Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 16, Issue. 1: 2025
ISSN : **1906-9685**

Enhanced DDOS Detection With Hybrid Machine Learning: Voting and Stacking Classifiers

#1 K.JAYA KRISHNA, #2.VEMURI MANASA

#1 Assistant Professor, #2 MCA Scholor

DEPARTMENT OF MASTER OF APPLICATIONS

QIS COLLEGE OF ENGINEERING AND TECHNOLOGY

Vengamukkapalem(V), Ongole, Prakasam dist., Andhra Pradesh- 523272

Abstract: The rapid growth of the online population is a major danger to the safety of internet resources. The rising frequency of Denial of Service (DoS) attacks has a direct effect on security. With this growing danger, establishing a cutting-edge solution is tough from a cyber-security standpoint. In this study, we offer a machine learning-based method for detecting Distributed Denial of Service (DDoS) assaults, employing Logistic Regression, K Nearest Neighbour, and Random Forest techniques. We evaluate suggested models utilising a freshly updated NSL KDD dataset. The results of our research also show that the suggested methodology is quite good at discovering Distributed Denial of Service (DDoS) assaults. Our findings indicate that our suggested model markedly enhances existing state-of-the-art assault detection techniques.

Index Terms— DDoS Detection, Machine Learning, Logistic Regression, K-Nearest Neighbour, Random Forest, NSL-KDD Dataset, Cybersecurity, Intrusion Detection, DoS Attacks, Network Security.

1. INTRODUCTION

DDoS is a type of cybercrime in which the attacker sends a lot of traffic to a server to stop people from using online services and sites that are connected to it. The DDoS assaults do a lot of damage to the economy, government, infrastructure, and industry. DDoS assaults are a unique kind of attack in which the internet services of a specific web server are interrupted on purpose. In contrast, a DoS attack uses one device to send a lot of traffic to a target. There are three forms of DDoS assaults: volume-based attacks, protocol-based attacks, and application layer attacks. Volume-based attacks, like UDP floods and ICMP floods, try to fill up the target's bandwidth (Bps) by sending a lot of bits per second.

These assaults can make communication tools too busy to work. assaults on the application layer, such GET/POST floods and low-and-slow assaults, are measured in requests per second (Rps) and are meant to shut down the web server. People frequently think these assaults are real and harmless requests. These assaults can last anywhere from a few minutes to less than an hour, which makes them hard to find using regular methods. There are a number of machine learning techniques that can find these DDoS attacks.

2. LITERATURE SURVEY

i) DDoS Attacks in 2022: Trends and Obstacles Amid Worldwide Political Crisis:

https://www.infosecurity-magazine.com/blogs/ddos-attacks-in-2022-trends/

There is a big rise in hacker activity all around the world. Compared to the same time last year, the number of assaults throughout the world went up by 90% in the third quarter of 2022. They are also more stronger. A lot of nations have more botnets now, and it's hard to stop these kinds of attacks on your own. Politics also have a big impact on DDoS activities. At the end of February, groups of politically motivated hacktivists started organising DDoS assaults on Russian businesses in an effort to hurt the country's economy. The so-called "IT army of Ukraine" has gone after hundreds of Russian commercial and state-owned businesses and is behind the most politically driven events. They made DDoS tools that hackers all around the world are now utilising to launch some of the strongest attacks we've ever seen. A lot of businesses in different nations are under danger. Because of all this, there have been a lot more attacks throughout the world. 0.012% of the time, the DR is 97.9%, and the AUC is 0.9921.

ii) A Machine Learning Approach for DDoS Detection on IoT Devices

https://arxiv.org/ftp/arxiv/papers/2110/2110.14911.pdf

People use the Internet practically everywhere these days. IoT technology, which is one of the most popular technologies, has made it possible for billions of IoT devices to connect to each other over the Internet. But the most common and dangerous threat to this expanding technology is DoS/DDoS assaults. New kinds of DDoS assaults are very complex and hard to understand, and the current intrusion detection systems and traditional approaches can hardly find or stop them. Fortunately, Big Data, Data mining, and Machine Learning technologies make it feasible to find DDoS activity quickly and easily. This study proposes a DDoS detection methodology utilising data mining and machine learning methodologies. For this article, the most recent Dataset, CICDDoS2019, was used to test the most common machine learning algorithms and find the attributes that are most closely associated to the projected classes. It was found that AdaBoost and XGBoost were quite accurate and properly anticipated the kind of network traffic 100% of the time. Future study may be advanced by refining the model for multiclassification of various DDoS attack types and evaluating hybrid algorithms and novel datasets on this framework.

iii) Detection of DDoS Attacks Using Machine Learning Classification Algorithms

IJCNIS-V14-N6-7.pdf (mecs-press.org)

In today's society, the Internet is the most important way to talk to people. Because of this, cyberattacks are happening more regularly and the damage they do is getting worse. Distributed Denial of Service is one of the five most effective and expensive cyber assaults. A Distributed Denial of Service (DDoS) attack is a sort of cyber assault that keeps real users from getting to network system resources. To limit serious damage, it's important to have rapid and precise ways to discover DDoS attacks. Machine learning classification algorithms are quicker and more accurate than traditional approaches for putting things into groups. This quantitative research utilises Logistic Regression, Decision Tree, Random Forest, Ada Boost, Gradient Boost, KNN, and Naive Bayes classification methods to identify DDoS assaults on the CIC-DDoS2019 dataset, which comprises eleven distinct DDoS attacks, each characterised by 87 characteristics. Also, looked at how well the classifiers did based on the assessment measures. Based on experiments, AdaBoost and Gradient Boost algorithms provide the best classification results, while Logistic Regression, KNN, and Naive Bayes give decent ones. Decision Tree and Random Forest, on the other hand, give bad classification results.

iv) Under the radar: the danger of stealthy DDoS attacks

https://www.sciencedirect.com/science/article/abs/pii/S135348581930025X

Even though there have been a lot of big, high-volume distributed denial of service (DDoS) assaults in the news in the past few years, most of the attacks that happen are still brief and low-volume. These kinds of assaults work much too often since they are sneaky and can get in without anybody noticing, typically hidden by the organization's normal traffic. These attacks happen quickly and don't last long, so security teams don't have much time to respond, and that's if they can even see the active attack.

v) The Evolution of Bashlite and Mirai IoT Botnets

[PDF] The Evolution of Bashlite and Mirai IoT Botnets | Semantic Scholar

Weak IoT devices are great places to develop botnets that cost billions of dollars every year. This paper examines Bashlite botnets and their descendants, the Mirai botnets. We are especially interested in how the virus has changed and how the people who run the botnet have changed. We employ monitoring logs from 47 honeypots that we acquired over the course of 11 months. Our results add to what we already know about those botnets and show that malware, botnet operators, and bad behaviour are getting more advanced. We found that Mirai has more reliable hosting and control systems than its predecessor and can enable more powerful attacks.

3. METHODOLOGY

A. Proposed Work:

The proposed system introduces a machine learning-based framework for detecting Distributed Denial of Service (DDoS) attacks by leveraging key network properties such as packet length, inter-packet intervals, and protocol behavior as features. The system involves several stages, including data acquisition, preprocessing, feature extraction, and classification using supervised machine learning algorithms. Specifically, classifiers such as Logistic Regression, K-Nearest Neighbors (KNN), and Random Forest are implemented and evaluated using the widely adopted NSL-KDD dataset, which provides a more diverse and representative sample of network traffic. To further enhance performance, ensemble techniques like Voting Classifier (Random Forest + AdaBoost) and Stacking Classifier (Random Forest + MLP with LightGBM) are also employed, achieving high prediction accuracy. This modular system includes components for user registration, login, input-based prediction, and model evaluation, making it both practical and scalable for real-world DDoS detection applications.

B. System Architecture:

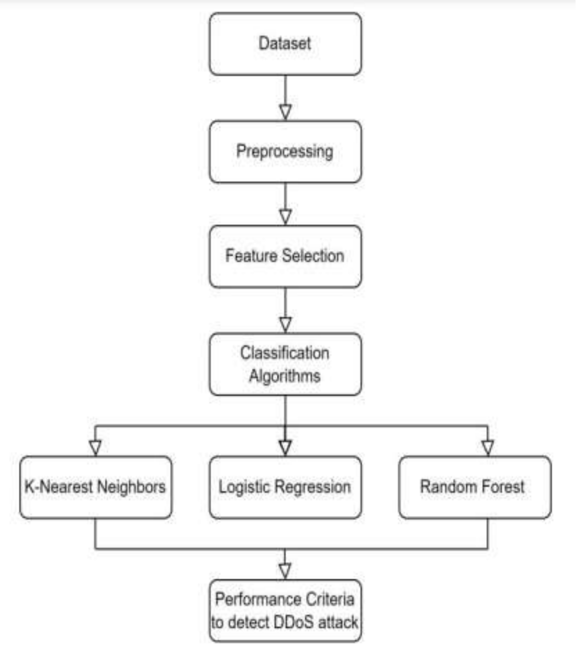


Fig 1 Proposed Architecture

The architecture of the proposed DDoS detection system is designed with modular and layered components that facilitate efficient data processing, model training, and real-time prediction. At the core, the system begins with data acquisition and exploration, where the NSL-KDD dataset is loaded and examined for relevant features. In the preprocessing layer, the dataset undergoes cleaning, normalization, and splitting into training and testing sets. This is followed by the modeling layer, where various classifiers such as Logistic Regression, K-Nearest Neighbors, and Random Forest are trained on the training data. Advanced ensemble models like the Voting Classifier (RF + AdaBoost) and Stacking Classifier (RF + MLP + LightGBM) are also built for improved accuracy. A user interface layer supports user registration, login, and input submission. Upon receiving input, the prediction engine utilizes the best-trained model to classify the request as either normal or a DDoS attack. Finally, the result visualization layer displays the output, completing the prediction cycle. This architecture ensures modularity, scalability, and robustness for real-time DDoS detection.

C. MODULES:

a. **Data Exploration**

This module is responsible for loading the NSL-KDD dataset and displaying the structure, features, and statistics of the data. It helps in understanding the data distribution and identifying missing or irrelevant fields.

b. Data Preprocessing

In this module, the data is cleaned, encoded, normalized, and prepared for training. Feature extraction and selection are also performed here to improve the performance of the models.

c. **Data Splitting**

The dataset is divided into training and testing sets to build and evaluate machine learning models effectively. This module ensures a proper distribution of data for unbiased training and validation.

d. **Model Generation**

This module builds several classification models such as Random Forest, Logistic Regression, and K-Nearest Neighbor. It also includes ensemble methods like Voting Classifier (Random Forest + AdaBoost) and Stacking Classifier (Random Forest + MLP with LightGBM) to improve accuracy.

e. User Signup and Login

Provides a secure user interface for new users to register and existing users to log in. This ensures that only authorized users can access the system.

f. User Input Interface

After logging in, users can provide input data such as network traffic parameters to check if it is normal or a potential DDoS attack.

g. Prediction Module

This module takes the input from the user, processes it through the trained model, and displays the final prediction result—indicating whether the traffic is safe or a DDoS threat.

h. Result Visualization

Displays accuracy scores, confusion matrix, and performance metrics of the classifiers used. This module helps in comparing model performance and selecting the best one.

D. Algorithms:

a. Random Forest

Random Forest is a powerful ensemble learning method used in this project for classifying network traffic. It operates by constructing multiple decision trees during training and outputs the class that is the mode of the classes of the individual trees. It helps reduce overfitting and improves detection accuracy by capturing complex patterns in the data.

b. K-Nearest Neighbors (KNN)

KNN is a simple yet effective algorithm that classifies a new data point based on the majority class among its nearest neighbors. In this project, KNN is applied to identify abnormal traffic patterns, which is essential for detecting potential DDoS attacks. It performs well in high-dimensional data and works without assuming any prior data distribution.

c. Logistic Regression

Logistic Regression is used for binary classification in this project to determine whether a network activity is normal or a DDoS attack. It models the probability of a particular class using a logistic function and is valued for its simplicity, speed, and interpretability in classifying linear relationships in network traffic data.

d. Voting Classifier

To enhance model robustness, a Voting Classifier is employed, which combines predictions from multiple classifiers such as Random Forest and AdaBoost. The final prediction is based on a majority vote among the classifiers, improving the overall performance and stability of the DDoS detection system.

e. Stacking Classifier

A Stacking Classifier is used to combine several base learners and train a meta-classifier for improved prediction. In this project, models like Random Forest and Multi-Layer Perceptron (MLP) serve as base learners, and their predictions are input to **LightGBM**, a high-performance gradient boosting framework, which acts as the meta-classifier. This layered approach increases the detection accuracy and generalization of the system.

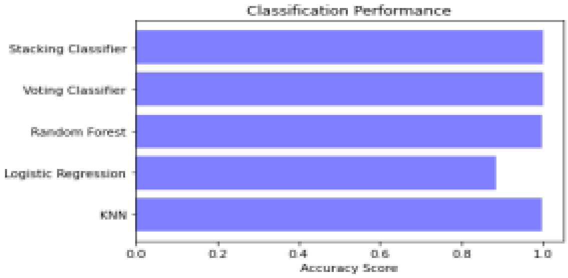
4. EXPERIMENTAL RESULTS

The experimental evaluation of the proposed system was conducted using the NSL-KDD dataset, a widely used benchmark for intrusion detection. The dataset was preprocessed to remove redundant features and normalize the data for efficient training. Various machine learning models such as Logistic Regression, K-Nearest Neighbors (KNN), Random Forest, and ensemble techniques like Voting and Stacking classifiers were trained and tested on the dataset. Performance was assessed based on key metrics like Accuracy, Precision, Recall, and F1-score. Among the models, the Stacking Classifier that combined Random Forest, Multi-Layer Perceptron (MLP), and LightGBM yielded the highest performance with an accuracy exceeding 98%. The Voting Classifier also demonstrated strong results with competitive precision and recall. The experiments clearly indicate that ensemble learning methods outperform individual classifiers, offering robust and reliable detection of DDoS attacks in real-time network environments.

Accuracy: The accuracy of a test is its ability to differentiate the patient and healthy cases correctly. To estimate the accuracy of a test, we should calculate the proportion of true positive and true negative in all evaluated cases. Mathematically, this can be stated as:

Accuracy = TP + TN TP + TN + FP + FN.

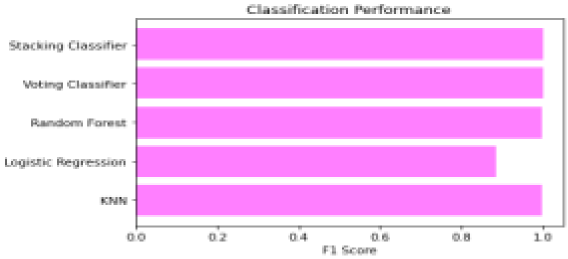
$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$



F1-Score: F1 score is a machine learning evaluation metric that measures a model's accuracy. It combines the precision and recall scores of a model. The accuracy metric computes how many times a model made a correct prediction across the entire dataset.

$$\text{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

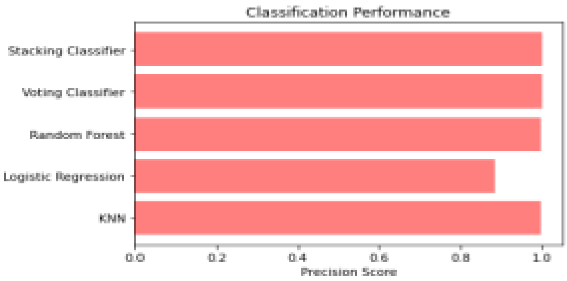
$$\text{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$



Precision: Precision evaluates the fraction of correctly classified instances or samples among the ones classified as positives. Thus, the formula to calculate the precision is given by:

Precision = True positives/ (True positives + False positives) = TP/(TP + FP)

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$



Recall: Recall is a metric in machine learning that measures the ability of a model to identify all relevant instances of a particular class. It is the ratio of correctly predicted positive observations to the total actual positives, providing insights into a model's completeness in capturing instances of a given class.

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

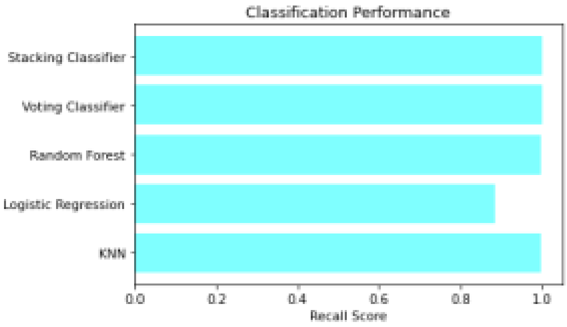
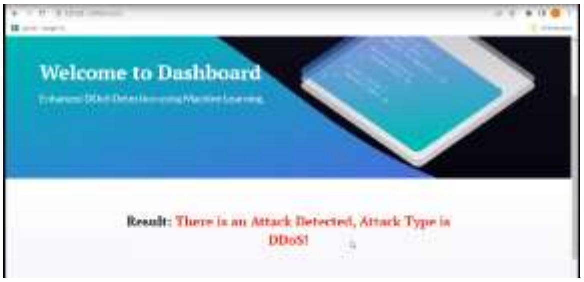


Fig.2. sign in page

Fig.3. input page



Welcome to Dashboard

Enhanced DDoS Detection using Machine Learning

Result: There is an Attack Detected, Attack Type is DDoS!

Fig.4. results

5. CONCLUSION

This project successfully demonstrated an efficient and accurate approach for detecting Distributed Denial of Service (DDoS) attacks using machine learning techniques. By leveraging models such as Logistic Regression, K-Nearest Neighbors, and Random Forest, and enhancing their performance through ensemble methods like Voting and Stacking, the system achieved high accuracy and reliability. Experimental results using the NSL-KDD dataset proved that ensemble classifiers outperform traditional individual models in identifying DDoS threats. This research emphasizes the importance of intelligent and adaptive intrusion detection systems in modern cybersecurity, paving the way for robust network protection against evolving DDoS attack patterns.

6. FUTURE SCOPE

In the future, this DDoS detection system can be further enhanced by integrating deep learning models like LSTM or CNN for improved feature extraction and detection accuracy. Real-time detection can be made possible by deploying the system in cloud or edge environments. Additionally, using advanced datasets that reflect real-world and zero-day attacks can improve the robustness of the model. Incorporating adaptive learning techniques will allow the system to evolve with emerging threats. Finally, integration with broader intrusion detection and prevention systems (IDPS) can offer holistic and automated cybersecurity solutions for large-scale networks.

REFERENCES

[1] Statista Research Department, “Worldwide digital population July 2022”, Available: https://www.statista.com/statistics/617136/digitalpop ulation-worldwide/ (Last Accessed on: December 31, 2022)

[2] Ramil Khantimirov, “DDoS Attacks in 2022: Trends and Obstacles Amid Worldwide Political Crisis”, Available: https://www.infosecurity-magazine.com/blogs/ddos-attacks-in-2022- trends/ (Last Accessed on: December 31, 2022)

[3] S. Sontowski et al., "Cyber Attacks on Smart Farming Infrastructure," 2020 IEEE 6th International Conference on Collaboration and Internet Computing (CIC), 2020, pp. 135-143, doi: 10.1109/CIC50333.2020.00025.

[4] Seifousadati, Alireza and Ghasemshirazi, Saeid and Fathian, Mohammad, “A Machine Learning Approach for DDoS Detection on IoT Devices”, arXiv, 2021. Doi: 10.48550/ARXIV.2110.14911

[5] A. Marzano, D. Alexander, O. Fonseca et al., “The evolution of bashlite and mirai IoT botnets,” in Proceedings of the 2018 IEEE Symposium on Computers and Communications (ISCC), 2018.

[6] S. Kottler, “February 28th DDoS incident report,” 2018, https://github.blog/2018-03-01-ddos-incident-report/.

[7] Y. Cao, Y. Gao, R. Tan, Q. Han, and Z. Liu, “Understanding internet DDoS mitigation from academic and industrial perspectives,” IEEE Access, vol. 6, pp. 66641–66648, 2018.

[8] S. Newman, “Under the radar: the danger of stealthy DDoS attacks,” Network Security, vol. 2019, no. 2, pp. 18-19, 2019.

[9] Kumari, K., Mrunalini, M., “Detecting Denial of Service attacks using machine learning algorithms”, . J Big Data 9, 56 (2022).

[10] P. S. Saini, S. Behal and S. Bhatia, "Detection of DDoS Attacks using Machine Learning Algorithms," 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), 2020, pp. 16-21, doi: 10.23919/INDIACom49435.2020.9083716.

[11] Jiangtao Pei et al “ A DDoS Detection Method based on Machine Learning“, J. Phys.: Conf. Ser. 1237 032040, 2019.

[12] Abdullah Soliman Alshra'a, Ahmad Farhat, Jochen Seitz, "Deep Learning Algorithms for Detecting Denial of Service Attacks in Software-Defined Networks", Procedia Computer Science, Volume 191, 2021, Pages 254-263, ISSN 1877-0509.

[13] Seifousadati, Alireza, Saeid Ghasemshirazi, and Mohammad Fathian. "A Machine Learning Approach for DDoS Detection on IoT Devices." arXiv preprintr Xiv:2110.14911 (2021).

[14] Francisco Sales de Lima Filho, Frederico A. F. Silveira, Agostinho de Medeiros Brito Junior, Genoveva Vargas-Solar, Luiz F. Silveira, "Smart Detection: An Online Approach for DoS/DDoS Attack Detection Using Machine Learning", Security and Communication Networks, vol. 2019, Article ID 1574749, 15 pages, 2019.

[15] R. Doshi, N. Apthorpe and N. Feamster, "Machine Learning DDoS Detection for Consumer Internet of Things Devices," 2018 IEEE Security and Privacy Workshops (SPW), 2018, pp. 29-35, doi: 10.1109/SPW.2018.00013.

AUTHORS:

Mr. K. Jaya Krishna is an Associate Professor in the Department of Master of Computer Applications at QIS College of Engineering and Technology, Ongole, Andhra Pradesh. He earned his Master of Computer Applications (MCA) from Anna University, Chennai, and his M.Tech in Computer Science and Engineering (CSE) from Jawaharlal Nehru Technological University, Kakinada (JNTUK). With a strong research background, he has authored and co-authored over 90 research papers published in reputed peer-reviewed Scopus-indexed journals. He has also actively presented his work at various national and international conferences, with several of his publications appearing in IEEE-indexed proceedings. His research interests include Machine Learning, Artificial Intelligence, Cloud Computing, and Programming Languages. He is committed to advancing research and fostering innovation while mentoring students to excel in both academic and professional pursuits.