ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 2(03), March 2023, pp. 1-9

Article Title: A DDoS Attack Categorization and Prediction Method Based On Machine Learning

A DDoS Attack Categorization and Prediction Method Based On Machine Learning

S. Ragul[1], S. Tamilselvi[2*]

[1] Assistant professor,Department of EEE,Chettinad college of Engineering and technology,Karur, ragulsambath22@gmail.com.

[2] Associate professor,Department of EEE, SSN College of Engineering, Kalavakkam, tamilselvis@ssn.edu.in.

ABSTRACT

Online services are the basis of the current, digital world. Operations using distributed denial of service (DDoS) remain a significant threat to the availability of online services. DDoS attacks have become increasingly dangerous with the rapid development of computers and communications technology. At present, there has not yet been a detection method with satisfactory detection accuracy due to the diversity of DDoS attack modes and the large volume of traffic involved in DDoS attacks. Therefore, this paper proposes a machine-learning-based method for detecting DDoS attacks. The proposed technique created a confusion matrix to identify the model performance after applying the machine learning models. The first classification uses the KNN classifier technique for both Precision (PR) and Recall (RE) then the Precision (PR) and Recall are both classified using the DNN classifier technique in the second classification (RE). According to the experiment results, the suggested machine learning-based DDoS attack detection system has a high detection rate for the current DDoS attack which is most prevalent and the results are implemented by using python software.

Keywords: Distributed denial of service, KNN classifier, precision and recall, DNN classifier, Python.

1 Introduction

Currently, Internet services are expanding rapidly, and the Internet is becoming inseparable from all aspects of modern life due to the rapid development of the Internet [1]. Internet services are now crucial for both people and commercial enterprises. As the demand for network-based services has increased, attackers have ramped up their attacks on these services to prevent companies from responding to legitimate users [2]. The enormous rise in network connectivity and vulnerability exposure resulted in the development of Software Defined Networks (SDN). Network management and performance typically increase as a result of SDN because it is software-based applications. The SDN is vulnerable to security risks and attack threats because of its centralized architecture .The Distributed Denial of Service (DDoS) attack is one frequent attack [3-4]. In a DDoS attack, different requests are sent to the target online assets using IP spoofing in order to overload the site's capacity to handle multiple requests simultaneously and prevent it from functioning effectively and efficiently even for the

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 2(03), March 2023, pp. 1-9

network's legitimate users [5]. Various DDoS attacks, including smurf attacks, HTTP, POST and SQL injection DoS, are active in networks right now. By sending an enormous amount of useless traffic to the web server, they all cause a denial of service to network services [6]. The controller makes use of the flexible and multifaceted characteristics of SDN network design to build a DDoS attack detection system. The controller extracts network traffic characteristics from statistical flow table data and then employs a support vector machine (SVM) approach to detect attack traffic [7]. A network of compromised connected devices used for cyber-attacks and virtually managed by a Command and Control Center is known as a botnet (C&C). Common instances include personal computers, cell phones, unsecured IoT devices and even resources from public cloud services. Attackers' compromise a device using malware and other methods, making it into a "zombie" in the attacker's botnet [8-9]. The authors suggested alternative classification algorithms because the existing algorithms have numerous defects [10]. Prediction depends heavily on machine learning. Machine learning-based DDoS assault detection has also advanced to a certain extent. The primary machine learning algorithms used for DDoS attack detection, which are the support vector machine, hidden Markov model as well as naive Bayesian algorithm [11-12]. To further improve accuracy and effectiveness, a method which blends model optimization with a number of machine learning classifiers will be used. Utilizing machine learning data mining methods is also essential to improving the data's quality [13-14]. When trained on mixed DDoS-attack traffic across blockchain network layers, machine learning approaches are unable to effectively identify the distinctive core properties. [15].

In this paper, the proposed a thorough simulation of a machine learning technique to identify and forecast the various DDoS attacks, which is built around the pre-processing, data analysis and classification, and performance matrix modules. The KNN classifier and DNN classifier technique is utilized in this study. As a result, the suggested machine learning-based DDoS attack detection system has a high detection rate for the current DDoS attack which is most prevalent and the results are implemented by using python software.

2 Proposed Methodology

The proposed study employs machine learning to categories and forecast various DDoS attack types. An analysis of the input dataset takes place during the first stage of data pre-processing. Standard Scaler is used to pre-process the datasets and Data pre-processing procedures are used to manage with the unnecessary data in the dataset. The KNN and DNN algorithms are used in the machine learning classifier approach. The first classification employs the KNN classifier technique for both Precision (PR) and Recall (RE).Precision (PR) and Recall are both classified using the DNN classifier technique in the second classification (RE).The data are then assessed to produce a forecast result.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 2(03), March 2023, pp. 1-9

Article Title: A DDoS Attack Categorization and Prediction Method Based On Machine Learning

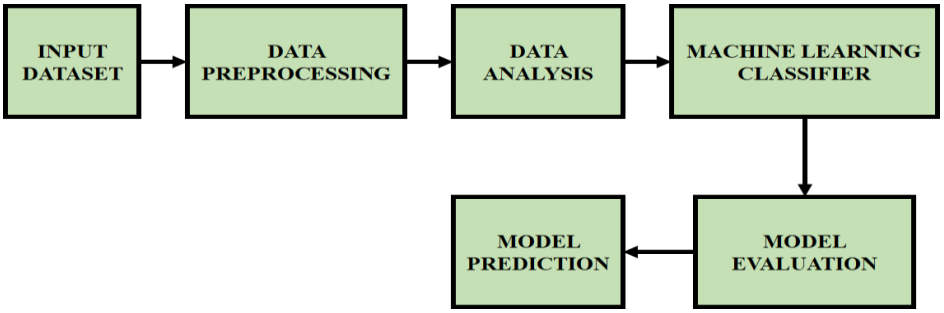

Figure1: *Block diagram for the proposed work*

2.1 Input Dataset

One must consider the dataset being utilized to guarantee the precision of intrusion detection systems. Secure network resilience is required due to the exponential growth of modern networks and applications. By selecting the appropriate training and testing datasets, it's possible to get accomplished.

2.2 Data Preprocessing

Data preprocessing, which is a component of data preparation, refers to any type of processing conducted on raw data in order to make it suitable for another data processing action. Generally, it has been a key initial phase of the information mining process. Because it is obvious that possessing better statistics is more important than having strong models, since the quality of the data is extremely crucial and because data preparation is required. Preprocessing helps to make the data consistent by eliminating any duplicates or anomalies, standardizing the data for comparison and improving the accuracy of the results. Machine learning data preparation methods can be broadly divided into two categories:

- Data cleaning
- Data transformation

2.2.1 Data cleaning

Data cleaning is the process of removing erroneous, damaged, badly structured, duplicate, or incomplete data from a dataset. Combining data from many sources will lead to data duplication or inaccurate categorization in a wide range of situations.

2.2.2 Data transformation

Without changing the substance of the datasets, data transformation is the technological process of transforming data from one format, standard, or structure to another. This is often carried out to make the data more usable by apps or users or to improve the quality of the data.

2.3 Data Analysis

Although there are many individuals, organizations, and experts who have different approaches to data analysis, most of them have been distilled into a single, overarching idea. Data analysis is the process of changing, cleaning and processing raw data to provide useful, applicable data

that allows companies to make informed decisions. The technique helps to reduce the risks involved in decision-making by providing useful data and information, typically displayed in charts, graphics, tables, and graphs.

2.4 Logistic Regression (LR) Classifiers

Logistic regression is a typical example of supervised learning. Logistic regression is a simple and effective classification technique. It is used to determine or estimate the probability that a binary (yes/no) event will take place. Using the logistic regression classification method to calculate the precision (PR), F1 score, and recall values (RE). Logistic regression is used widely in the social sciences, the majority of the medical fields and machine learning. Also, logistic regression will over fit, particularly if the model has a lot of predictor variables. Correction is frequently used to minimize excessive coefficients in the parameters when a model has high dimensionality.

2.5 KNN Classifiers

One of the most basic types of machine learning algorithms, KNN is mostly used for classification. The classification is based on the way a surrounding data point is classified. KNN classify the new updated data points depending on similar to the previously stored data points. The KNN parameter 'k' specifies which of the closest neighbours should be included in the voting process. KNN determines the distances between a query and each example in the data, determines the K examples closest to the query and then votes for the label with the highest frequency.

2.5.1 K-data mining concept

The term "KNN" refers to classifications which use K's most recent historical records to combine to identify new records. For the past 40 years, pattern recognition has extensively researched a well-known statistical technique called KNN. The test sample X is classified as the category which appears the most frequently among the recent K training samples by the KNN as it develops from the test sample X, steadily and increasing the area until it contains K training samples.

2.5.2 Mathematical model for KNN algorithm

Based on the assumption of objects with comparable neighbours will have similar prediction values, the nearest neighbour method of prediction is used. Finding the k points closest to the unknown sample in the multidimensional space Rn is the basic principle of the nearest neighbour algorithm, and the class of the unknown sample is determined using the categories of the k points. These k points are the unknown samples' k closest neighbours. All instances are taken to be points in dimensional space by the method. According to the usual Euclidean distance, the closest neighbour of an instance is determined. The eigenvector of x should be:

$$<a_1(X), a_2(X), \dots, a_n(X) > \quad (1)$$

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 2(03), March 2023, pp. 1-9

Article Title: A DDoS Attack Categorization and Prediction Method Based On Machine Learning

Where,

a_r(x) Represents the rth attribute value of instance x.

The distance between the two instances X_i and X_j is defined as$d(X_i, Y_j)$,

$$d(X_i, Y_j) = \sqrt{\sum_{r=1}^{n}(ar(Xi) - ar(xj))2} \quad (2)$$

2.6 Model Evaluation

This evaluation technique helps to determine the algorithm which will produce the desired results for the supplied dataset. Similarly, it is referred to as "Best Fit" in terms of machine learning. Using the same input dataset, it compares the performance of various machine learning models. The technique of evaluation focuses heavily on the way the model predicts the final results. Out of all the various algorithms that employ in the stage, they select the one which provides greater accuracy for the input data and it is regarded as the best model because it is more accurately forecasts the result.

2.7 Model Prediction

Many data models have been developed to satisfy the needs because data analysis varies from industry to firm according to their requirements. The division of data analytics known as predictive modelling employs probability and data mining to forecast outcomes. The amount of predictors which are extremely helpful in predicting future decisions is used to build each model. After receiving the data for a certain prediction, an analytical model is created.

3 Results and Discussion

The proposed technique **is** thorough simulation of a machine learning technique to identify and forecast the various DDoS attacks, which is built around the pre-processing, data analysis and classification, and performance matrix modules. And the results are done by the python software as shown in the following figures. As a result the suggested machine learning-based DDoS attack detection system has a high detection rate for the current DDoS attack which is most prevalent.

The SDN dataset contains a variety of DoS attack methods which initiated at various OSI model layers. A number of DDoS attack scenarios, including TCP-SYN flood, UDP flood, and ICMP flood attacks, are also included in the SDN dataset. The input dataset is shown in figure 2.

	dt	switch	src	dst	pktcount	bytecount	dur	dur_nsec	tot_dur	flows
0	11425	1	10.0.0.1	10.0.0.8	45304	48294064	100	716000000	1.010000e+11	3
1	11605	1	10.0.0.1	10.0.0.8	126395	134737070	280	734000000	2.810000e+11	2
2	11425	1	10.0.0.2	10.0.0.8	90333	96294978	200	744000000	2.010000e+11	3
3	11425	1	10.0.0.2	10.0.0.8	90333	96294978	200	744000000	2.010000e+11	3
4	11425	1	10.0.0.2	10.0.0.8	90333	96294978	200	744000000	2.010000e+11	3
5	11425	1	10.0.0.2	10.0.0.8	90333	96294978	200	744000000	2.010000e+11	3
6	11425	1	10.0.0.1	10.0.0.8	45304	48294064	100	716000000	1.010000e+11	3
7	11425	1	10.0.0.1	10.0.0.8	45304	48294064	100	716000000	1.010000e+11	3
8	11425	1	10.0.0.1	10.0.0.8	45304	48294064	100	716000000	1.010000e+11	3
9	11425	1	10.0.0.2	10.0.0.8	90333	96294978	200	744000000	2.010000e+11	3

pktrate	Pairflow	Protocol	port_no	tx_bytes	rx_bytes	tx_kbps	rx_kbps	tot_kbps	label
451	0	UDP	3	143928631	3917	0	0.0	0.0	0
451	0	UDP	4	3842	3520	0	0.0	0.0	0
451	0	UDP	1	3795	1242	0	0.0	0.0	0
451	0	UDP	2	3688	1492	0	0.0	0.0	0
451	0	UDP	3	3413	3665	0	0.0	0.0	0
451	0	UDP	1	3795	1402	0	0.0	0.0	0
451	0	UDP	4	3665	3413	0	0.0	0.0	0
451	0	UDP	1	3775	1492	0	0.0	0.0	0
451	0	UDP	2	3845	1402	0	0.0	0.0	0
451	0	UDP	4	354583059	4295	16578	0.0	16578.0	0

Figure 2: *Input dataset*

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 2(03), March 2023, pp. 1-9

Article Title: A DDoS Attack Categorization and Prediction Method Based On Machine Learning

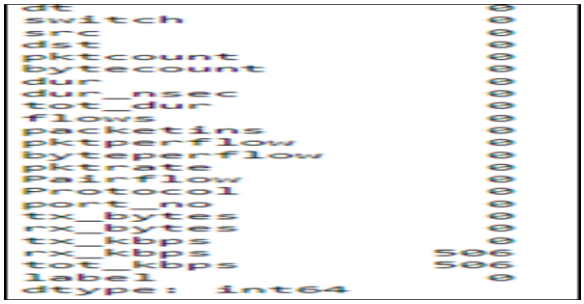

Figure 3: *Data preprocessing*

The null values are displayed in Figure 3. Any type of processing performed on raw data to get it ready for another data processing operation is referred to as data preprocessing, which is a part of data preparation.

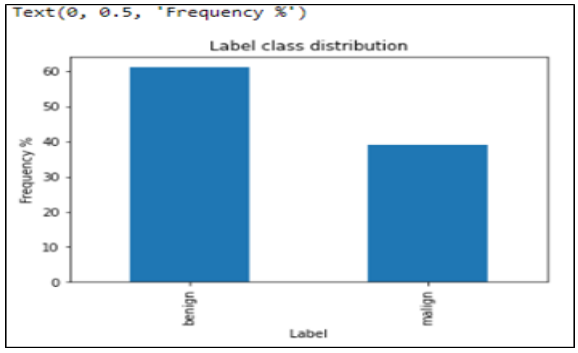

Figure 4: *Label class distribution*

Figure 4 illustrates the label class distribution, from the observation of the figure it is noted that the frequency of the benign is 60%, which is greater when compared to the malign.

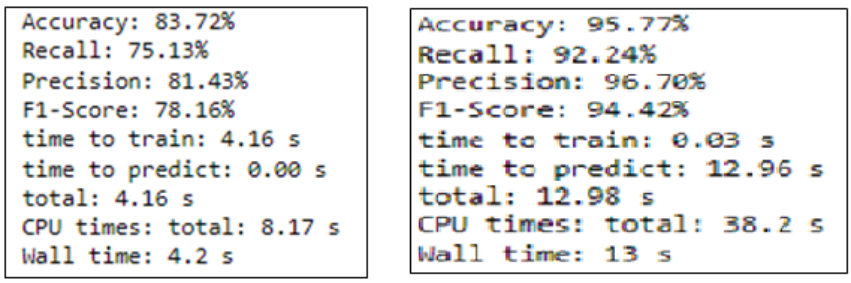

Figure 5: *(a) and (b) Performance matrix for Logistic regression and KNN*

A logistic regression performance matrix and KNN is shown in Figure 5 (a) and (b). Logistic regression is used to determine or forecast the probability that a binary (yes/no) event will occur. Data is imbalanced then other methods like ROC/AUC perform better in evaluating the model performance. Performance for KNN is illustrated above.

ISSN (Online): 2583-7052

International Journal of Advanced Trends in Engineering and Management (IJATEM)
Vol. 2(03), March 2023, pp. 1-9

Article Title: A DDoS Attack Categorization and Prediction Method Based On Machine Learning

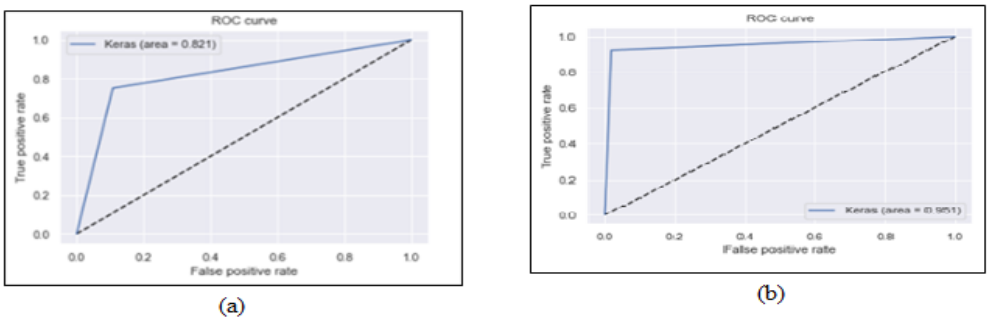

Figure 6: *(a) and (b) ROC graph for logistic regression and KNN*

A Logistic Regression and KNN for ROC Curve is depicted in Figure 6 (a) and (b). The receiver operating characteristic curve (ROC curve) is a graph which displays the way a classification model performs across all categorization levels. Two parameters are plotted on this curve: % True Positive. Rate of False Positives

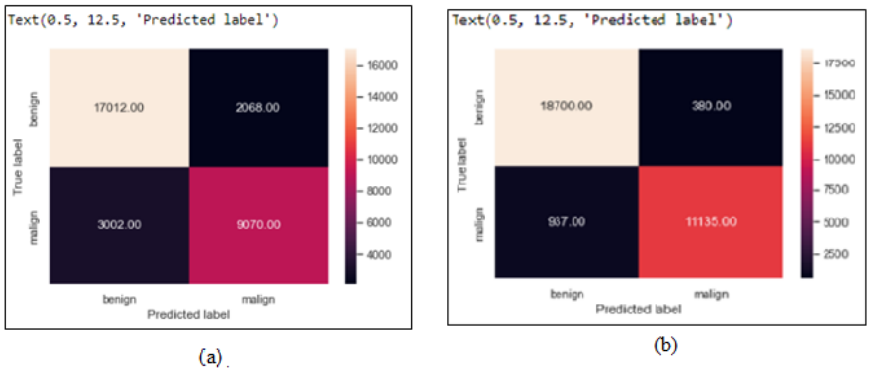

Figure7: *(a) and (b) confusion matrix for logistic regression and KNN*

Figure 7 (a) and (b) shows the confusion matrix for logistic regression and KNN. The creation of a confusion matrix, which is a 2 by 2 table which contrasts the predicted values from the model with the actual values from the test dataset

4 Conclusion

One of the most well-known and major cyber-attacks in recent memory is the DDoS attack. The victim's resources are intended to be reduced by the DDoS attack. The attacker floods the victim's side with traffic. As a result, these services would not be employed for a certain period of time, making it impossible to provide the service to legitimate clients. This work offers a thorough simulation of a machine learning technique to categories and forecast various DDoS attacks. Our suggested method is based on the following modules: pre-processing, classification and analysis of the data, and performance matrix. The first classification uses the KNN classifier technique for both Precision (PR) and Recall (RE) then the Precision (PR) and Recall are both classified using the DNN classifier technique in the second classification (RE).

According to the experiment results, the suggested machine learning-based DDoS attack detection system has a high detection rate for the current DDoS attack which is most prevalent and the results are done by the python software as illustrated in above figures.

References

1. Kalutharage Chathuranga Sampath; Xiaodong Liu; Christos Chrysoulas; Nikolaos Pitropakis; Pavlos Papadopoulos, Year: 2023, "Explainable AI-Based DDOS Attack Identification Method for IoT Networks," Computers, Vol: 12, no: 2, pp. 32.

2. Amrish, R., K. Bavapriyan, V. Gopinaath, A. Jawahar, and C. Vinoth Kumar, Year: 2022, "DDoS detection using machine learning techniques," Journal of IoT in Social, Mobile, Analytics, and Cloud, Vol: 4, no: 1, pp. 24-32.

3. Chuang, Hsiu-Min, Fanpyn Liu, and Chung-Hsien Tsai, Year: 2022, "Early Detection of Abnormal Attacks in Software-Defined Networking Using Machine Learning Approaches," Symmetry, Vol: 14, no: 6, pp. 1178.

4. Gaurav, Akshat, Varsha Arya, and Domenico Santaniello, Year: 2022, "Analysis of machine learning based DDoS attack detection techniques in software defined network," Cyber Security Insights Magazine (CSIM), Vol: 1, no: 1, pp. 1-6.

5. Mohmand Muhammad Ismail; Hameed Hussain; Ayaz Ali Khan; Ubaid Ullah; Muhammad Zakarya; Aftab Ahmed; Mushtaq Raza; Izaz Ur Rahman; Muhammad Haleem, Year: 2022, "A machine learning-based classification and prediction technique for DDoS attacks," IEEE Access, Vol: 10, pp. 21443-21454.

6. Zeinalpour Alireza; Hassan A. Ahmed, Year: 2022, "Addressing the Effectiveness of DDoS-Attack Detection Methods Based on the Clustering Method Using an Ensemble Method," Electronics, Vol: 11, no: 17, pp. 2736.

7. Perez-Diaz Jesus Arturo; Ismael Amezcua Valdovinos; Kim-Kwang Raymond Choo; Dakai Zhu, Year: 2020, "A Flexible SDN-Based Architecture for Identifying and Mitigating Low-Rate DDoS Attacks Using Machine Learning," in IEEE Access, Vol: 8, pp. 155859-155872.

8. Aljuhani Ahamed., Year: 2021, "Machine Learning Approaches for Combating Distributed Denial of Service Attacks in Modern Networking Environments," in IEEE Access, Vol: 9, pp. 42236-42264, 2021.

9. Biswas Rajorshi; Sungji Kim; Jie Wu, Year: 2021, "Sampling Rate Distribution for Flow Monitoring and DDoS Detection in Datacenter," in IEEE Transactions on Information Forensics and Security, Vol: 16, pp. 2524-2534.

10. Mishra Anupama, Year: 2022, "Prediction Approach against DDoS Attack based on Machine Learning Multiclassfier," arXiv preprint arXiv, pp. 2204.12855.

11. Perez-Diaz; Jesus Arturo; Ismael Amezcua Valdovinos; Kim-Kwang Raymond Choo; Dakai Zhu, Year: 2020, "A Flexible SDN-Based Architecture for Identifying and Mitigating Low-Rate DDoS Attacks Using Machine Learning," in IEEE Access, Vol: 8, pp. 155859-155872.

12. Agarwal Ankit; Manju Khari; Rajiv Singh, Year: 2021, "Detection of DDOS attack using deep learning model in cloud storage application," Wireless Personal Communications, pp. 1-21.

13. Dong Shi; Mudar Sarem, Year: 2020, "DDoS Attack Detection Method Based on Improved KNN with the Degree of DDoS Attack in Software-Defined Networks," in IEEE Access, Vol: 8, pp. 5039-5048.

14Tayyab Mohammad; Bahari Belaton; Mohammed Anbar, Year: 2020, "ICMPv6-Based DoS and DDoS Attacks Detection Using Machine Learning Techniques, Open Challenges, and Block chain Applicability: A Review," in IEEE Access, Vol: 8, pp. 170529-170547.

15. Zhijun Wu; Xu Qing; Wang Jingjie; Yue Meng; Liu Liang, Year: 2020, "Low-Rate DDoS Attack Detection Based on Factorization Machine in Software Defined Network," in IEEE Access, Vol: 8, pp. 17404-17418.