IJARSCT

ISSN (Online) 2581-9429

International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)

International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal

Impact Factor: 7.301

Volume 3, Issue 11, May 2023

Detection and Prevention of DDoS Attack using Ensemble Model

Prof. R. N. Muneshwar[1], Mr. Mahale Swami[2], Mr. Muley Pranav[3],
Mr. Joshi Lakshmikant[4], Mr. Sambyal Suryansh[5]
Department of Information Technology[1,2,3,4,5]
Amrutvahini College of Engineering, Maharashtra, India

Abstract: *In the current cyber world, one of the most severe cyber threats are distributed denial of service (DDoS) attacks, which make websites and other online resources unavailable to legiti- mate clients. It's a cyberattack aimed at overwhelming a server with malicious traffic, causing a website to shut down temporarily or permanently. It's typically executed using malware- infected devices called bots, and their cluster is referred to as a botnet. These bots include lap- tops, smartphones, smart TVs, wearable devices, thermometers, security cameras, in-vehicle infotainment systems, etc. So, what industries do DDoS attackers target? They commonly target the gaming, software and technology, media and entertainment, finance, and internet and telecom industries. It is different from other cyber threats that breach security parameters; how- ever, DDoS is a short-term attack that brings down the server temporarily. So we came up with a solution to decrease the impact of DDoS attack and thus proposed a model which can predict whether input given is a attack file or normal file based on the dataset (CICDDOS2019) it is trained on.To mitigate the impact of DDoS attacks, we propose a model that predicts whether an input file is an attack file or a normal file based on the CICDDOS2019 dataset it is trained on. We have developed an ensemble of machine learning classifiers, including KNN-DT, KNN-RF and DT-RF, to enhance the accuracy and robustness of the prediction model. By accurately identifying attack files, organizations can take proactive measures to protect their servers and mitigate the effects of DDoS attacks.*

Keywords: Distributed Denial of Service (DDoS), Deep Learning, CNN, KNN, Decision Tree, Random Forest etc.

I. INTRODUCTION

Distributed denial of service (DDoS) attacks are currently the most prevalent and sophisticated threat for organizations, and are increasingly difficult to prevent. In 2018, for example, GitHub was hit with one of the largest DDoS attacks ever [4]. This impactful attack comes in one of the most highlighted cyberattacks of the current cyber age, shaking the ground basis of one of the pillars (availability) of the CIA security triad. Attackers use thousands of dump terminals, machines, and botnets to concurrently launch DDoS attacks that subsequently exhaust the target system main resources, making the entire services unavailable. There are a potentially extreme number of legitimate and powerful tools available, which can be abused to launch DDoS attacks on large and small scales accordingly. In another recent DDoS attack [4], attackers misused the legitimate Memcached tool, whose primary purpose is to reduce strain over the underlying VOLUME 4, 2016 1 network resources. The attacker abused Memcached objects and spoofed IP addresses, allowing Memcached responses to be directed to the targetaddresses with 126.9 million packets/second to largely consume target resources. Moreover, the use of spoofed IPs makes the trace-back next to impossible [5] in DDoS attacks. Therefore, the efficient and early detection, mitigation, and prevention of DDoS attacks remain a challenging task. However, strong novel measures can be taken towards timely detection, to allow subsequent countermeasures to prevent or mitigate sophisticated DDoS attacks [6]. There have been interest in utilizing artificial learning approaches (e.g., machine learning and deep learning techniques) to prevent or mitigate sophisticated DDoS attacks [7]–[9], although designing efficient and effective DDoS mitigation strategies remain an ongoing challenge. In this work, a deep CNN framework is proposed for efficient and early detection of DDoS attacks in SDNs, and a deep CNN ensemble mechanism is designed to detect varied Flow based DDoS attacks. In comparison to this solution, related state-of-the-

DOI: 10.48175/IJARSCT-10596

art deep learning (DL) based ensembles and hybrid approaches (i.e., RNN, LSTM, RL) are applied to the same tasks for verification. Table 1 provides a snapshot of the given ensembles and hybrid approaches leveraged in this work.To further enhance the accuracy and robustness of our detection model, we have built an ensemble of machine learning classifiers. This ensemble consists of KNN-DT, KNN-RF, and DT-RF models. The KNN-DT model combines the K-Nearest Neighbors (KNN) algorithm with a Decision Tree (DT) classifier, while the KNN-RF model combines KNN with a Random Forest (RF) classifier. Similarly, the DT-RF model combines a Decision Tree with Random Forest. The ensemble approach leverages the strengths of these individual classifiers to improve the overall detection performance and handle various types of DDoS attacks effectively.To evaluate the effectiveness of our proposed solution, we compare it against state-of-the-art deep learning-based ensembles and hybrid approaches, including recurrent neural networks (RNN), long short-term memory (LSTM), and reinforcement learning (RL). These techniques are commonly applied for DDoS attack detection and mitigation. Table 1 provides an overview of the ensembles and hybrid approaches utilized in this work, showcasing the diverse range of techniques employed to address the inherent challenges of DDoS attacks. By harnessing the power of advanced machine learning and deep learning techniques, along with our ensemble of classifiers, we aim to significantly enhance the detection and prevention capabilities against sophisticated DDoS attacks. Timely detection and subsequent countermeasures are crucial in reducing the impact of these attacks and ensuring the stability and availability of online services. In the following sections, we present the methodology, experimental setup, and results of our proposed deep CNN framework and ensemble mechanism for DDoS attack detection in SDNs. These findings contribute to the ongoing efforts to develop robust and proactive strategies against DDoS attacks in the ever-evolving cyber landscape.

II. LITRETAURE SURVEY

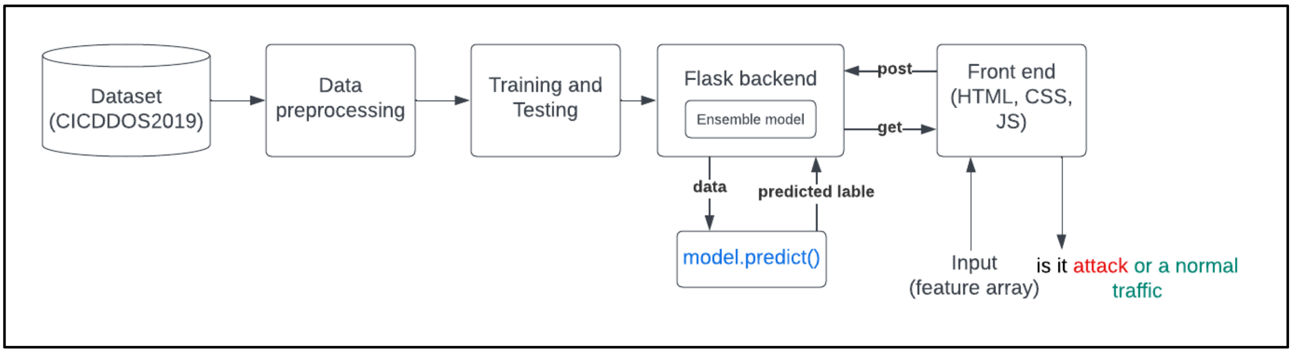
Juan Fernandoand Gabriel Enrique Taborda Blandon, A Deep Learning- Based Intrusion Detection and Preventation System for Detecting and Preventing Denial-of-Service Attacks: The contributions that have been obtained in this research are summarized below. Among ML and DL neural network techniques, Deep Learning is the most suitable for detecting DoS attacks, according to the SLR. According to the criteria defined in this study, the DL algorithms of Deep Feed Forward, Recurrent Neural Network and Long Short Term Memory are the ones that report the best behavior to train models to detect DoS attacks. In this research, the DFNN algorithm was chosen to develop the training model, which, when trained with the CICDDOS2019 data set (which was adapted to recognize two classes: malicious and benign), yielded an accuracy of 0.9994, an accuracy of 0.9995, a recall of 0.999, and an F1 score of 0.9993.

Dnan Akhunzada, Iqra Mustafa, Tanil Bharat Patel, A Deep CNN Ensemble Framework for Efficient DDoS Attack Detection in Software Defined Networks: Contemporary innovative research and novel cyber security solutions are indispensable to properly secure the new era of digitization. This paper proposed an efficient and scalable deep CNN ensemble framework to address the issue of the most prevalent and sophisticated DDoS attack detection in SDNs. also evaluated proposed framework with benchmark deep learning ensembles and hybrid state-of-the-art algorithms on a flowbased SDN dataset. The proposed algorithm demonstrates improvements both in detection accuracy and computational complexity. Finally, We endorse varied deep learning ensemble based detection and prevention mechanisms for the emerging large-scale distributed networks.

Ahmed Ramzy Shabaan, Mohmed Hussein, DDoS attack detection and classification via Convolutional Neural Network (CNN): Distributed Denial of Service is considered one of the most serious and widespread threats that faced by those responsible for securing networks. In this paper, five different classification algorithms were proposed and implemented to detect and classify DDoS attack. All models are built and trained by using two different datasets. Convolutional Neural Network (CNN) is commonly used in image processing and classification field. CNN is introduced to classify normal traffic from DDoS attack. According to analysis and results, we found that CNN performed better than other classifiers with accuracy of 99

Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, and C. Wang, Machine Learning and Deep Learning Methods for Cybersecurity: This paper presents a literature review of ML and DL methods for network security. The paper, which has mostly focused on the last three years, introduces the latest applications of ML and DL in the field of intrusion detection. Unfortunately, the most effective method of intrusion detection has not yet been established. Each approach to implementing an intrusion detection system has its own advantages and disadvantages, a point apparent from the discussion of comparisons among the various methods. Thus, it is difficult to choose a particular method to implement an intrusion detection system over the others. Datasets for network intrusion detection are very important for training and testing systems. The ML and DL methods do not work without representative data, and obtaining such a dataset is difficult and time-consuming. However, there are many problems with the existing public dataset, such as uneven data, outdated content and the like. These problems have largely limited the development of research in this area. Network information update very fast, which brings to the DL and ML model training and use with difficulty, model needs to be retrained long-term and quickly. So incremental learning and lifelong learning will be the focus in the study of this field in the future.

III. METHODOLOGY



Step 1: The data is collected from the database, cleaned and pre-processed, including removing missing data, label grouping, and data normalization using min-max.
Step 2: After data pre-processing, the data is split into training and testing.
Step 3: The ML model is built and evaluated using recall and AUC; when the ML model reaches the optimal performance.
Step 4: After that DDoS attack feature is given as input to our model which is kept at the flask back end.
Step 5: Based on the dataset it makes the decision whether given input is DDoS attack or a normal traffic.

IV. PROPOSED WORK

1.**Decision Tree**: Decision Tree is a powerful algorithm used for solving classification and regression problems. It is a tree-structured classifier where the internal nodes represent the features of a dataset, and the branches represent the decisions. It can handle both categorical and continuous variables, making it efficient for real-world problems. Decision Tree is popular because of its visualization capability as it can be easily understood, and the logic behind the decisions made by the model can be easily explained. This algorithm is useful in various applications such as finance, healthcare, marketing, and many more. A Decision tree consists of two nodes, Decision and Leaf nodes. Decision nodes contain multiple branches and are used to make decisions through testing. Leaf nodes do not contain further branches but represent the output of those decisions. The tree structure is formed by connecting these nodes using edges. The decision tree is a commonly used method for decision-making and classification in various fields, including machine learning, finance, and medicine. It is easy to interpret and visualize, making it a popular choice for data analysis.

2.**Linear Regression:** Using ML algorithm linear regression and SVM are applied on a group different of clusters. Linear regression separate the data point from a single line. We get accuracy improved of the data. It is used to estimate real values(cost of houses, number of calls, total sales etc) based on continuous new variable(s). Here, we build up connection among free and ward factors by the fitting a best line. This best fit line is known as relapse line and spoken to by a direct condition and represented by a linear equation;

$$Y = a * X + b$$

These coefficients a and b are determined dependent on limiting the aggregate of squared contrast of separation between information focuses and relapse line.

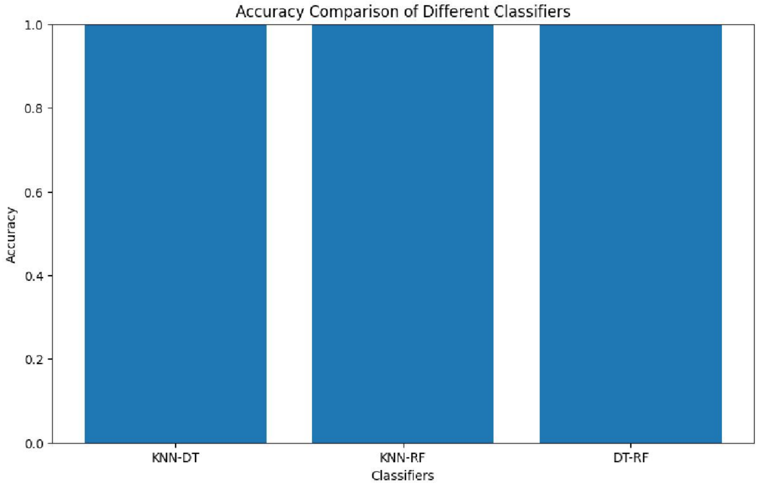
DOI: 10.48175/IJARSCT-10596

3.**Random Forest:** The random forest algorithm is a powerful tool in machine learning that uses ensemble learning to solve complex regression and classification problems. It consists of many decision trees, each of which makes a prediction based on a different subset of input variables. This approach helps to reduce overfitting and increases the accuracy of the results. Random forests have proven to be effective in many applications, including image and speech recognition, fraud detection, and medical diagnosis. They are widely used in industry and academia, and continue to be an active area of research and development.

4.**KNN:** K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique. K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories. K-NN algorithm stores all the available data and classifies a new data point based on the similarity. This means when new data appears then it can be easily classified into a well suite category by using K- NN algorithm. K-NN algorithm can be used for Regression as well as for Classification but mostly it is used for the Classification problems. K-NN is a non-parametric algorithm, which means it does not make any assumption on underlying data. It is also called a lazy learner algorithm because it does not learn from the training set immediately instead it stores the dataset and at the time of classification, it performs an action on the dataset. KNN algorithm at the training phase just stores the dataset and when it gets new data, then it classifies that data into a category that is much similar to the new data. Example: Suppose, we have an image of a creature that looks similar to cat and dog, but we want to know either it is a cat or dog. So for this identification, we can use the KNN algorithm, as it works on a similarity measure. Our KNN model will find the similar features of the new data set to the cats and dogs images and based on the most similar features it will put it in either cat or dog category

V. EXPERIMENTAL RESULTS

Classifier ensemble models trained were compared to see which performed better on datasets. On evaluation metrics, some classifier ensemble models performed well, while others performed poorly. The tree-based ensemble models most improved, since those models contain more hyperparameters to be optimized with a more extensive search space area, which significantly affects the classifier's performance. In contrast, the regression-based ensemble models didn't improve much since those models depend more on data distribution.



VI. CONCLUSION

Hence, we built a robust and effective DDoS attack detection system by implementing a machine learning ensemble model. Our system combines various classifiers, including KNearest Neighbors (KNN), Decision Trees (DT), Random Forest (RF), and others, to leverage their individual strengths and achieve a high level of accuracy in identifying and mitigating DDoS attacks.The construction of our ensemble model involved careful selection and integration of the different classifiers. Each classifier was trained on relevant features extracted from network traffic data, enabling them to capture different aspects of attack patterns and behaviors. By combining the predictions of multiple classifiers, we aimed to enhance the overall detection capabilities of the system. Through extensive experimentation and evaluation,

DOI: 10.48175/IJARSCT-10596

we achieved an impressive accuracy rate of almost 99% with our ensemble model. This accuracy rate indicates the system's ability to accurately differentiate between legitimate network traffic and malicious DDoS attacks, minimizing the occurrence of false positives and false negatives. Our built DDoS attack detection system provides network administrators with a proactive defense mechanism, allowing them to swiftly respond to and mitigate potential attacks

REFERENCES

[1] A. Abdellatif, H. Abdellatef, J. Kanesan, C. -O. Chow, J. H. Chuah and H. M. Gheni, "An Effective Heart Disease Detection and Severity Level Classification Model Using Machine Learning and Hyperparameter Optimization Methods," in IEEE Access, vol. 10, pp. 79974-79985, 2022, doi: 10.1109/ACCESS.2022.3191669

[2] Juan Fernando and Gabriel Enrique Taborda Blandon,"A Deep Learning- Based Intru- sion Detection and Preventation System for Detecting and Preventing Denial-of-Service Attacks," IEEE Access, 2022

[3] Adnan Akhunzada, Iqra Mustafa, Tanil Bharat Patel, "A Deep CNN Ensemble Framework for Efficient DDoS Attack Detection in Software Defined Network," IEEE Access, 2020

[4] Y. Xin, L. Kong, Z. Liu, Y. Chen, Y. Li, H. Zhu, M. Gao, H. Hou, and C. Wang, "Ma- chine learning and deep learning methods for cybersecurity," IEEE Access, vol. 6, pp. 35365–35381, 2018

[5] Ahmed Ramzy Shabaan, Mohmed Hussein, "DDoS attack detection and classification via Convolutional Neural Network (CNN)," IEEE Access, 2019

DOI: 10.48175/IJARSCT-10596