A Survey on DDoS attack in SDN-based Cloud Network using Machine Learning Algorithm

Dr.Kaladevi A C[(1)], Asma Begum M[(2)],Fathima S K[(3)]

Professor[(1)], PG Student[(2)],Asst.Prof[(3)]

Department of CSE, Sona College of Technology, Salem

***Abstract*-Cloud computing is an emerging technology that provides payable services and storage services to the legitimate cloud users. The problem of DDoS attack still prevails in SDN-based cloud; it leads to unavailability of services. SDN is a networking technology which separate control plane from date plane and also helps in detection of DDoS attacks. To overcome the problem of DDoS attacks, we introduce a hybrid algorithm (XGBoost and AdaBoost algorithm) that helps to detect the normal packet from attack packet. Some characteristics of SDN network would help to minimize the occurrence of DDoS attacks. The vulnerabilities in cloud can be solved by SDN network because of its advanced features like flexibility and dynamic programming environment. This paper gives clear ideas about machine learning algorithms that are used for classifying DDoS attacks. The performance evaluation metrics considered for hybrid algorithms are accuracy, execution time, false positive rate.**

***Index Terms*- Cloud Computing. Software Defined Network, Machine Learning, DDoS attacks.**

I. Introduction

Cloud computing technology used to deliver on- demand services with the help of broad network access. Due to the virtual development of technology, there occur some security issues. The most common problem is DDoS attacks, in which attackers targets the network resources or network servers to make it unavailable to its legitimate users [8]. SDN is a networking architecture which decouples the data plane from control plane. This network is adopted in data centres of cloud to solve the problem of Denial of service. This implementation of SDN in cloud network supports Networking-as-a-Service (NaaS) and provides many abstraction facilities to resolve the security problem including DDoS attacks. As cloud provides variety of service, this paper intended to provide Security-as-a-Service (SaaS) and Networking-as-a-Service (NaaS). SDN network provides complete protection for DDoS attacks because SDN has good features to defeat this attack [11]. The features include Separation of control plane from data plane. It simplifies the operational complexities when compared to traditional network. A centralized controller and global view of the whole network is responsible for implementation of policies and provides hardware abstraction to SDN applications. Software based traffic analysis: Because of software based traffic analysis, SDN greatly improves the DDoS attack detection and mitigation capabilities and dynamic updating of forwarding rules. The reason behind the occurrence of DDoS attacks is due to vulnerabilities present in cloud network. The vulnerabilities include broad network connectivity, pooling of resources, elasticity, metered service, and on-demand services. Broad network access can lead access can lead to sophisticated and immense DDoS attacks because of large number of users can access the network at same time and on-demand services may lead to botnets outbreak. Resource pooling provides service to

multiple clients because many users make request for the same resource and providers will make adjustment in resource and provide scalable services and this may lead victims to be more vulnerable to these kinds of attacks. The attackers make use of the advantages of rapid elasticity and measured service to launch more sophisticated attacks. By knowing these vulnerabilities, attackers launch DDoS attacks. The security to be provided in cloud network layer because SDN is a networking paradigm and DDoS attacks happens in network layer and it is specially focused on data plane and control plane of SDN network.

II. Background

Software defined Networking is a technology that supports Networking-as-a-Service (NaaS), in which it separates the control plane from data plane to provide various abstraction policies and flexible Networking support. In SDN, there is a kind of DDoS attacks known as table overflow attack. The data plane in SDN has switches which contains flow table entries. These flow tables contain some entries. Whenever a packet-in message arrives to flow table, it checks the flow table for matching entries. If the entries matches, the packet is kept and it is forwarded. If the packet entries don't match the available flow entries then the packet-in message is forwarded to the SDN controller. The controller will install rules and send it back to the data plane through North Bound API such as OpenFlow protocol. The vulnerability is that the size of flow table is minimum and handles up to 10000 entries. By using this vulnerabilities, the attackers as botnets, launch more unwanted packet requests to make service unavailable to legitimate packets [9].

Background of SDN:

SDN is a Networking technology that provides many abstraction facilities, flexibility and dynamic programming facilities to provide Networking service. There are three planes in SDN architecture. They are 1) Application plane 2) Control Plane 3) Data plane

1) Application plane: This Plane is an end user application plane.

2) Control Plane: this Plane contains SDN controller which as act as brain of the SDN network. SDN controller install rules and action for all packet-in message that come from data plane to control plane through open flow protocol. This protocol is used for communication especially for SDN network.

3) Data plane: the data plane act as forwarding plane. These Plane forward packets to other layers. The switches in data plane contain flow table and group table. The flow table is used to collect all the flows and acts as lookup table.

III. Machine learning techniques

Machine learning is a learning technique used for classification and regression. There are various kinds of machine learning techniques. They include supervised learning, unsupervised learning and semi- supervised learning. Supervised learning, in which they have labeled datasets therefore classification, is performed by learning the data from training phase. Unsupervised learning, in which the datasets are unlabelled datasets. This can done clustering techniques. This can be classified by considering the similarity measure of some instances and classify the new instances with the help of feature similarity. Semi-supervised learning is a combination of supervised and unsupervised learning [10].

Decision tree:

C5.0 is a type of decision tree algorithm used for classification. In [2], they focused on accuracy and time in comparison with c4.5 and naïve Bayes algorithm. By collecting the datasets from wire shark, a pre-processing phase is done to check for consistency of data. The types of DDoS attacks used for classification or detection of attacks are TCP flooding and UDP flooding. The attacker tool used to launch DDoS attacks is hping3. The c5.0 algorithm aims to provide high accuracy, faster detection rate, and low computational cost in comparison with c4.5 and naïve Bayes algorithm. It obtained 100% accuracy and 0.70 seconds detection rate.

Naive bayes:

Naïve bayes classifier is a supervised learning algorithm used to improve classification performance and also called as probabilistic classifier and makes prediction vastly and accurately. It is considered as a traffic classification method that is used to classify the normal packet and attack packet. It is used to identify those kinds of attacks that occur in uncertain situations. It is used to classify the packets efficiently when compared with other ML algorithms. In this algorithm [3] [7], they also used info gain feature selection method to reduce the number of parameters to decrease computation time. CAIDA dataset is used to classify the normal packet from attack packets. The parameters that are selected are MTI, POIP, TTL, ACK, SYN, and Time Stamp Field. This algorithm gives better accuracy of about 92.34% [3].

Support Vector Machine (SVM):

SVM is used for classification of packets. The main aim of this paper is to enhance the accuracy of detection of attacks and less positive rate. It is a non-probabilistic algorithm. They used DARPA datasets for classification. It is a supervised algorithm that helps to analyse the datasets and use the recognized patterns for classification of attacks. This algorithm helps to learn the pattern with small amount of training samples and will generate accurate classification by minimizing false positive rate. This is used to classify the unknown samples by learning from trained samples. SVM classifier helps to find a global optimum accuracy. In [5], they compared SVM with other algorithms such as naive bayes, bagging, j48 and random forest. SVM obtains an accuracy of about 95.11% and have very lower false positive rate of 0.008 when compared with naive bayes, bagging, j48 and random forest. As SVM takes more time for training and gives better results in case of traffic characteristics[5].

Random forest:

Random forest classifier is an ensemble based classifier in which decision tree algorithm is also used [13]. It is a supervised learning algorithm and also considered as a high accurate classifier for large datasets. It provides less error to detect denial of service and objective of this algorithm in DDoS attack detection is used to improve the accuracy of classifier. Random forest algorithm is used along with information gain feature selection algorithm. Information gain is used to select the particular features that aim to increase the accuracy of the classifier. Here, NSL-KDD dataset is used for testing and training phase. Firstly, the datasets are analysed using information gain model. After that select the proper feature set are selected, random forest algorithm is applied to classify the normal packet classification accuracy [13].

TABLE: 1

Comparison of algorithm with Accuracy in percentage, False Positive Rate (FPR), Training Time in Seconds.

SL.NO	ALGORITHMS	ACCURACY	FALSE POSITIVE RATE	TRAINING TIME
1	k-NN	99.05	2.74	23.933(ms)
2	Naive Bayes	90.14	0.02	3
3	Support Vector Machine	95.11	0.008	120
4	Random Forest	96.33	0.018	3.59
5	Self-Organising Maps	98.24	2.14	2.810
6	GBDT	97.69	0.013	21.67
7	Extreme Gradient Boosting	98.53	0.008	11.07

Self-Organizing Maps (SOM):

It is an unsupervised learning algorithm, in which they don't require any labelled data but there is a need to know about the characteristics of input data to be given. In [14], the proposed a hybrid algorithm by combining SOM and K-NN, both are type of artificial neural networks. This hybrid algorithm contains the accuracy of k-Nearest Neighbour and uses SOM for training to speed the classification. This paper aims at providing the both accuracy and computational overhead. SOM means mapping from high dimensional space to usually two-dimensional space and also helps in visualization of high dimensional data. K-NN is one of the simplest learning algorithms and is specially used for low dimensional. Both SOM and K-NN is based on generalization method in which performs generalization on all available cases and classify new instances based on similarity measure or feature similarity. The computational cost for classifying is new instances are high because it performs all computational attempts in classification. It is also classified based on nearest distance neighbour. The hybrid algorithm has obtained a detection rate of 98.24% as on Table: 1. and false positive rate of about 2.14 is achieved when compared with K-NN, SOM distributed- neurons and SOM distributed-center [14].

Extreme Gradient Boosting (XGBoost):

In [15], Extreme gradient boosting (XGBoost) is used as detection algorithm or classification algorithm in SDN network. The aim of this algorithm is improve higher classification accuracy, false positive rate, and fast speed in comparison with gradient boosting algorithm, random forest and SVM. This algorithm is an ensemble based classifier which combines weak classifier to form a strong classifier. The dataset used for XGBoost algorithm is KDD Cup 1999 dataset to classify the packets and also used maximum information gain, chi square statistics for DDoS detection and important features extracted among 41 features are 9 features. These 9 features are considered for classification. In [15], they have used interconnected cloud concepts, they created two cloud networks, by attacking one cloud network then the other cloud network is also under DDoS attacks. In this algorithm, the tuple with

misclassification error is considered as an important parameter for next classifier using error rate calculation. The process is repeated until the tuple is correctly classified. XGBoost has obtained an accuracy of 98.53%, false positive rate of 0.008 and training time of 11.07 as on Table:1.

Proposed algorithm:

The proposed algorithm is a hybrid algorithm that is XGBoost and AdaBoost algorithm. XGBoost is an ensemble based learning algorithm which combines certain weak classifiers to form a strong classifier. AdaBoost algorithm is also an ensemble based learning algorithm, which work same as XGBoost but when AdaBoost in addition with XGBoost will improves its performance by considering metrics such as accuracy, false positive rate. When an algorithm misclassifies a tuple, that particular tuple is sent to next classifier to perform correctly and gives a correct result without any misclassification. The misclassified tuple can be correctly classified by assigning correct amount of weight in each iteration and finally all weight is summed up to obtain good classification accuracy.

IV. Conclusion

In this paper, we conclude by saying that the proposed hybrid algorithm is better used for classifying the attack packets from normal packet. The environment used for SDN-based cloud network is Mininet and POX controller is used to detect traffic and to minimize the DDoS attacks. This performance evaluation metrics considered for the proposed method is accuracy in comparison with other algorithm such as SVM, Random forest and XGBoost algorithm, false positive rate.

REFERENCES

[1] Anku Jaiswal, Chidananda Murthy, Madhu BR, “Prevent DDoS attack in Cloud using Machine Learning”, International Journal of Advanced research in Computer Science and Software Engineering, 2016.

[2] Hariharan M, Abhishek H,K and Prasad B G, “ DDoS attack detection using C5.0 Machine Learning Algorithm”, I.J. Wireless and microwave Technologies.

[3] Hema V and Emilin Shyni C, “ DoS Attack Detection based on Naïve Bayes Classifier”, Middle east Journal of Scientific Research 23(Sensing, Signal Processing and Security): 398- 405, 2015.

[4] Janitza Punto, Kilhung Lee, “SDN-based DoS attack detection and mitigation system gor Cloud environment”, International Journal of Computer Systems, 2018.

[5] Kokila R T, Thamarai Selvi S, Kannan Govindarajan, “DDoS detection and analysis in SDN-based environment using Support Vector Machine Classifier”, International Conference on Advanced Computing, 2014.

[6] Krishna Reddy V, Sreenivasulu D, “Software- Defined Networkin with DDoS attacks in Cloud Computing”, International Journal of Innovative Technologies”,2016.

[7] Mahadev, Vinod kumar, Himani Sharma, “Detection and analysis of DDoS attack at application layer using Naïve Bayes classifier”, International Journal of Computer Engineering and Technology (IJCET).

[8] Marwane Zekri, Said El Kafhali Noureddine Aboutabit, “DDoS attack detection using Machine Learning Techniques in Cloud Computing Environments”, IEEE, 2017.

[9] Narmeen Zakaria Bawany, Jawwad A, Shamsi, Khaled Salah, “DDoS Attack detection and mitigation using SDN: Methods, Practices and Solutions”, Springer, 2017.

[10] Niharika Sharma, Amit Mahajan, Vibhakar Mansotra, "Machine Learning techniques used in detection of DDoS attacks: A literature review", international Journal of Advanced research in Computer Science and Software Engineering, 2016.

[11] Qiao Yan and Richard yu F, "Distributed Denial of Service attacks in Software-Defined Networking with Cloud Computing", IEEE Communication Magazines, 2015.

[12] Qiao Yan, Richard Yu F, "Software-Defined Networking and Distributed Denial of Service (DDoS) Attacks in Cloud Computing Environments: A survey, some research issues and challenges", IEEE Communications Surveys &Tutorials, 2016.

[13] Sanjay Agarwal, Reena Singh Rajput, "DoS attack detection using random Forest Classifier with Information Gain", IJEDR, 2017.

[14] Tran Manh Nam, Phan Hai Phong, "Self- Organizing map-based approaches in DDoS flooding detection using SDN", IEEE, 2018.

[15] Zhou Chen, Fu Joang, Yijun Cheng, XIn Gu, Weirong Liu, " XGBoost Classifier for DDoS attack detection and analysis in SDN-based Cloud", IEEE International Conference on Big Data and Smart Computing, 2018.