows, the IDS system has full visibility of the attributes of the E2E connection, such as source IP address, source port, destination IP address, destination port and transport protocol, etc. These attributes are used to train the proposed GCN-IDS model which is put online to defend the infrastructure by identifying DDoS flows and classifying them when the proposed GCN-IDS model generalizes well. Notably, similar assumptions are made by [8].

IV. Proposed GCN-Empowered IDS Model

GCNs are variations of deep neural networks that exploit the non-Euclidean characteristic of graphs such as their irregular structures along with aggregate node information from neighborhoods convolutionally. In doing so, they have achieved notable performance across multiple and distinguishable areas such as computer vision, natural language processing, and science. As follows, we present our GCN methodology for IDS.

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote a graph where $\mathcal{V} = \{v_1, v_2, \cdots, v_{|\mathcal{V}-1|}, v_{|\mathcal{V}|}\}$ is the set of nodes and $\mathcal{E}$ is the set of edges. Considering Fig. 1, the nodes $v_i \in \mathcal{V}, \forall i \leq |\mathcal{V}|$ represent the IP addresses involved in an E2E connection between a legitimate or malicious device in the C&C infrastructure and a device in the critical infrastructure. This topological information, i.e., the graph connectivity is captured by the adjacency matrix $\mathcal{A} \in \mathbb{R}^{|\mathcal{V}| \times |\mathcal{V}|}$ in such a way that if the devices v_i and v_j are communicating, i.e., $e_{ij} = (v_i, v_j) \in \mathcal{E}$, then $\mathcal{A}_{ij} = 1$. Otherwise, $\mathcal{A}_{ij} = 0$. Finally, let $\mathcal{X} \in \mathbb{R}^{|\mathcal{V}| \times d}$ denote the node attribute matrix and d the signals of the graph.

The statistical characterization of the flow in the edge $e_{ij} = (v_i, v_j) \in \mathcal{E}$ defines whether the communication is benign or malicious for the binary classification and the type of DoS attack for multi-class classification. For this reason, we abstract the solution as an edge classification problem. Therefore, the aim of leveraging the graph convolution is to uproot the edge embeddings that represent the communication flow.

It is important to mention that the node classification approach, which detects and blocks malicious nodes, would render an inadequate solution since IP address blocking and blacklisting only work temporarily and will not be able to stop advanced persistent threats (APT).

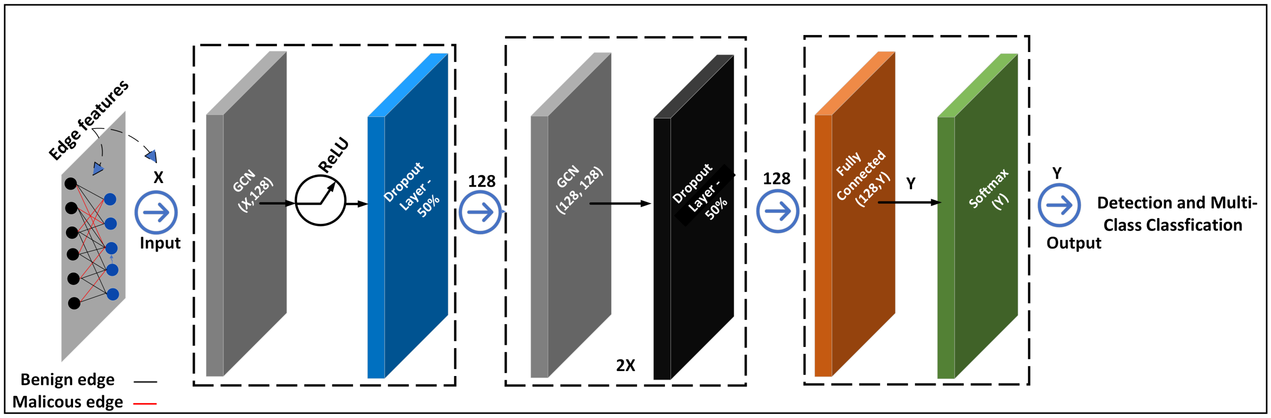

Fig. 2. Proposed GCN-Empowered DDoS Detection System Architecture

For this reason, they are ineffective practices in modern cybersecurity.

In this paper, we adopt the solution proposed by Kipf and Welling for the GCN layer, which is specified as follows [15]:

$$\mathcal{H} = \tilde{\mathcal{D}}^{-\frac{1}{2}} \tilde{\mathcal{A}} \tilde{\mathcal{D}}^{-\frac{1}{2}} \mathcal{X} \mathcal{W} \tag{1}$$

where $\mathcal{H}$ is the GCN layer, $\tilde{\mathcal{A}} = \mathcal{A} + \mathbf{I}$ is the adjacency matrix of the undirected graph $\mathcal{G}$ with added self-connections, $\mathbf{I}$ is the identity matrix, $\tilde{\mathcal{D}}$ is the diagonal degree matrix of $\tilde{\mathcal{A}}$, and $\mathcal{W}$ is the matrix of trainable data of the layer.

Based on Eq. (1), Fig. 2 depicts the proposed GCN IDS architecture for detection and multi-class DDoS attack classification. The rectified linear unit $\mathrm{ReLU}(\cdot) = \max(0, \cdot)$ activation function and dropout layers are used to add non-linearity and prevent overfitting, respectively. In addition to the GCN layer, the model leverages a fully connected layer with a log softmax for multi-class classification.

V. Performance Metrics

The performance of the proposed GCN-empowered IDS is quantified using the precision, recall, and F-1 score metrics for each class, along with the overall accuracy of the model. These metrics are derived based on the following quantities:

- True positive (TP): it reports the observations that are correctly classified.
- True negative (TN): it reports the observations that are correctly classified as a negative instance.
- False positive (FP): it reports the observations that are incorrectly classified.
- False negative (FN): it reports the observations that are incorrectly classified as negative instances.

1) Precision: It informs the quality of the positive predictions made by the IDS. Precision is given by

$$Precision = \frac{TP}{TP + FP} \tag{2}$$

2) Recall: It quantifies how often the IDS correctly identifies the TPs from all the actual positive samples in the dataset. Recall is given by

$$Recall = \frac{TP}{TP + FN} \tag{3}$$

3) F-1 Score: It is the harmonic mean between the precision and recall metrics. The F-1 score informs about the IDS reliability since it encapsulates precision and recall. The F-1 $Score$ score is given by

$$F\text{-}1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{4}$$

4) Accuracy: Accuracy reports how well the IDS model can predict the outcome. Accuracy is expressed by

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5}$$

VI. Data PreProcessing

A. Dataset

In this paper, we evaluated the proposed IDS using the dataset known as “CIC-IDS2017” [16]. This dataset was generated considering an attacker network with two clients running Windows and Kali Linux as well as a victim network with twelve clients operating based on Windows, Linux, and MacOS. Entries in the dataset are the packets transmitted between these networks that were parsed using the CICFlowMeter application, which, in turn, extracted over 80 network flow features for each packet. Although multiple attacks are reported in the dataset, our focus was on the subset reporting multiple types of DoS attacks. More specifically, we considered the *DoS/DDoS* subset that contains 792,703 packets and categorizes four DoS attacks, namely, Slowloris, Slowhttptest, Hulk, and GoldenEye.

B. Implementation

From the CSV file, the data is transformed into tensors. Recall that we are considering an edge classification problem and, therefore, the topological information is represented by the communicating nodes with their corresponding IP addresses. Moreover, the network flow (statistical information) is represented as an edge. In practical terms, it means that each source and destination IP address in the CSV file becomes a node while each row in the CSV file becomes an edge.

Pandas - the Python data science library - is leveraged for data manipulation. In terms of pre-processing, we dropped the columns with *Not a Number (NaN)* or *infinite* - *"Flow Bytes/s"* and *"Flow Packets/s"* columns. The *LabelEncoder* class from the *sklearn* library is used to turn source and destination IP addresses into numerical labels. On the other hand, *PyTorch* converts the encoded source and destination IP labels to tensors. To deflate the overfitting risk, attributes like *Flow ID, Source IP, Source Port, Destination IP, Protocol, Timestamp* are discarded. The remaining attributes of the network flow are taken as input. In terms of training and testing, we consider 80% for training and 20% for testing. The final data is stored in a *PyTorch Geometric* Data object before being passed to the GCN model. Finally, we use the Adam algorithm with a learning rate of 0.001, 1000 epochs, and dropout rate of 0.5.

VII. Results

TABLE I
MULTI-CLASS CLASSIFICATION REPORT

Traffic	$Precision$	$Recall$	$F1 - Score$	**Support**
Benign	0.9992	0.9888	0.9940	87916
GoldenEye	0.9947	0.9806	0.9876	2113
Hulk	0.9795	0.9992	0.9893	46118
Slowhttptest	0.9552	0.9928	0.9737	1117
slowloris	0.9909	0.9671	0.9789	1126
Accuracy			0.9920	138390
Macro avg	0.9839	0.9857	0.9847	138390
Weighted avg	0.9922	0.9920	0.9920	138390

Table I depicts the results for the proposed GCN IDS. As shown, the metric values concentrate in the range between 95% and 99+%, which underscores the high performance and dependability of the model in terms of detection and multi-class classification.

Independently of the type of traffic, $Precision$ and $Recall$ are consistent over 95% leading to a robust F-1 $Score$ - the most important indicator due to the class imbalance illustrated in the **Support** column. Undeniably, one of the most important features of an IDS system is to detect benign traffic properly. In this respect, the proposed one achieves more than 99% - F-1 $Score$. This result is corroborated by the confusion matrix in Fig. 3 which indicates that 86932 packets out of 87916 are correctly identified as benign packets.

Furthermore, the proposed GCN IDS unequivocally distinguishes among the attack types. The worst-case scenario corresponds to the Slowhttptest $Precision$ metric which marks approximately 95.5%. Fig. 3 shows that only 8 packets predicted to be Slowhttptest were misclassified out of 1117. Among them, 5 packets are benign and 3 are Slowloris. This result is likely because of the small number of samples of this attack type - 1117 out of 138390 entries in the dataset. Not surprisingly, Slowloris comes in second with the worst metric being approximately 97.8%. Future research should be conducted to synthetically augment the number of samples for these attacks using Generative AI. Finally, Table II compares the proposed model against two recent breakthroughs that also leverage GNN and use the same dataset. As seen, the proposed model stands at the same level of $Accuracy$, i.e., above 99%, with the benefit of being simple.

TABLE II
PERFORMANCE COMPARISON

Model	**Accuracy**
GoGDDoS [7]	0.9933
FTG-NET [6]	0.9914
Our Proposal	**0.9920**

Remarks: As shown, our model presents excellent results despite its simplicity which is relevant because a dependable IDS is vital for organizations to trigger the appropriate cybersecurity strategy. For instance, when operating under a defensive approach, the security team might immediately drop the connection while an offensive counterintelligence strategy might allow the continuity of the attack to collect intelligence about its modus operandi. In this respect, an inspection of the diagonal

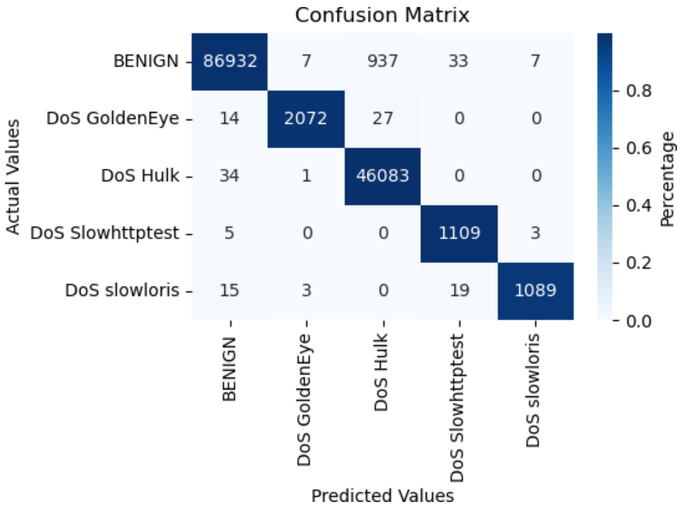

Fig. 3. Confusion Matrix.

values of the confusion matrix in Fig. 3 along with an analysis of Table I confirm the high performance and dependability of the proposed model which qualifies it as a potential candidate for being adopted by organizations.

VIII. Conclusion

This paper presents a simple and highly efficient GCN IDS for DDoS attacks. Using the well-known “CIC-IDS2017”, the proposed model achieved high performance and dependability considering the discussed metrics and previous works. For future works, we are considering

- To augment the underrepresented attack types in such a way that we end up with a balanced dataset and, consequently, a better training/testing setup for the model.
- To add attention mechanisms to empower the detection capabilities of the model;

References

[1] “Protecting Your Business From Cyber Attacks: The State of DDoS Attacks Q1 & Q2, 2023”, Zayo Group, https://go.zayo.com/zayo-ddos-protection-ebook/ (accessed Feb. 20, 2024).

[2] H. Liao et al., “A Survey of Deep Learning Technologies for Intrusion Detection in Internet of Things,” *IEEE Access*, vol. 12, pp. 4745-4761, 2024, doi: 10.1109/ACCESS.2023.3349287.

[3] X. Ma et al., “A Comprehensive Survey on Graph Anomaly Detection With Deep Learning,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12012-12038, 1 Dec. 2023, doi: 10.1109/TKDE.2021.3118815.

[4] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, “Graph Convolutional Networks: A Comprehensive Review,” Computational Social Networks, vol. 6, no. 1, pp. 1-23, 2019, https://doi.org/10.1186/s40649-019-0069-y.

[5] Y. Li et al., “GraphDDoS: Effective DDoS Attack Detection Using Graph Neural Networks,” *in Proc. of IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, 2022, pp. 1275-1280, doi: 10.1109/CSCWD54268.2022.9776097.

[6] L. Barsellotti, L. De Marinis, F. Cugini and F. Paolucci, “FTG-Net: Hierarchical Flow-to-Traffic Graph Neural Network for DDoS Attack Detection,” *in Proc. of 2023 IEEE 24th International Conference on High Performance Switching and Routing (HPSR)*, 2023, pp. 173-178, doi: 10.1109/HPSR57248.2023.10147929.

[7] Y. Li, Z. Zhou, R. Li, F. Shi, J. Guo and Q. Liu, “GoGDDoS: A Multi-Classifier for DDoS Attacks Using Graph Neural Networks,” *in Proc. of 2023 IEEE Symposium on Computers and Communications (ISCC)*, 2023, pp. 1462-1467, doi: 10.1109/ISCC58397.2023.10218316.

[8] G. Duan, H. Lv, H. Wang and G. Feng, “Application of a Dynamic Line Graph Neural Network for Intrusion Detection With Semisupervised Learning,” *IEEE Transactions on Information Forensics and Security*, vol. 18, pp. 699-714, 2023, doi: 10.1109/TIFS.2022.3228493.

[9] A. El Kamel, ”A GNN-Based Rate Limiting Framework for DDoS Attack Mitigation in Multi-Controller SDN,” *IEEE Symposium on Computers and Communications (ISCC)*, 2023, pp. 893-896, doi: 10.1109/ISCC58397.2023.10218204.

[10] Y. Cao, H. Jiang, Y. Deng, J. Wu, P. Zhou and W. Luo, ”Detecting and Mitigating DDoS Attacks in SDN Using Spatial-Temporal Graph Convolutional Network,” *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 6, pp. 3855-3872, 1 Nov.-Dec. 2022, doi: 10.1109/TDSC.2021.3108782.

[11] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), 2018.

[12] R. Doriguzzi-Corin, S. Millar, S. Scott-Hayward, J. Martínez-del-Rincón and D. Siracusa, “Lucid: A Practical, Lightweight Deep Learning Solution for DDoS Attack Detection,” *IEEE Transactions on Network and Service Management*, vol. 17, no. 2, pp. 876-889, June 2020, doi: 10.1109/TNSM.2020.2971776

[13] N. A. E. Kuadey, G. T. Maale, T. Kwantwi, G. Sun and G. Liu, “DeepSecure: Detection of Distributed Denial of Service Attacks on 5G Network Slicing—Deep Learning Approach,” *IEEE Wireless Communications Letters*, vol. 11, no. 3, pp. 488-492, March 2022, doi: 10.1109/LWC.2021.3133479

[14] P. Kisanga, I. Woungang, I. Traore and G. H. S. Carvalho, “Network Anomaly Detection Using a Graph Neural Network,” *International Conference on Computing, Networking and Communications (ICNC)*, USA, 2023, pp. 61-65, doi: 10.1109/ICNC57223.2023.10074111.

[15] T. N. Kipf and M. Welling, “Semi-supervised classification with graph convolutional networks”, CoRR, abs/1609.02907, 2016, http://arxiv.org/abs/1609.02907

[16] I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, “Toward Generating a New Intrusion Detection Dataset and Intrusion Traffic Characterization”, 4th International Conference on Information Systems Security and Privacy (ICISSP), Portugal, January 2018