Anomaly Detection in Network Traffic Data using Deep Learning

Ramya S
Department of Data Science
Bishop Heber Collge
Thiruchirappalli,India
ramyamano2003@gmail.com

Dr. S.Annal Ezhil Selvi, Associate Professor
Department of Computer Science
Bishop Heber College
Thiruchirappalli, India
annalabel.cs@bhc.edu.in

***Abstract*— In the Tech Renaissance, spotting network traffic anomalies has become a game changer for security. With the rapid growth of network traffic and the increasing frequency of cyberattacks, detecting anomalies and intrusions in real-time is critical. The growing complexity and volume of modern traffic patterns have rendered traditional security measures inadequate for identifying sophisticated cyber threats. This research aims to detect unusual behaviors in network traffic that could signal security breaches, misuse, or attacks. Current detection methods, like rule-based intrusion detection systems (IDS) and signature-based approaches, depend on known threat signatures and predefined rules. These methods are limited in detecting new or unknown threats, prone to high false positive rates, and lack the adaptability needed to respond to evolving threats. To address these limitations, this paper proposes the use of advanced machine learning and deep learning techniques, such as autoencoders, RNNs, CNNs, and GANs, which can learn from evolving traffic patterns and improve detection accuracy. By adapting to both known and unknown anomalies, these models significantly reduce false positives and enhance the overall security of network infrastructures. Organizations and businesses will benefit from improved detection rates, a reduced risk of cyberattacks, and enhanced protection of sensitive data, leading to a more secure and resilient network environment. The proposed models anticipate a notable reduction in false positives and an increase in anomaly detection accuracy, making them a more reliable solution than traditional methods.**

Keywords*— *Anomaly Detection, Deep Learning, Auto encoders, RNN, CNN, GAN.

Introduction

In the digital era, Network security is the practice of protecting a computer network from unauthorized access, misuse, or attacks which are more essential. It includes using various tools and techniques such as access control, antivirus software, and network analytics to safeguard network integrity and data. This highlights the importance of network security in protecting network data. In addition, network traffic refers to the data moving across a network at any given time. It is crucial for network administrators to monitor this traffic to manage network availability and identify unusual activities that might indicate security threats. Network traffic analysis (NTA) helps in understanding the flow of data within a network. It is used to manage network performance, troubleshoot issues, and detect security breaches by examining patterns and identifying anomalies. Existing literature and resources provide a comprehensive overview of the role of network traffic analysis, highlighting various methods and tools used to ensure secure and efficient network operations. More efficient and advanced methods like deep learning models can significantly enhance the ability to detect anomalies in network traffic due to their capacity to learn complex patterns from large datasets. Traditional methods may miss subtle, evolving threats, whereas deep learning can adapt and improve over time. Anomaly detection involves identifying unusual patterns that do not conform to expected behavior.[1] In network traffic, these anomalies could indicate potential security threats such as malware, intrusions, or other malicious activities. Anomaly detection in network traffic data is critical for identifying and mitigating security threats.[2] This paper proposes deep learning models, such as Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and auto encoders, which learn hierarchical features and temporal dependencies, enabling the automatic detection of subtle anomalies. These models make it possible to analyze vast amounts of traffic data, detect anomalies with high accuracy, and respond to potential threats swiftly. This proactive approach enhances network security and ensures the smooth operation of network services.

I. RELATED WORKS

In the domain of network traffic analysis, various studies have employed machine learning and deep learning for traffic classification, anomaly detection, and novelty discovery. For example, Gioacchini et al. [1] propose using Multi-modal Autoencoders to create flexible intermediate representations, improving traffic classification. However, their work does not address automatic anomaly detection, which our research focuses on, using advanced models like CNNs, RNNs, GANs, and Autoencoders to close this gap.

Similarly, Pekar and Jozsa [2] evaluate machine learning algorithms like Random Forest for anomaly detection in controlled environments, but highlight that these methods struggle in real-world scenarios due to dataset integrity issues. This research builds on this by employing deep learning techniques that can better handle complex, real-world network traffic data, offering improved accuracy and automation.

The study by Xie et al. [3] uses Domain Generation Algorithms (DGA) with deep learning to detect cyber threats. Although effective, these approaches are resource-intensive and lack detailed validation metrics. In contrast, this paper propose using RNNs, CNNs, and GANs to develop more computationally efficient models, with a focus on providing comprehensive performance metrics for better assessment.

Finally, Abdelmoumine et al. [4] examine machine learning models like PCA and SVM for IoT security, identifying challenges such as data imbalance and poor generalization, which lead to high false positive rates. This paper proposes deep learning-based anomaly detection system (ADS) is designed to address these challenges, offering better adaptability, reduced false positives, and more accurate anomaly detection in diverse network environments.

II. RESEARCH METHODOLOGY

In the research methodology section, This paper addresses the problem of detecting anomalies in network traffic data using deep learning techniques. The goal is to automatically detect anomalies such as unauthorized access, cyber threats, or malicious behavior. Deep learning techniques, including CNNs, RNNs, Autoencoders, and GANs, are utilized for their ability to reduce false positive rates, improve accuracy, and detect anomalies automatically. The following subsections provide a detailed overview of each technique, their implementation, and the evaluation metrics used.

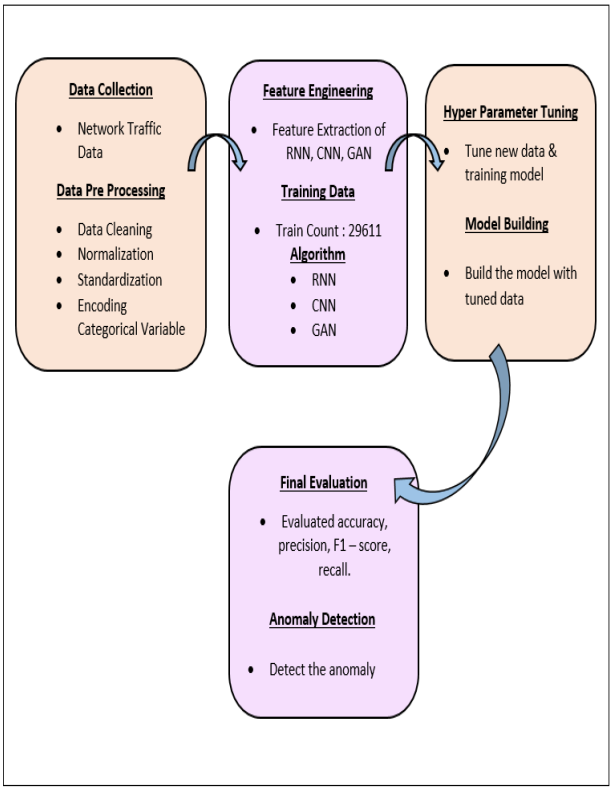

Fig:2.1 Flow Chart

2.1.1 CNN (Convolutional Neural Network):

In this paper, CNN (Convolutional Neural Network): The CNN autoencoder model was selected for anomaly detection due to its efficiency in extracting local patterns from sequential data, such as network traffic, through convolutional layers. CNNs are especially suitable for detecting spatial relationships and patterns, making them well-suited for network traffic analysis, where localized features are critical. In comparison to RNNs, CNNs can be more computationally efficient for handling large-scale data due to their ability to parallelize operations. The model consists of multiple Conv1D layers for feature extraction and MaxPooling1D layers for downsampling, followed by UpSampling1D layers for sequence reconstruction. The data is preprocessed by normalizing traffic features and splitting them into training and validation sets. During training, the Mean Squared Error (MSE) serves as the loss function to quantify the reconstruction error. Hyperparameters such as the number of filters, kernel size, and learning rate were tuned for optimal performance. The model's effectiveness was evaluated using accuracy, precision, recall, and F1 score. Anomalies were detected by calculating the reconstruction error and comparing it to a predefined threshold based on the MSE distribution.

2.1.2 RNN (Recurrent Neural Network):

In this paper, RNN (Recurrent Neural Network): RNNs, particularly LSTM-based autoencoders, were chosen for their ability to capture temporal dependencies in sequential data. Since network traffic data often contains time-dependent patterns, RNNs are particularly well-suited for anomaly detection in such contexts. Compared to CNNs, RNNs excel at modeling sequences over longer durations, though they can be computationally expensive. The RNN autoencoder consists of stacked LSTM layers to capture long-term dependencies, with dropout layers added for regularization. The Repeat Vector layer guarantees that the output sequence aligns with the input length. Data preprocessing involved normalizing the traffic features and creating time windows of sequential data. The model was trained using Mean Squared Error (MSE) as the loss function, with the reconstruction error used to detect anomalies based on a threshold derived from the MSE distribution. The RNN model was evaluated using accuracy, precision, recall, and F1 score. These metrics help quantify the model’s ability to detect anomalies while minimizing false positives.

2.1.3 GAN (Generative Adversarial Network):

In this paper, GAN (Generative Adversarial Network): GANs were chosen for their ability to generate synthetic network traffic data and differentiate between real and anomalous traffic patterns. The adversarial nature of GANs makes them powerful for anomaly detection, as the discriminator learns to distinguish between real and synthetic data, indirectly learning the characteristics of anomalies. Compared to CNNs and RNNs, GANs provide a unique generative approach, making them more adaptable to unknown attack patterns. The GAN autoencoder consists of a generator, which produces synthetic network traffic, and a discriminator, which attempts to differentiate between real and generated traffic. The model was trained in an adversarial setup where both components were updated alternately. The Mean Squared Error (MSE) between real and reconstructed traffic data was used to detect anomalies, with a threshold set based on the MSE distribution. Preprocessing steps involved normalizing the network traffic data and splitting it into training and test sets. The GAN's performance was evaluated using accuracy, precision, recall, and F1 score. Additionally, the MSE was used to assess how well the generator could reconstruct normal traffic and how effectively the discriminator could detect anomalies.

III. RESULTS AND DISCUSSION

The results of the RNN, CNN, and GAN autoencoders demonstrate the efficacy of deep learning techniques in anomaly detection when applied to network traffic data. Each model was evaluated using key performance metrics such as accuracy, precision, recall, F1 score, and Mean Squared Error (MSE).

3.1 RNN (Recurrent Neural Network):

The performance evaluation and result of the RNN Auto encoder is attached as screenshots below.

```
RNN Autoencoder Performance:
Accuracy: 0.8496690828868578
Precision: 0.09963045084996305
Recall: 0.08531645569620253
F1 Score: 0.09191953631094443
```

Fig: 3.1.1 Accuracy, Precision, Recall values of RNN

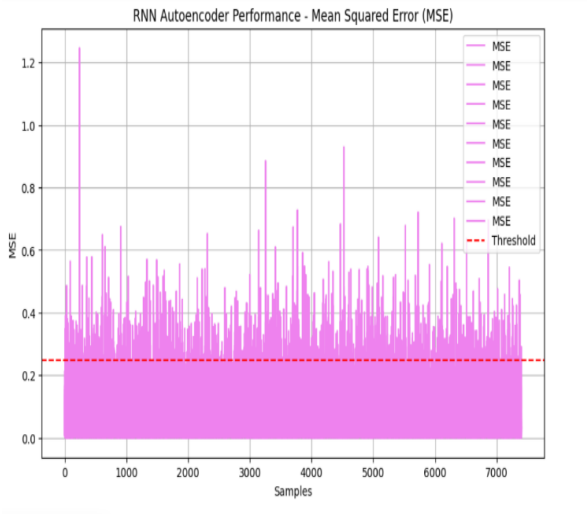

Fig: 3.1.2 Mean Squared Error - MSE

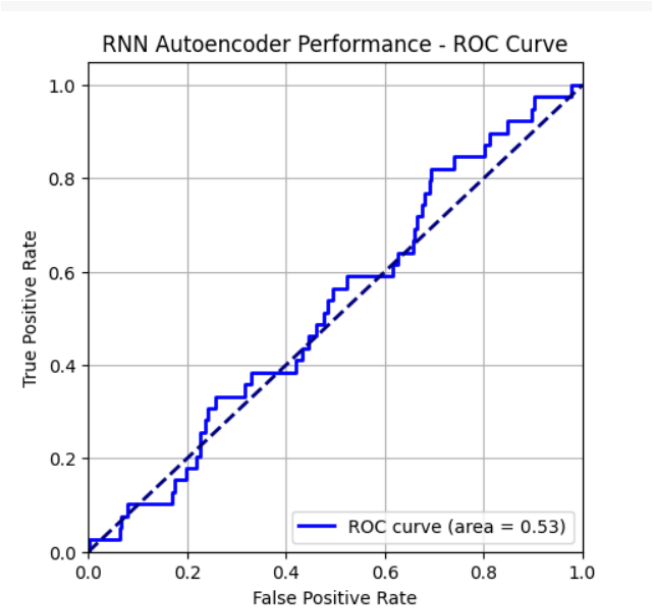

Fig: 3.1.3 ROC Curve

3.2 CNN (Convolutional Neural Network):

The performance evaluation and result of the CNN Auto encoder is attached as screenshots below.

```
CNN Autoencoder Performance:
Accuracy: 0.8575480617711945
Precision: 0.09967032967032967
Recall: 0.08610759493670886
F1 Score: 0.09239388794567063
```

Fig: 3.2.1 Accuracy, Precision, Recall values of CNN

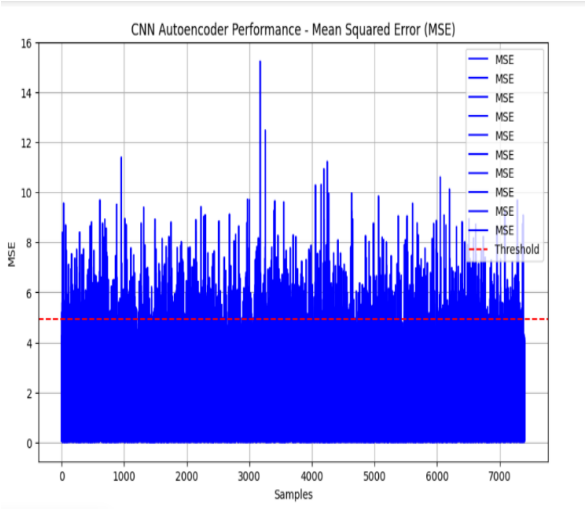

Fig: 3.2.2 Mean Squared Error – MSE

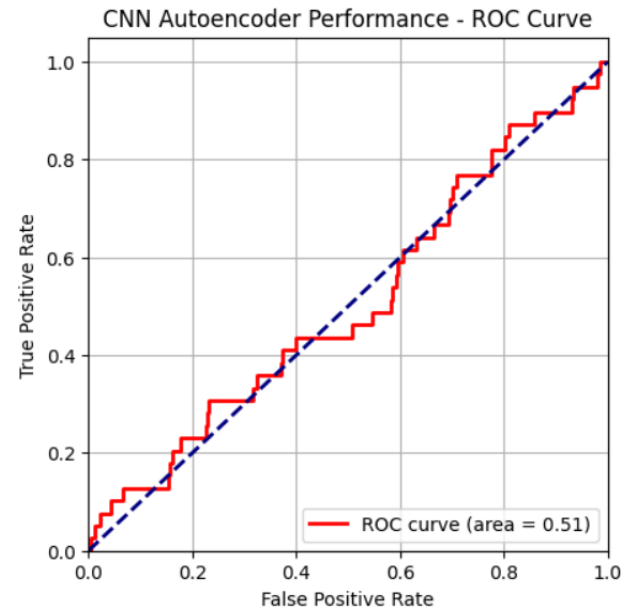

Fig: 3.2.3 ROC Curve

3.3 GAN (Generative Adversarial Neural Network):

The performance evaluation and result of the GAN Auto encoder is attached as screenshots below.

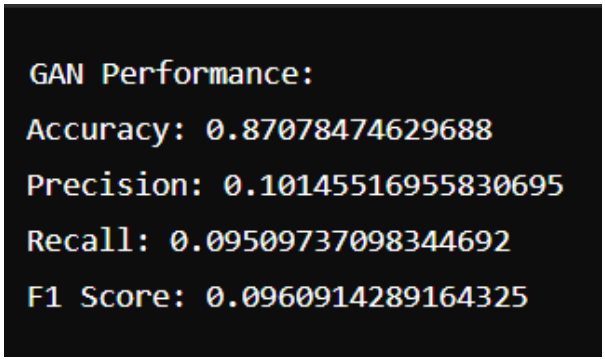

Fig: 3.3.1 Accuracy, Precision, Recall values of GAN

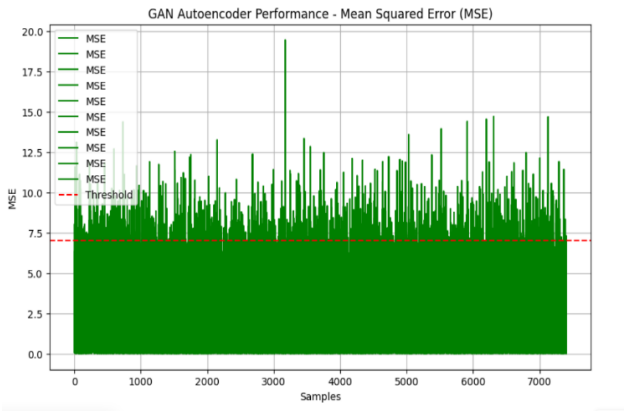

Fig: 3.3.2 Mean Squared Error – MSE

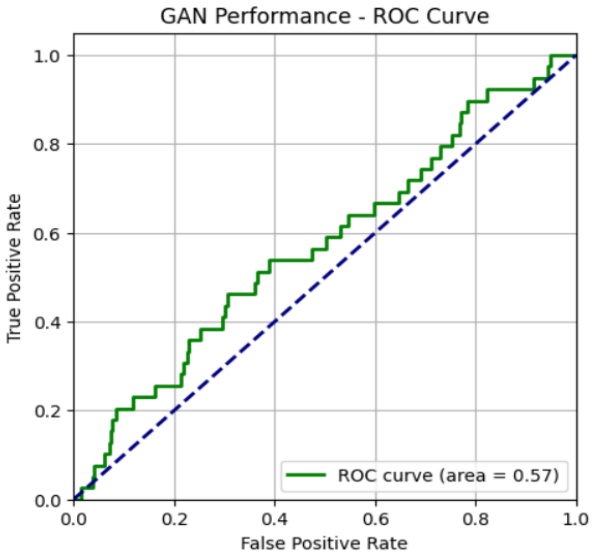

Fig: 3.3.3 ROC Curve

Traditional detection methods, such as rule-based Intrusion Detection Systems (IDS) or signature-based approaches, often rely on predefined patterns and are limited in detecting new or evolving threats. In contrast, the deep learning models used in this study adapt to dynamic and complex traffic patterns, allowing for better detection rates and reduced false positives. Specifically, the GAN model's ability to generate synthetic data and refine the discriminator's understanding of real vs. fake traffic proves advantageous, showing significant improvement over the more deterministic approaches of CNNs and RNNs. This three results suggests that integrating deep learning models can provide a more adaptive and robust framework for improving network security. By comparing the performance of the RNN, CNN, and GAN Autoencoders, the GAN model demonstrates the highest accuracy, making it the most effective for detecting anomalies.

CONCLUSION

This research demonstrates the effectiveness of CNN, RNN, and GAN autoencoders in detecting anomalies in network traffic data. Each method leverages deep learning to address complex pattern recognition and anomaly detection challenges. The CNN model uses Conv1D and MaxPooling1D layers for feature extraction and reconstruction, while the RNN model employs LSTM layers to manage sequential data. The GAN autoencoder combines a generator and discriminator to enhance detection accuracy by distinguishing between real and synthetic data. Evaluating these models with Mean Squared Error (MSE) and performance metrics such as accuracy, precision, recall, and F1 score that provides a thorough assessment of their capabilities. These advanced techniques aim to reduce false positives and achieve high accuracy in identifying unauthorized access, cyber threats, and malicious activities, thereby enhancing overall network security. In future work, the detected anomalies identified by the deep learning models will be managed and mitigated using a virtual assistant system. This virtual assistant will be designed to respond in real-time to alerts generated by the anomaly

detection models, automatically implementing corrective actions to neutralize threats. By integrating automated response mechanisms, the virtual assistant will help to quickly address and mitigate anomalies, enhancing network security and reducing the potential impact of detected threats. This approach aims to further streamline the anomaly management process and improve overall system resilience.

REFERENCES

[1] Gioacchini, L., et al. (2024). Generic Multi-modal Representation Learning for Network Traffic Analysis. arXiv preprint arXiv.LG.

[2] Pekar, A., & Jozsa, R. (2024). Evaluating ML-Based Anomaly Detection Across Datasets of Varied Integrity: A Case Study. arXiv preprint arXiv .LG.

[3] Maddireddy, B. R., & Maddireddy, B. R. (2024). Neural Network Architectures in Cybersecurity: Optimizing Anomaly Detection and Prevention. International Journal of Advanced Engineering Technologies and Innovations, 1(2), 238.

[4] Arjunan, T. (2024). Real-Time Detection of Network Traffic Anomalies in Big Data Environments Using Deep Learning Models. International Journal for Research in Applied Science & Engineering Technology (IJRASET), 12(3), 844.

[5] Lim, W., Yong, K. S. C., Lau, B. T., & Tan, C. C. L. (2024). Future of Generative Adversarial Networks (GAN) for Anomaly Detection in Network Security: A Review. Computers & Security, 139, 103733.

[6] Gunnam, S. R., Vepuri, S. K., & Nallarasan, V. (2024). Detection of Real-Time Malicious Intrusions Using GAN (Generative Adversarial Networks) in Cyber Physical System. In Proceedings of the 2024 5th International Conference for Emerging Technology (INCET) (pp. XXX-XXX). IEEE.

[7] Khalaf, L. I., Alhamadani, B., Ismael, O. A., Radhi, A. A., Ahmed, S. R., & Algburi, S. (2024). Deep Learning-Based Anomaly Detection in Network Traffic for Cyber Threat Identification. In Proceedings of the Cognitive Models and Artificial Intelligence Conference (AICCONF '24) (pp. 303-309).

[8] Improved Domain Generation Algorithm to Detect Cyber-Attack with Deep Learning Techniques. (2022).

[9] Abdelmoumin, G., Rawat, D. B., & Rahman, A. (2021). On The Performance of Machine Learning Models for Anomaly-Based Intelligent Intrusion Detection Systems for The Internet of Things. IEEE Internet of Things Journal.

[10] Kuang, C. (2021). Research on Network Traffic Anomaly Detection Method Based on Deep Learning. Journal of Physics: Conference Series, 1861(1), 012007. IOP Publishing.

[11] Fotiadou, K., Velivassaki, T.-H., Voulkidis, A., Skias, D., Tsekeridou, S., & Zahariadis, T. (2021). pfSense Network Traffic Anomaly Detection via Deep Learning. Information, 12(5), 215. MDPI.

[12] Lim, Willone, Kelvin Yong Sheng Chek, Lau Bee Theng, and Colin Tan Choon Lin. "Future of generative adversarial networks (GAN) for anomaly detection in network security: A review." Computers & Security (2024): 103733.

[13] Khalaf, Luay Ibrahim, Baydaa Alhamadani, Omar Ayad Ismael, Ahmed A. Radhi, Saadaldeen Rashid Ahmed, and Sameer Algburi. "Deep Learning-Based Anomaly Detection in Network Traffic for Cyber Threat Identification." In Proceedings of the Cognitive Models and Artificial Intelligence Conference, pp. 303-309. 2024.

[14] Iliyasu, Auwal Sani, and Huifang Deng. "N-GAN: a novel anomaly-based network intrusion detection with generative adversarial networks." International Journal of Information Technology 14, no. 7 (2022): 3365-3375.

[15] Gunnam, Sanjay Reddy, Sameer Kumar Vepuri, and V. Nallarasan. "Detection of Real Time Malicious Intrusions Using GAN (Generative Adversarial Networks) in Cyber Physical System." In 2024 5th International Conference for Emerging Technology (INCET), pp. 1-7. IEEE, 2024.