International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: x Issue: y
DOI: (do not edit)
Article Received: (do not edit) **Revised**: (do not edit) **Accepted**: (do not edit) **Publication**: (do not edit)

Denial of Service (DoS) Attack Detection Using Machine Learning

1. Prof.Barkha Kumari
Dept.of Artificial Intelligence
GHRCEM,PUNE
Pune, India
barkha.kumari@raisoni.net

2.Onkar Jyotiling Mane
Dept.of Artificial Intelligence
GHRCEM,PUNE
Pune, India
onkar.mane.ai@ghrcem.raisoni.net

3.Rutuja Nandkumar Ohol
Dept.of Artificial Intelligence
GHRCEM,PUNE
Pune, India
rutuja.ohal.ai@ghrcem.raisoni.net

Sanket Vinodkumar Chaudhary
Dept.of Artificial Intelligence
GHRCEM,PUNE
Pune, India
sanket.chaudhary.ai@ghrcem.raisoni.net

Abstract— *Denial of Service (DoS) attacks are often used in security hacking techniques to interfere with geographical networks or make computer resources unusable. Using publicly available datasets, we apply the Gradient Boosting,Support Vector Machine,Naive Bayes,K-Nearest Neighbors to detect and mitigate Denial of Service (DoS) assaults. According to our experimental results, the Gradient Boosting,Support Vector Ma chine performed better in terms of accuracy and balanced accuracy than both the Naive Bayes method and k-nearest neighbors, particularly when working with datasets that had a little imbalanced distribution.*

Keywords- :*Denial of Service, Cybersecurity,Machine Learning, Gradient Boosting, Support Vector Machines (SVM), Naive Bayes, K-Nearest Neighbors (K-NN)*

I. INTRODUCTION

In the current digital age, the security of the internet and its networked systems is of paramount importance. However, as the cyber landscape continues to evolve, one type of attack has remained a major challenge for network security- denial of service attacks (DoS). These malicious activities, which seek to interfere with the normal operation of computer systems and services, are becoming increasingly sophisticated and widespread. As distributed computing and connected devices become more widespread, the attack surface has increased, making traditional methods of detection and prevention less effective. As a result, there is an urgent need for novel approaches that can effectively respond to the ever changing nature of cyber threats

Machine Learning has proven to be a disruptive technology in a variety of areas, and its capabilities in network security are no exception. In recent years, there has been a great deal of attention paid to the potential of machine learning algorithms to identify and mitigate distributed denial of-service (DoS) attacks. By allowing for automated and data-based decision making, ML has the potential to significantly improve the accuracy and response time of DoS detection. This project entitled "DoS attack detection using machine learning," seeks to investigate the use of ML techniques for the detection and remediation of DoS attacks. This report will provide an explanation of the rationale for this project, the objectives of the project, and the methodology and tools to be used.

This report will discuss the importance of this research in network security and the wider digital environment, high lighting the advantages and implications of using Machine Learning (ML) for the detection of DoS attacks. It will also cover the methodology of the project, the data collection and preprocessing methods, the use of ML algorithms, as well as the evaluation metrics used to evaluate the performance of the DoS detection system proposed. Finally, it will cover the expected results, potential issues, and the broader consequences of this research, emphasizing the need to further develop our capabilities in protecting critical network infrastructure from the ongoing DoS attacks.

II. MOTIVATION

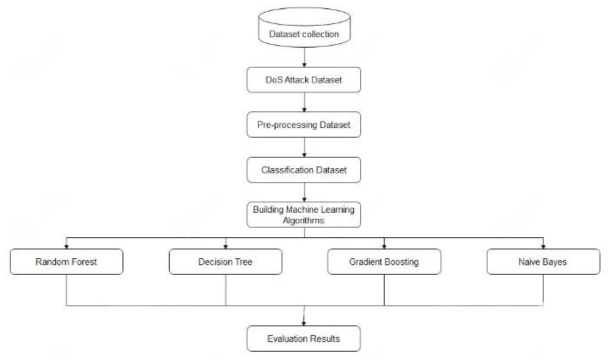
For many years, denial of service (DoS) assaults has been a recurring issue in network security. Denial-of-service (DoS) and traditional DoS attacks still pose a threat even with the advancement of numerous detection and defense mechanisms. DoS attacks are particularly becoming a significant issue going forward because they are happening more often by the day. According to a report, a DoS monitoring system recorded an incredible 57,116 attacks in the third quarter of 2022, the majority of which came from the United States. In 2022, the data comprised 45.95 percent from the second quarter and 39.60 percent from the third quarter. Notably, well-known IT organizations have also been the subject of denial-of-service (DoS) assaults in recent years. The worst incident, which targeted AWS in February 2020, produced an incredible 2.3 terabits per second (Tbps) of traffic. In 2018, GitHub too fell victim to a denial-of-service assault.

DoS attackers employ the transmission control protocol/user datagram protocol (TCP/UDP) to take advantage of partially opened connections. They then use internet protocol (IP) spoofing to transmit multiple forged payloads normally. In contrast, in an application layer denial-of-service assault, at tackers send multiple requests to exhaust popular programs like the domain name system (DNS), HTTP, and so on. At the network level, there is no way to distinguish these requests from those made by authorized users.

The proportion of false requests in these assaults is higher than that of real queries. Since the connections are already established and the requests appear to be coming from authorized users, it is difficult to identify these assaults. This paper suggests a machine learning-based method for detecting denial of-service (DoS) assaults. An intelligent protection system that can identify attacks is required because it is very difficult to manually monitor network traffic in order to keep the system safe from attackers. Because it is a straightforward yet efficient method for getting superior outcomes, the suggested system performs noticeably better than current ones.

III. METHODOLOGY

Python has surpassed all other programming languages in popularity among data scientists thanks to its user-friendliness, simplicity, and quick prototyping. To read and process data reliably and effectively, it offers strong statistical and numerical tools like NumPy and pandas. Additionally, it features a useful machine learning program called scikit-learn. In this work, we'll use the aforementioned programmers to create a model. For the creation of an effective machine learning model, Python provides various libraries which helps to develop application easily. We chose Python as our implementation language in this article to carry out data analytics and machine learning since the large library shows that Python has mature support for data science.



A. Dataset

The Dataset that we used in building the machine learning model, is taken from Kaggle titled ”Dos Attack Detection”. The subject of this study is detecting the DoS attack. The dataset consists of a large number of samples data that makes it suitable to evaluate the detection accuracy. The Overall Data has 3,46,869 instances and 78 columns out of which 159295 belong to benign, 55180 for DoS Slow Loris followed by 132394 for DoS Hulk. We narrowed down the dataset to focus on two specific classes: “DoS Slow Loris” and ”Benign.” Benign represents the normal traffic on the network while DoS slow Loris represents a DoS attack bringing down the total instances to 214475 and 78 columns.

B. Data Collection

In this paper, we used the dataset from the Kaggle website that related very popular and well-known IoT attacks. The DoS attack. we divide the dataset into training set and test set respectively. The training set of the dataset accounts for 75% of each total sample, and the test set accounts for 25% of each total sample. And then we also test 20% (Testing) / 80% (Training) and 30% (Testing) / 70% (Training). The aim is to build a network intrusion detector, a predictive model that can distinguish between “bad” connections (intrusions or attacks) and “good” normal connections.

C. Data Pre-Processing

To make the dataset more readable, accessible, and to not contain any null values, we applied two way. In the first one, we fill missing values in the dataset of the mean value for certain columns. The alternate is that we use the marker encoder system to convert the data types for specific columns to numeric data types because machine literacy algorithms can only deal with numeric data types. Now the dataset is ready to be used as an input to the algorithms. The elimination of unnecessary background information significantly increases the efficiency of parameter identification.

D. Machine Learning Models

This exploratory paper delves into the ever-evolving field of machine learning models, examining their importance, uses, and most recent developments. Our research demonstrates the transformative power of machine learning across a range of algorithms and provides insight into the fundamental concepts of the field. We also shovel into important issues of model interpretability, bias, and ethical considerations, and illuminate the need for responsible AI development. also, we give a comprehensive overview of state- of- the- art ways and arising trends, pressing the ongoing quest to exploit the full eventuality of machine knowledge models in the digital age. our swiftly evolving digital world.

a) Gradient Boosting

The Gradient Boosting is an ensemble algorithm that was developed to solve classification and regression tasks. It merges several weak learners into a single strong learner. These are Gradient Boosting techniques, in which each tree is run separately, producing independent forecasts, which are then combined to make a final model's prediction. The number of weak learners is determined as number of estimators parameter. The model's prediction is integrated in classification problems like detecting the DoS by selecting the class label with the most votes from all trees.

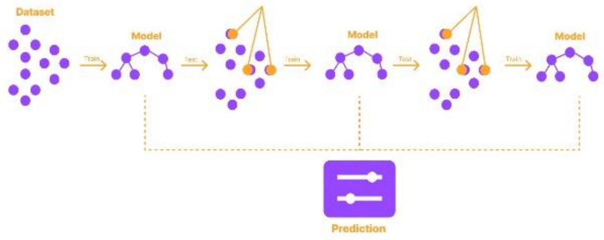
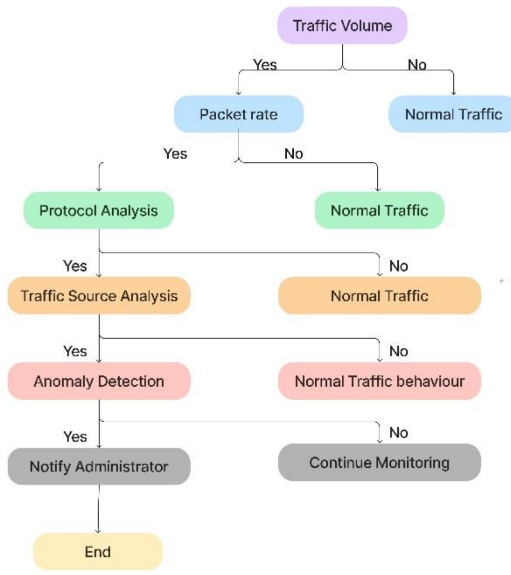


Fig: - Gradient Boosting

b) Decision Tree

A decision tree is a classifier expressed as a recursive partition of the in-stance space. The decision tree consists of nodes that form a rooted tree meaning it is a directed tree with a node called "root" that has no incoming edges. All other nodes have exactly one incoming edge. A node with outgoing edges is called an internal or test node. All other nodes are called leaves (also known as terminal or decision nodes). In a decision tree, each internal node splits the instance space into two or more sub-spaces according to a certain discrete function of the input attributes values.



c) Naive Bayes

Naive Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems. It is mainly used in text classification that includes a high-dimensional training dataset. Naive Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions. It is a probabilistic classifier, which means it predicts on the basis of the probability of an object. Some popular examples of Naive Bayes Algorithm are spam filtration, Sentimental analysis, and classifying articles.

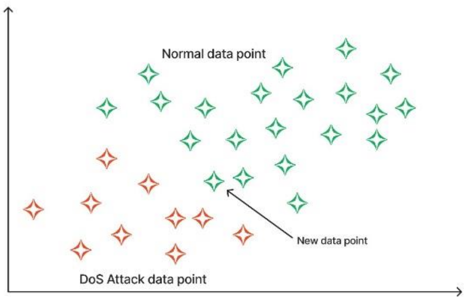
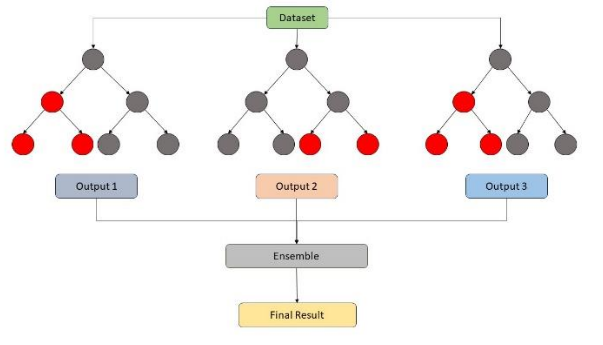


Fig: - Naïve Bayes

d) Random Forest

Random forest classifiers fall under the category of ensemble-based learning methods. They are known for being simple to implement, operationally fast, and successful across various domains. The fundamental principle of random forests involves constructing multiple "simple" decision trees during the training stage and making a majority vote (mode) across them during the classification stage. This approach helps mitigate the overfitting tendency of individual decision trees. In the training stage, random forests use bagging, a technique that involves repeatedly selecting random samples with replacement from the training set and fitting trees to these samples. Each tree is grown without pruning. The number of trees in the ensemble is a parameter learned automatically using the out-of-bag error. Random forests share simplicity and good performance with algorithms like naïve Bayes and k-nearest neighbors.



IV. Result

The Denial of Service (DoS) Attack Detection system achieved a high attack recognition rate of 87 percent, pro cessed data faster than the manual system, and received positive user feedback. The system's error rate was low and most errors were due to insufficient data. Compared to existing systems, the system offered advantages in accuracy and speed, and further improvements. These findings demonstrate the potential of an Denial of service (DoS) attack detection system to improve the cyber space with the help of new age machine learning techniques for the organizations and individuals. Abbreviations and Acronyms

V. Conclusion

Security threats are evolving and getting more hidden and complicated. Detecting malicious security threats and attacks have become a huge burden to cyberspace. We should apply proactive prevention and early detections of security vulnerabilities and threats rather than patching security holes afterwards. To analyze large amount of data to find out suspicious behaviors, threat patterns, and vulnerabilities and to predict and prevent future cybersecurity threats are a challenge. Machine Learning (ML) is a powerful instrument to take up such challenge.

In this paper, we used dataset to classify denial of ser vice attack by using Gradient Boosting, Support Vector Machine, Naïve Bayes, K-Nearest Neighbors and compare their performance. The experimental results show that the Gradient Boosting, Support Vector Machine performed better than Naive Bayes and K-Nearest Neighbors algorithm in the dataset with slightly imbalanced distribution.

Acknowledgment

I would like to convey my heartfelt gratitude to Mr. Rishabh Singh for his tremendous support and assistance in the completion of our project. I would also like to thank our Project Mentor and Guide Miss.Barkha Kumari for providing me with this wonderful opportunity to work on a project Denial of Service (DoS) Attack Detection Using Machine Learning. The completion of the project would not have been possible without their help and insights.

Future Work

Future work for this report can focus on the implementation and testing of the proposed system for Denial of service (Dos) attack detection and machine learning techniques. This system can be tested on a larger scale and in different organizations and institutions to determine its effectiveness and potential for implementation on a wider scale.

Another area for future work is the improvement of the accuracy and efficiency of the proposed system. This can be achieved through the use of advanced machine learning algorithms and the incorporation of feedback mechanisms to continually improve the system's performance.

In addition, research can be conducted on the potential for integrating blockchain technology into the proposed system to further enhance its security and reduce the risk of cyber-attacks.

Lastly, the proposed system can be extended to other type cyber-attacks, such as Man in the Middle, DDos to provide a more comprehensive solution for organizations. This can further streamline their workflows and reduce the need for manual intervention.

References

[1] Y. Mirsky, N. Kalbo, Y. Elovici, and A. Shabtai, "Vesper: Using echo analysis to detect man-in-the-middle attacks in LANs," IEEE Transactions on Information Forensics and Security, vol. 14, no. 6, pp. 1638–1653, Jun. 2019, doi: 10.1109/TIFS.2018.2883177. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[2] P. S. Saini, S. Behal, and S. Bhatia, "Detection of DDoS attacks using Machine learning algorithms," in 2020 7th International Conference on Computing for Sustainable Global Development (INDIACom), Mar. 2020, pp. 16–21, doi: 10.23919/INDIACom49435.2020.9083716.

[3] Chunhui Bao, Yifei Pu, and Yi Zhang,"Fractional-Order Deep Backprop agation Neural Network," Computational Intelligence and Neuroscience, Vol. 2018, Article ID 7361628, 2018.

[4] V. Christina, S. Karpagavalli, G. Suganya, “Email spam filtering using supervised machine learning techniques,” Int. J. Comput. Sci. Eng., 02 (09) (2010), pp. 3126-3129.

[5] S. Banerjee and P. S. Chakraborty, “Proposed approach to de tect distributed denial of service attacks in software defined net work using machine learning algorithms,” International Journal of Engineering Technology, vol. 7, no. 2.8, p. 472, Mar. 2018, doi:10.14419/ijet.v7i2.8.10488.

[6] Jaeyeon Jung, Balachander Krishnamurthy, Michael Rabinovich, “Flash Crowds and Denial of Service Attacks: Characterization and Impli cations for CDNs and Web Sites,” Proc. of World Wide Web, May 2002,Honolulu, USA, pp. 293-304.

[7] S. Balaji and R. Seshadri, “Attack prevention and attack detection strategies by comparing different DDos models,” International Journal of Computer Applications, vol. 129, no. 14, pp. 24–27, Nov. 2015, doi: 10.5120/ijca2015907094.

[8] J. Cheng, J. Yin, Y. Liu, Z. Cai, and C. Wu, “DDoS Attack Detection Using IP Address Feature Interaction,” Proc. of 2009 International Conference on Intelligent Networking and Collaborative Systems, pp. 113-118.