"Review on Xgboost "

Shaziya Shaheen [1], Prof.Naziya Pathan [2], Prof. Anuja Ghasad[3]
[1] M-Tech.scholar,Departement of computer science and engg.,RTM Nagpur University,Nagpur(MS),India.
[2] Asstt.Prof.,Departement of computer science and engg,RTM Nagpur University,Nagpur(MS),India.
[3] Asstt.prof.,Departement of computer science and engg,RTM Nagpur University,Nagpur(MS),India.

Abstract: The marriage of cloud and software defined network (SDN) can work out the challenge which exist in the typical cloud platform such as the private cloud isolation of user, network flow control. But in SDN based cloud, the SDN controller which manages the whole system is vulnerable to distributed-denial-of service (DDoS) attack, causing paralysis of the entire network. It is critical for SDN controller to be quick-speed, low false positive, and high precise against attack detection. In this paper, we are proposing gradient boosting (XGBoost), as detection method in SDN based cloud. In addition, we will use the POX as SDN controller, build SDN topology using Mininet to validate that our method performs higher accuracy, lower false positive rate, fast-speed and has scalability.

Keyword: SDN based cloud, XGBoost, DDoS attack detection, SDN Controller,Mininet.

1. INTRODUCTION

Today, information security solutions fall into two categories: analyst driven, or unsupervised machine learning driven. Analyst-driven solutions rely on rules determined by fraud and security experts. Experts derive these rules based on their experiences, and on the intuitions developed during investigations and analyses of previous successful attacks. This expert-driven process usually results in a high rate of undetected attacks (false negatives), and introduces a delay between the detection of new attacks and the implementation of countermeasures necessary to prevent future instances of these attacks. Moreover, bad actors often figure out current rules, and design newer attacks such that they can avoid being blocked or detected. We present XGBOOST, an analyst-in-the-loop security system where XGBoost is put together with state-of-theart machine learning to build a complete end-to-end XGBoost based solution. The system presents four key features: a big data behavioral analytics platform, an outlier detection system, a mechanism to obtain feedback from security analysts, and a supervised learning module. In the system we use a machine learning and big data combination system which outperforms other non-machine learning systems and provides a solution for network attack detection on real time datasets.

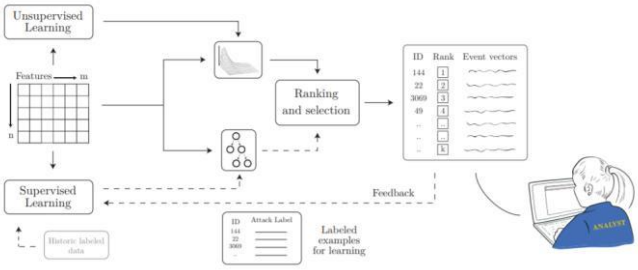

Fig:1 shows the Flow chart of the proposed system

2. RELATED WORK

In SDN based cloud, cloud computing is extended from server centralization and virtualization as well as storage centralization and virtualization to network centralization and virtualization [1]. Companiess like Google have already adopted SDN in their cloud data centers. However, the combination of SDN and Cloud has introduced potential cyber security risks. How to find a quick and accurate detection method for DDoS attack in SDN controller is extraordinary significant. Extensive existed researches for DDoS detection encountered challenges of scalability, accuracy, false positive rate and speed. In this paper, we consider a challenging problem in SDN controller for high accuracy, fast, low false positive rate of DDoS attack detection. In the SDN architecture, the controller acts as the SDN brain once attacked, and the entire network is paralyzed as well as the service unavailable. The key issue is how to find fast and accurate algorithm for detecting DDoS attacks in SDN controllers as soon as possible, and the algorithm is scalable for fast-growing network traffic. For DDoS attack detection, machine learning is an effective method which can detect it

against the controller. The previous researches in recent years mainly include the following methods. P. Xiao et al. [2] present an effective detection approach based on CKNN (K-arest neighbors traffic classification with correlation analysis) to detect DDoS attacks and a grid-based method named r-polling method for reducing training data involved in the calculation. The approach exploits correlation information of training data to improve the classification accuracy and reduce the overhead caused by the density of training data. In 2016, A. Saied, RE. Overill, and T. Radzik [3] select an Artificial Neural Network (ANN) algorithm to classify the exception flow from the normal flow which based on specific characteristic features (Patterns). W. Feng et al. [4] combine the SVM method with Self-Organized Ant Colony Network (CSOACNs) to take the advantages of both while avoiding their weaknesses. They take KDD 99 dataset for evaluation. In [5], authors propose a DDoS attack detection method based on hybrid heterogeneous multi-classifier ensemble learning and design a heuristic detection algorithm based on Singular Value Decomposition (SVD) to construct our detection system, which can achieve high accuracy and low false positive Rate (FPR) but having low speed. Although there have been a number of studies on DDoS attack detection, many of above methods are based on traditional networks and may encountered challenges of scalability, accuracy, false positive and speed. To address the speed, scalability, accuracy and false positive issue. We use the XGBoost method as detecting algorithm. Differentiate the network flows based on certain features related to traffic characteristics and categorize them as malicious or benign to the controller [6]. First, XGBoost classifier is a boosting classifier which combines hundreds of tree models with lower classification accuracy, and it generates a high accurate and low false positive model through the constant iteration of the model. Second, XGBoost uses two techniques to accelerate the algorithm. On the one hand, with the help of OpenMP that XGBoost can automatically utilize the multicore of single machine CPU for parallel computation; on the other hand, XGBoost defines a data matrix class-Dmatrix, which will be pre-processed at the beginning of the training.

3. PROPOSED SYSTEM

System Model In this paper, in view of the limited equipment configuration, the testing scenario was simply implemented based on [10] topology. We will use Mininet to build a SDN based cloud network and the POX as the controller for SDN. Fig. 1 depicts the topology used for the system. We will consider an attack scenario in several interconnected clouds, where a device (switch, host) in one cloud suffers a malicious DDoS attack on other cloud devices (switches, hosts) We will select Windows logs and other datasets available on the UCI repository for the same purpose.First we will Collect data from various datasets followed by preprocessing of data where outliers detection and removal will be performed,We will then use the processed data to perform classification using big data and machine learning lastly we will add machine learning to improve detection accuracy

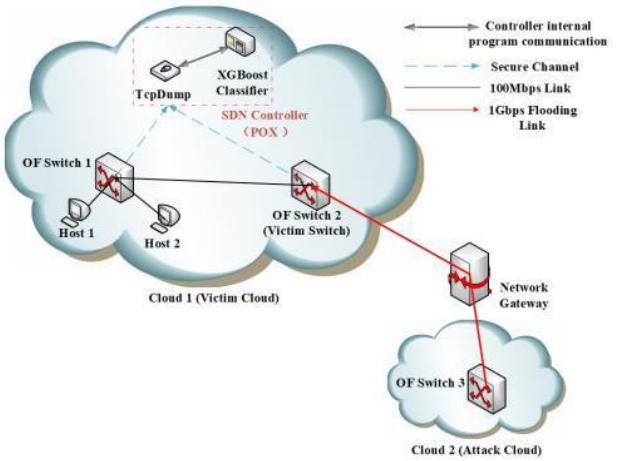

Fig. 1. System Model of SDN-based Cloud With Attack

4. Research Methodologies

4.1 Xgboost

XGboost is an optimized distributed gradient boosting library designed to be highly **efficient**, **flexible** and **portable**. It implements machine learning algorithms under the Gradient Boosting framework. XGBoost provides a parallel tree boosting (also known as GBDT, GBM) that solve many data science problems in a fast and accurate way. The same code runs on major distributed environment (Hadoop, SGE, and MPI) and can solve problems beyond billions of examples.XGBoost stands for “Extreme Gradient Boosting”, where the term “Gradient Boosting” originates from the paper *Greedy Function Approximation: A Gradient Boosting Machine*, by Friedman. This is a tutorial on gradient boosted trees, and most of the content is based on these slides by Tianqi Chen, the original author of XGBoost. The **gradient boosted trees** has been around for a while, and there are a lot of materials on the topic.

XGBoost is used for supervised learning problems, where we use the training data (with multiple features) xixi to predict a target variable yiyi. Before we learn about trees specifically, let us start by reviewing the basic elements in supervised learning.

4.2 Software define Networking(SDN)

SDN is short for software defined networking. Software defined networking (SDN) is an approach to using open protocols, such as OpenFlow, to apply globally aware software control at the edges of the network to access network switches and routers that typically would use closed and proprietary firmware.Software defined networking offers numerous benefits including on-demand provisioning, automated load balancing, streamlined physical infrastructure and the ability to scale network resources in lockstep with application and data needs. As noted on Enterprise Networking Planet, coupled with the ongoing virtualization of servers and storage, SDN ushers in no less than the completely virtualized data center, where end-to-end compute environments will be deployed and decommissioned on a whim.

5. CONCLUSION

SDN based cloud has received more and more attention in recent years as a new type of cloud. Cloud security is one of the most important network security. Therefore, the security of SDN based cloud must be solved. This paper Proposed techniques of solving the DDoS attack of controller in SDN based cloud. The machine learning algorithm is proposed to detect DDoS attacks by analyzing attack traffic patterns. By comparing performance, we can see that the proposed XGBoost algorithm has higher accuracy and lower false positive rate than other algorithms. In addition, We will compare the XGBoost in high-speed Internet environment.

6. REFERENCES

[1] Q. Yan, FR. Yu, Q. Gong, and J. Li "Software-Defined Networking(SDN) and Distributed Denial of Service(DDoS) Attacks in Cloud Computing Environments: A Survey, Some Research Issues, and Challenges," IEEE Communications Surveys and Tutorials, vol. 18, no. 1, pp. 602-622, 2016. [2] P. Xiao, W. Y. Qu, H. Qi, and Z. Li, "Detecting DDoS attacks against data center with correlation analysis," Computer Communications, vol. 67, no. C, pp. 66-74, 2015. [3] A. Saied, RE. Overill, and T. Radzik, "Detection of known and unknown DDoS attacks using artificial neural networks," Elsevier Science Publishers B. V, vol. 172, no. C, pp. 385-393, 2016. [4] W. Feng, Q. Zhang, G. Hu, and J. Huang, "Mining network data for intrusion detection through combining SVMs with ant colony networks," Future Generation Computer Systems, vol. 37, no. 7, pp. 127-140, 2014. [5] B. Jia, X. Huang, R. Liu, and Y. Ma, "A DDoS Attack Detection Method Based on Hybrid Heterogeneous Multiclassifier Ensemble Learning," Journal of Electrical and Computer Engineering, vol. 2017, no. 2, pp. 1-9, 2017. [6] NZ. Bawany, JA. Shamsi, and K. Salah, "DDoS Attack Detection and Mitigation Using SDN: Methods, Practices, and Solutions," Arabian Journal for Science and Engineering, vol. 42, no. 2, pp. 425-441, 2017. [7] T.Chen, and C.Guestrin, "Xgboost: A Scalable Tree Boosting System," in KDD, pp. 785-794, 2016. [8] N. Dayal, and S. Srivastava, "Analyzing behavior of DDoS attacks to identify DDoS detection features in SDN," in International Conference on Communication Systems and Networks(COMSNETS), pp.274-281, 2017. [9] Y. Jarraya, T. Madi, and M.Debbabi, "A survey and a layered taxonomy of software-defined networking," IEEE Communications Surveys and Tutorials, vol. 16, no. 4, pp. 1955-1980, 2014. [10] R. Braga, E. Mota, A. Passito, "Lightweight DDoS flooding attack detection using NOX/OpenFlow," IEEE Conference on Local Computer Networks, vol. 8, no. 1Suppl1, pp. 408-415, 2010. [11] Knowledge Discovery and Data mining (KDD) Cup 1999 dataset available from: http://kdd.ics.uci.edu/databases/kddcup99/kddcup99.html.

Shaziya Shaheen Wakeel Parvez (B.E) is pursuing M.Tech. degree in Computer Science and Engineering from Nuva college of Engineering,RTM Nagpur University,Nagpur,India.