International Journal of Computer Sciences and Engineering
Vol.**12**, Issue.**4**, pp.**12-16**, April **2024**
ISSN: 2347-2693 (Online)
Available online at: www.ijcseonline.org

ISSN: 2347-2693 (E)

Research Article

An Enhanced Intrusion Detection System Using Edge Centric Approach

Bhavya Lahari Vaddempudi[1*], Amrutha Tulabandu[2], S.N.B. Tanuja Reddy[3], Deepika Leela Pudi[4], Venkata Narayana Yerininti[5]

[1,2,3,4,5]Dept. of Information Technology, Vasireddy Venkatadri Institute of Technology, Guntur, India

**Corresponding Author: vaddempudilahari@gmail.com,*

Received: 28/Feb/2024; **Accepted**: 03/Apr/2024; **Published**: 30/Apr/2024. **DOI**: https://doi.org/10.26438/ijcse/v12i4.1216

Abstract: In the ever-evolving landscape of Cybersecurity, the detection and mitigation of network intrusions and anomalous activities remain formidable challenges. Conventional methods for identifying threats often encounter difficulties in scaling up and adapting swiftly, as they heavily rely on labeled network data. Furthermore, a narrow focus on individual data points may inadvertently overlook critical details at the packet level, thus exposing vulnerabilities that malicious actors can exploit. To confront these ongoing challenges head-on, Graph Neural Networks (GNNs) emerge as a promising solution. Their innate ability to comprehend complex network structures equips them with the capability to provide deeper insights into the dynamics of network traffic. By harnessing the power of GNN, it autonomously detects and comprehends intrusions and anomalies, surpassing the limitations of conventional techniques. Through experimentation and evaluation on real-world datasets, the proposed system demonstrates promising results in accurately identifying and classifying network intrusions.

Keywords: Cyber Security, Network intrusions, Graph Neural Networks (GNN), Packet level analysis, Performance metrics, Effectiveness Evaluation, Bot-Iot.

1. Introduction

Over the past 20 years, information technology has advanced dramatically and taken the place of other knowledge sources.The Internet is the most beneficial information source for nearly everyone, from the novice to the expert, since it provides access to the most recent information and cutting-edge technology. Yet with the usefulness and importance of the Internet, the misuse and the cybercrime become more evident [1]. Any illegal behavior involving computers or networks is referred to as "cybercrime". Moreover, Cybercrime includes criminal acts committed online [2].

There are various kinds of cyber-crimes which are happening in day-to-day life. But the people are not aware of all such types. The majority of people just have knowledge of viruses and worms and hacking. They are not aware of phishing, defamation, identity theft, cyber stalking etc [3]. Cybercrime presents challenges such as data breaches, identity theft, ransomware attacks, and phishing scams, leading to financial loss, reputational damage, and privacy violations for individuals and organizations. These crimes exploit vulnerabilities in technology and human behavior, requiring robust cybersecurity measures, law enforcement efforts, and international cooperation to combat effectively. Additionally, emerging threats like IoT attacks and supply chain vulnerabilities continue to pose new challenges for cybersecurity professionals worldwide.

Figure 1: Types of cybercrime [4].

Cybercrime presents challenges such as data breaches, identity theft, ransomware attacks, and phishing scams, leading to financial loss, reputational damage, and privacy violations for individuals and organizations. These crimes exploit vulnerabilities in technology and human behavior, requiring robust cybersecurity measures, law enforcement efforts, and international cooperation to combat effectively. Additionally, emerging threats like IoT attacks and supply chain vulnerabilities continue to pose new challenges for cybersecurity professionals worldwide.

Machine learning is an integral part of artificial intelligence, which is used to design algorithms based on the data trends and historical relationships between data. Machine learning is used in various fields such as bioinformatics, intrusion

detection, Information retrieval, game playing, marketing, malware detection, image deconvolution and so on [5]. ML algorithms enable applications to draw conclusions and make predictions based on existing data without human supervision, leading to quick, near-optimal solutions even in problems with high dimensionality [6].

Traditionally, known threats have been identified using Network Intrusion Detection Systems (NIDS); but, as assaults get more sophisticated, machine learning models like Graph Neural Networks (GNN) are becoming more and more prevalent. [7]. Edge intelligence has arisen as a promising computing paradigm for supporting miscellaneous smart applications that rely on machine learning techniques [8].Many underlying relationships among data in several areas of science and engineering, e.g., computer vision, molecular chemistry, molecular biology, pattern recognition, and data mining, can be represented in terms of graphs [9].

2. Related Work

Review of IDS Methodologies: The various intrusion detection methodologies, including traditional ML algorithms like K-nearest neighbors (KNN), Support Vector Machines (SVM), Random Forests (RF), as well as DL techniques such as deep Autoencoders and Long Short-Term Memory (LSTM) cells.

Ansam Khraisat et al.'s work on the evolution of malware sophistication can be expanded to examine the latest trends and techniques used by cybercriminals to evade detection. In their paper "Evolution of Malware: A Comprehensive Analysis [10]. David Pujol-Perich et al.'s work highlights the potential of GNNs in achieving high accuracy in NIDS. They provide a comprehensive survey of GNN-based NIDS, discussing their advantages and limitations, and highlighting the need for large amounts of labeled data and the potential for adversarial attacks.

Miltiadis Allamanis et al.'s work on using structured graph-based code encoding for program semantics can be further explored to investigate its potential in malware detection and other security applications. They propose a graph-based neural network model for learning program semantics, achieving state-of-the-art results on several code understanding tasks.

3. Theory

Graph Neural Networks
A kind of neural networks called Graph Neural Networks (GNNs) were designed to operate on data that is organized as graphs. Nodes, which stand for entities or objects, and edges, which represent connections or relationships between nodes, make up graphs. GNNs use the graph structure to carry out a number of functions, such as graph formation, node classification, link prediction, and graph classification.

3.1 Why GNN?
Graphs representing network traffic data are analyzed using Graph Neural Networks (GNNs). The ability of GNNs to capture intricate interactions between network components, which encapsulates relationships between entities makes it possible to identify abnormalities and questionable activity. GNNs operate by propagating information through the nodes of a graph, leveraging both node features and edge connections to learn representations that capture the complex relational dependencies within the data. This enables GNNs to perform tasks such as node classification, link prediction, and graph classification across various domains including social networks, bioinformatics, recommendation systems, and knowledge graphs. GNNs have garnered significant attention due to their ability to effectively model structured data, offering promising avenues for advancing machine learning techniques in scenarios where understanding relationships is crucial. The project intends to improve network security by utilizing GNNs to identify and mitigate cyber threats in real-time, eventually securing digital assets and guarding against possible security breaches.

They utilize a message-passing mechanism where nodes iteratively exchange information with their neighbors, allowing for the aggregation of local neighborhood information and enrichment of node representations. GNNs learn continuous vector representations (embeddings) for each node, capturing both structural and feature information along with their neighbors.

3.2 Graph Representation Learning for IDS: Graph Neural Networks (GNNs) are gaining attention for intrusion detection due to their ability to capture complex structural patterns in data, particularly in graph-structured data representing interactions between entities like hosts and processes.

3.3 BoT-IoT - Data Set
In the UNSW Canberra Cyber Range Lab, a realistic network environment was designed to create the BoT-IoT dataset. Botnet and legitimate traffic coexisted in the network environment. The source files for the dataset are available in a variety of forms, including csv files, produced argus files, and original pcap files. The files were separated, based on attack category and subcategory, to better assist in labelling process.

4. Experimental Method/Procedure/Design

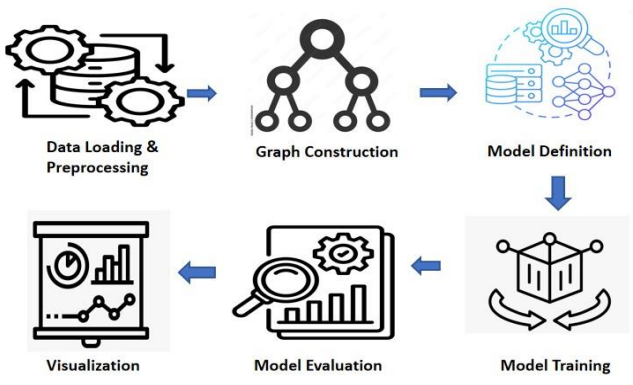

Figure 2: System Architecture

Procedure

Data Preprocessing: To address the network security classification task, we initiated the process by preprocessing the raw network data obtained from a CSV file. This involved several sub-steps such as removing irrelevant columns, renaming the target variable, and handling class imbalance by sampling data from each class. Categorical variables were encoded, and numerical features were scaled using standardization techniques. Subsequently, the pre-processed data was split into training and testing sets to facilitate model evaluation.

Graph Construction: With the preprocessed data in hand, we proceeded to construct a graph representation suitable for training a Graph Neural Network (GNN). Using the NetworkX library, we constructed a directed graph where nodes represented network entities such as IP addresses, and edges represented connections between them. This NetworkX graph was then converted into a DGL (Deep Graph Library) graph to leverage the functionalities provided by DGL for GNN training. Initial node and edge features were assigned to the graph to initialize the model.

Model Definition: Following graph construction, we defined the architecture of the Graph Neural Network (GNN) model. The model consisted of multiple SAGE (Graph SAGE) layers for message passing and an MLP (Multi-Layer Perceptron) predictor for classification. We defined the loss function (Cross-Entropy Loss) and selected the Adam optimizer for training the model parameters.

Model Training: The GNN model was trained iteratively over multiple epochs. Each training iteration involved a forward pass to compute predictions, followed by the computation of loss using the defined loss function. Back-Propagation was then performed to update the model parameters, with the Adam optimizer adjusting the weights to minimize the loss. Throughout the training process, we monitored training accuracy to assess the model's convergence and performance.

Model Evaluation: Once the model was trained, we evaluated its performance on the test dataset. Performance metrics such as accuracy, precision, recall, and F1-score were computed to assess the model's classification performance. Additionally, a confusion matrix was visualized to gain insights into the model's classification results and identify any potential misclassifications.

Visualization and Analysis: Lastly, we utilized UMAP (Uniform Manifold Approximation and Projection) for dimensionality reduction of learned embeddings. The embeddings were visualized to analyze clustering patterns and assess the separation of different network entities in the learned feature space.

5. Results and Discussion

The GNN model achieved significant accuracy in classifying network traffic, effectively distinguishing between different types of cyber attacks, we found distinct patterns for detecting DDoS, reconnaissance, and theft attacks. DDoS attacks were evident from traffic spikes, causing service disruptions. Reconnaissance showed up as unusual scanning activities, signaling attempts to identify vulnerabilities. Theft attacks were characterized by unauthorized access and data breaches as demonstrated by the confusion matrix and classification report.

Distinct Patterns for Different Attacks:
DDoS Attacks: These attacks were characterized by traffic spikes leading to service disruptions. The model effectively detected these anomalies associated with DDoS attacks.

Reconnaissance Attacks: Unusual scanning activities were observed, indicating attempts to identify vulnerabilities in the network. The model successfully identified these reconnaissance activities.

Theft Attacks: Unauthorized access and data breaches were evident, showcasing the model's capability to detect unauthorized access and data exfiltration attempts.

Confusion Matrix: The confusion matrix provides a visual representation of the model's performance in classifying different attack types. It illustrates the number of true positive, true negative, false positive, and false negative predictions, allowing for a comprehensive assessment of the model's accuracy.

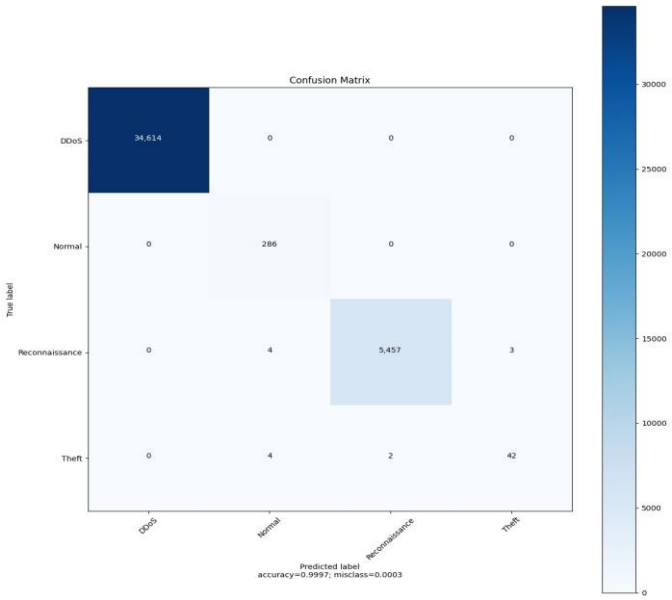

Figure 3: Confusion Matrix

	precision	recall	f1-score	support
DDoS	1.0000	1.0000	1.0000	34614
Normal	0.9728	1.0000	0.9862	286
Reconnaissance	0.9996	0.9987	0.9992	5464
Theft	0.9333	0.8750	0.9032	48
accuracy			0.9997	40412
macro avg	0.9764	0.9684	0.9722	40412
weighted avg	0.9997	0.9997	0.9997	40412

Figure 4: Performance Metrics

In the graphical representation, each metric (precision, recall, and F1 score) is represented by a bar, allowing for easy comparison. This presentation effectively communicates the performance of our model in a clear and visually appealing manner.

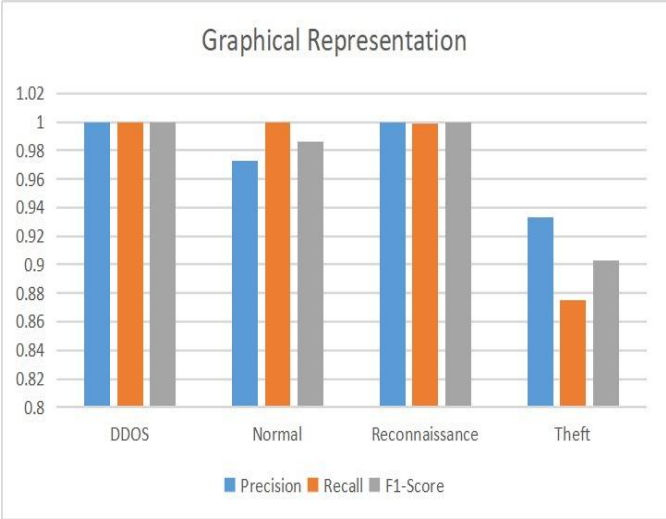

Figure 5: Graphical Representation

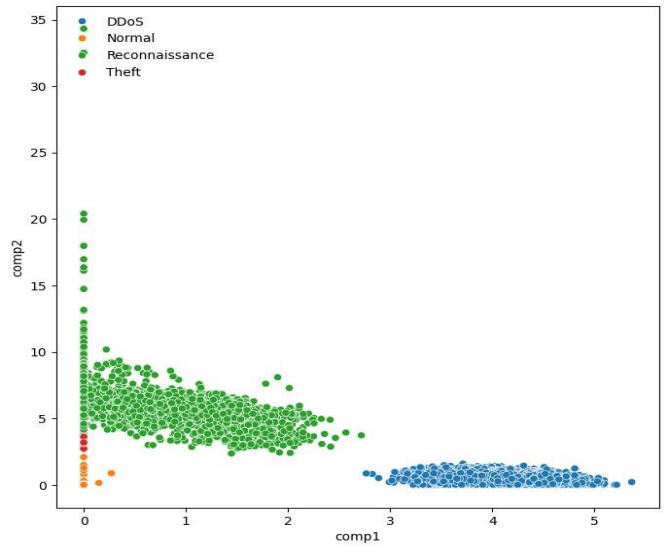

Figure 6: Visual Representation of Attacks

The Uniform Manifold Approximation and Projection (UMAP) visualization showcased meaningful clustering patterns in the learned node embeddings, indicating potential groupings of network addresses based on their features.
The table below shows the accuracy achieved by the GNN model on both datasets

Table 1: Comparision of both datasets

DATASET	ACCURACY
BOT-IOT	99.99%
TON-IOT	95.32%

Bot-IoT Dataset: The GNN model achieved an exceptionally high accuracy of 99.99% on the Bot-IoT dataset, indicating its outstanding performance in detecting network intrusions in this specific IoT network environment.

Ton-IoT Dataset: The model demonstrated strong performance with an accuracy of 95.32% on the Ton-IoT dataset, showcasing its capability to identify network intrusions in different IoT network scenarios.

The comparative analysis highlights the GNN model's robust performance in detecting network intrusions across diverse IoT network scenarios represented by the Bot-IoT and Ton-IoT datasets. The model's exceptionally high accuracy on the Bot-IoT dataset and strong performance on the Ton-IoT dataset underscore its effectiveness in network intrusion detection.

6. Conclusion and Future Scope

6.1 Conclusion

The outcomes of our experimental analysis reveal that the newly curated dataset not only surpasses previous ones but also encompasses a broader spectrum of cyber risks, thereby providing a more comprehensive foundation for our research. Employing Graph Neural Networks (GNNs) as the cornerstone of our approach, we aim to bolster network security by harnessing their capabilities in analyzing data and detecting threats with unparalleled efficacy.

GNNs excel in capturing the intricate relationships inherent in network data, enabling us to achieve thorough and nuanced threat detection. This is complemented by our system's capability for real-time monitoring, ensuring that potential threats are identified and addressed swiftly, thus mitigating potential risks before they escalate. Moreover, the adaptability of our solution to evolving network dynamics and its scalability to manage networks of varying sizes further solidify its effectiveness in safeguarding against cyber threats. By leveraging GNNs within our solution, we are not only advancing the state-of-the-art in network security but also contributing to the ongoing efforts to fortify Cyber Security measures in today's rapidly evolving digital landscape.

6.2 Future Scope

Preventive Measures: Incorporating proactive measures into the network intrusion detection system is imperative. By embedding prevention mechanisms, organizations can bolster their cyber defenses, mitigate vulnerabilities, and diminish the repercussions of security breaches, thus augmenting the overall resilience of their network infrastructure.

Optimized Model Architecture: Explore advanced GNN architectures, such as Graph Convolutional Networks (GCNs), Graph Attention Networks (GATs), or Graph Neural Networks (GGNs), to delve deeper into the intricate relationships within network data. Enhancing the model's architecture allows for more nuanced analysis, leading to heightened accuracy in threat detection and mitigation.

Dataset Enrichment: Broaden the dataset utilized for training and assessment by incorporating a diverse array of network traffic samples, including emerging threats and attack scenarios. This expansion ensures that the model is

exposed to a comprehensive range of potential threats, enhancing its adaptability and robustness in navigating evolving security landscapes.

Conflict of Interest
There isn't any conflict of interest between us.

Funding Source
None

Author's Contribution
Bhavya Lahari Vaddempudi took charge of data collection and preprocessing, ensuring the datasets were meticulously cleaned and prepared for analysis. Amrutha Tulabandu led the development of the Graph Neural Network (GNN) model, fine-tuning its parameters and architecture to achieve optimal performance. S.N.B Tanuja Reddy conducted comprehensive literature reviews, gathering valuable insights to inform the project's methodology and approach. Deepika Leela Pudi played a pivotal role in result analysis, interpreting the model's predictions and deriving meaningful conclusions.

Acknowledgements
We sincerely thank our guide, Venkata Narayana Yerininti, whose invaluable guidance and support were pivotal in navigating the complexities of the project, leading to its successful completion.

References

[1] A. R. Mathew, A. Al Hajj and K. Al Ruqeishi, "Cyber crimes: Threats and protection," 2010 International Conference on Networking and Information Technology, Manila, Philippines, pp.**16-18, 2010**. doi: 10.1109/ICNIT.2010.5508568.
[2] N. Aggarwal, M. Sehgal and A. Arya, "An empirical analysis of Cyber Crimes, their prevention measures, and laws in India," 2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC), Solan, Himachal Pradesh, India, pp.**570-575, 2022.** doi: 10.1109/PDGC56933.2022.10053354.
[3] Anupreet Kaur Mokha. A Study on Awareness of Cyber Crime and Security. Research J. Humanities and Social Sciences. October -December, Vol.**8**, Issue.**4**, pp.**459-464, 2017.** doi: 10.5958/2321-5828.2017.00067.5
[4] S. Angra and S. Ahuja, "Machine learning and its applications: A review," 2017 International Conference on Big Data Analytics and Computational Intelligence (ICBDAC), Chirala, Andhra Pradesh, India, pp.**57-60, 2017.** doi: 10.1109/ICBDACI.2017.8070809.
[5] D. Kataria et al., "Artificial Intelligence And Machine Learning," 2022 IEEE Future Networks World Forum (FNWF), Montreal, QC, Canada, pp.**1-70, 2022.** doi: 10.1109/FNWF55208.2022.00133.
[6] M. Gorricho-Segura, X. Echeberria-Barrio and L. Segurola-Gil, "Edge-based Analysis for Network Intrusion Detection using a GNN Approach," 2023 JNIC Cybersecurity Conference (JNIC), Vigo, Spain, pp.**1-7, 2023.** doi: 10.23919/JNIC58574.2023.10205384.
[7] L. Zeng, C. Yang, P. Huang, Z. Zhou, S. Yu and X. Chen, "GNN at the Edge: Cost-Efficient Graph Neural Network Processing Over Distributed Edge Servers," in IEEE Journal on Selected Areas in Communications, March, Vol.**41**, No.**3**, pp.**720-739, 2023**. doi: 10.1109/JSAC.2022.3229422.
[8] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner and G. Monfardini, "The Graph Neural Network Model," in IEEE Transactions on Neural Networks, Jan., Vol.**20**, No.**1**, pp.**61-80, 2009.** doi: 10.1109/TNN.2008.2005605.
[9] M. Schneider, J. Weißenbach and U. Schmid, "Substrate Temperature and Bias Voltage Dependent Properties of Sputtered AlN Thin Films for BAW Applications," 2018 IEEE 18th International Conference on Nanotechnology (IEEE-NANO), Cork, Ireland, pp.**420-425, 2018.** doi: 10.1109/NANO.2018.8626332.
[10] Ankita Sharma, "Review on Major Cyber security Issues in Educational Sector," International Journal of Computer Sciences and Engineering, Vol.**9**, Issue.**12**, pp.**26-29, 2021.**

AUTHORS PROFILE

Bhavya Lahari Vaddempudi pursuing IV B. tech in the stream of Information Technology at Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dist., Andhra Pradesh.

Amrutha Tulabandu pursuing IV B. tech in the stream of Information Technology at Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dist., Andhra Pradesh.

S.N.B Tanuja Reddy pursuing IV B. tech in the stream of Information Technology at Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dist., Andhra Pradesh.

Deepika Leela Pudi pursuing IV B. tech in the stream of Information Technology at Vasireddy Venkatadri Institute of Technology, Nambur, Guntur Dist., Andhra Pradesh.

Venkata Narayana Yerininti working as Assistant Professor in the stream of Information Technology at Vasireddy Venkatadri Institute of Technology, Nambur, Guntur District, Andhra Pradesh.

