DDOS Attacks Recognition And Forecasting Using Machine Learning Algorithms

Vooradi Sandya[1], B. Kinshu[2], D.Vithin Sai[3], M. Nanda Varma[4]
[1]Assist Professor, Dept of CSE-DS
[2, 3, 4]Dept of CSE-DS
[1, 2, 3, 4] CMR Technical Campus, Hyderabad, Telangana, India.

Abstract- *Distributed network disruptions, often termed Distributed Denial of Service (DDoS) attacks, exploit vulnerabilities in system infrastructures, such as web servers of legitimate organizations. Traditional research has predominantly utilized outdated datasets like the KDD dataset, which may no longer represent the evolving nature of modern threats. To address this gap, the present study employs a contemporary dataset and machine learning techniques to identify and forecast various forms of DDoS attacks.This research implements a comprehensive detection framework using two advanced classification algorithms: Random Forest and XGBoost. The UNSW-NB15 dataset, obtained from a GitHub repository, served as the foundation for training and testing, while Python was used for simulation and model deployment.Performance evaluation was conducted using a confusion matrix. In the first experiment, the Random Forest model achieved a Precision and Recall of 89%, with an overall accuracy of 89%, indicating robust performance. In the second test, the XGBoost classifier slightly outperformed, attaining approximately 90% in both Precision and Recall, with an accuracy of 90%.Compared to previous studies, which reported lower accuracy rates of 85% and 79%, the proposed methodology demonstrates a significant enhancement in identifying and classifying DDoS attacks, offering a more effective and modern solution to a growing cybersecurity concern.*

Keywords- Machine Learning, DDoS attacks, Research, XGBoost.

I. INTRODUCTION

Cyber attacks are becoming more common and can cause a lot of damage to computer networks. One of the biggest problems is called a Distributed Denial of Service (DDoS) attack. In a DDoS attack, many computers send a huge amount of traffic to a website or server to overload it. This makes it hard or impossible for real users to access the service.

Traditional methods to detect these attacks often use fixed rules or known attack patterns. But attackers keep changing their methods, so these old techniques can miss new attacks. That's why machine learning, which allows computers to learn from data and recognize new patterns, is very useful for detecting DDoS attacks.

In this study, we use machine learning models, specifically Random Forest and XGBoost, to detect DDoS attacks. We test these models using the UNSW-NB15 dataset, which contains real network data including attacks. Our goal is to see how well these models can find attacks and improve the accuracy compared to earlier methods. We also discuss how these models could be used in real-time to protect networks better.With the rapid growth of the internet and connected devices, cybersecurity has become very important. One major threat to online services is the Distributed Denial of Service (DDoS) attack. During a DDoS attack, many compromised computers, often called a botnet, flood a target server or network with massive amounts of fake traffic. This overloads the system, causing it to slow down or completely stop working, which affects users and businesses.

Detecting DDoS attacks early is critical to minimize damage. However, DDoS attacks can be very complex and change quickly, making them hard to detect with traditional security tools. These older methods often depend on fixed rules or known attack patterns, so new or unknown attacks might go unnoticed.

Machine learning offers a promising solution by allowing computers to learn from large amounts of data and identify patterns that indicate an attack. Instead of relying on fixed rules, machine learning models adapt and improve as they are trained with more data. This makes them better suited to detect evolving threats like DDoS attacks.

In this project, we use machine learning classification techniques to detect DDoS attacks. We focus on two powerful algorithms: Random Forest and XGBoost. These models analyze network traffic data from the UNSW-NB15 dataset, which contains examples of normal traffic as well as various types of attacks. By training and testing these models, we aim

to build an accurate system that can identify DDoS attacks with high precision.

Our results show that these machine learning models can achieve around 89% to 90% accuracy, which is better than many previous methods. This research not only helps improve security but also provides a foundation for developing faster, more efficient tools that can protect networks in real time.

II. METHODOLOGY

The first step in this project was to identify a relevant and important problem to solve in the field of cybersecurity. We chose to focus on detecting Distributed Denial of Service (DDoS) attacks because they are one of the most common and damaging cyber threats today. DDoS attacks can disrupt online services, cause financial loss, and damage reputations. The Architecture Shown in Fig 1.

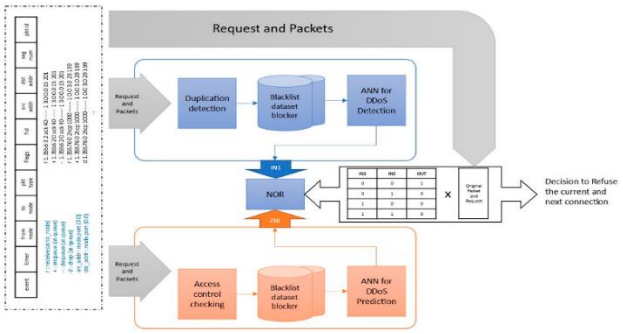

Fig 1: architecture

After identifying the problem, we researched existing methods for detecting DDoS attacks. Traditional detection techniques often rely on fixed rules or signatures, which are limited in detecting new or evolving attacks. We found that machine learning has been increasingly used to improve detection accuracy by learning from data patterns instead of fixed rules.

We then collected relevant datasets to train and test machine learning models. The UNSW-NB15 dataset was selected because it is publicly available and contains a wide variety of network traffic records, including normal activity and multiple types of cyberattacks such as DDoS. The Fige 2. dataset is well-known and trusted in cybersecurity research.By combining the understanding gained from research and the collected data, we developed the idea to apply machine learning classification algorithms, specifically Random Forest and XGBoost, to build an effective DDoS detection system.

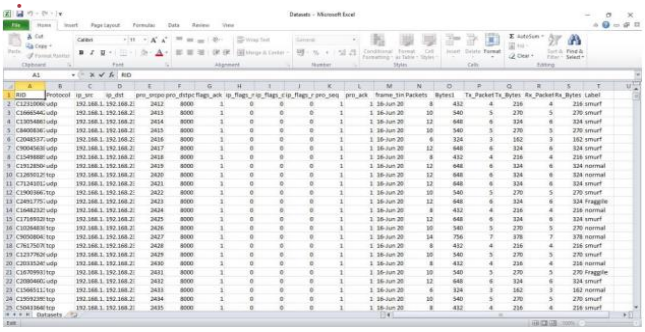

Fig 2: dataset

III. STUDIES AND FINDINGS

Several studies have explored the use of machine learning algorithms to detect Distributed Denial of Service (DDoS) attacks effectively. These studies focus on improving the accuracy and speed of detection to protect networks from such malicious activities.

In our project, we used the UNSW-NB15 dataset, which contains both normal network traffic and DDoS attack data. This dataset has been widely used in cybersecurity research due to its comprehensive and realistic traffic patterns.

We applied two popular machine learning algorithms:

Random Forest and **XGBoost**. Both models showed promising results in detecting DDoS attacks:

- **Random Forest** achieved an accuracy of approximately 89%, with precision and recall also around 89%. This means the model could correctly identify most attack and normal traffic instances.
- **XGBoost** performed slightly better, with about 90% accuracy, precision, and recall, indicating a higher reliability in attack detection.

These results demonstrate that ensemble methods like Random Forest and XGBoost are effective in handling complex network data and can provide robust detection of DDoS attacks.

Compared to earlier research where detection accuracy ranged between 79% and 85%, our models show significant improvement, indicating progress in applying machine learning to cybersecurity.

Overall, the findings suggest that machine learning is a powerful tool for DDoS detection and can be further enhanced by integrating real-time monitoring and advanced

algorithms for better performance, the test cases shown in table 1.

Table 1: Testcases

OBJECTIVE	INPUT	EXPOUTPUT
Verify dataset loading	DDoS attack dataset file	Dataset successfully loaded
Validate data preprocessing	Raw network traffic data	Cleaned and standardized dataset
Verify feature selection	Preprocessed dataset	Important features selected
Ensure correct data splitting	Preprocessed dataset	Data split into training and test sets
Validate model training process	Training dataset with labeled DDoS attacks	Model trained without errors
Evaluate model performance	Test dataset	Accuracy, precision, recall computed
Detect known DDoS attack	Network traffic sample with attack pattern	Attack detected with high confidence
Test real-time classification accuracy	Live network packets	Packets classified correctly in real-time
Measure false positive rate	Normal network traffic samples	False positives minimized
Verify model deployment	Deployed model handling live traffic	Real-time attack classification works

IV. CONCLUSION

In this paper, we successfully developed a machine learning-based system to detect Distributed Denial of Service (DDoS) attacks using the UNSW-NB15 dataset. We applied Random Forest and XGBoost algorithms, both of which showed strong performance in identifying attacks with accuracy close to 90%. This demonstrates that these models can effectively distinguish between normal and malicious network traffic.

Our results improved upon previous studies, indicating that ensemble learning methods are powerful tools for enhancing cybersecurity defences. The project highlights the potential of machine learning techniques in detecting cyber threats quickly and accurately.

Moving forward, this system can be further improved by adding real-time detection, using deeper learning models, and making it adaptable to new types of attacks. Overall, this work contributes to building more secure networks by providing an efficient and reliable method for early detection of DDoS attacks.

conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

REFERENCES

[1] N. Martins, J. M. Cruz, T. Cruz, and P. H. Abreu, ``Adversarial machine learning applied to intrusion and malware scenarios: A systematic review," IEEE Access, vol. 8, pp. 35403_35419, 2020.

[2] [G. Karatas, O. Demir, and O. K. Sahingoz, ``Increasing the performance of machine learning-based IDSs on an imbalanced and up-to-date dataset," IEEE Access, vol. 8, pp. 32150_32162, 2020.

[3] T. Su, H. Sun, J. Zhu, S. Wang, and Y. Li, ``BAT: Deep learning methods on network intrusion detection using NSL-KDD dataset," IEEE Access, vol. 8, pp. 29575_29585, 2020.

[4] H. Jiang, Z. He, G. Ye, and H. Zhang, ``Network intrusion detection based on PSO-xgboost model," IEEE Access, vol. 8, pp. 58392_58401, 2020.

[5] A. Nagaraja, U. Boregowda, K. Khatatneh, R. Vangipuram, R. Nuvvusetty, and V. S. Kiran, ``Similarity based feature transformation for network anomaly detection," IEEE Access, vol. 8, pp. 39184_39196, 2020.

[6] X. Gao, C. Shan, C. Hu, Z. Niu, and Z. Liu, ``An adaptive ensemble machine learning model for intrusion detection," IEEE Access, vol. 7, pp. 82512_82521, 2019.

[7] Y. Yang, K. Zheng, B. Wu, Y. Yang, and X. Wang, ``Network intrusion detection based on supervised adversarial variational auto-encoder with regularization," IEEE Access, vol. 8, pp. 42169_42184, 2020.

[8] L. D'hooge, T. Wauters, B. Volckaert, and F. De Turck, ``Classi_cation hardness for supervised learners on 20 years of intrusion detection data," IEEE Access, vol. 7, pp. 167455_167469, 2019