International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 4 Issue 6, September-October 2020 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Review Paper on Predicting Network Attack Patterns in SDN using ML

Dr. C. Umarani[1], Gopalshree Kushwaha[2]

[1]Associate Professor, [2]Student,
[1,2]Masters in Computer Applications, Jain (Deemed-to-be) University, Bangalore, Karnataka, India

ABSTRACT

Software Defined Networking (SDN) provides several advantages like manageability, scaling, and improved performance. SDN has some security problems, especially if its controller is defense-less over Distributed Denial of Service attacks. The mechanism and communication extent of the SDN controller is overloaded when DDoS attacks are performed against the SDN controller. So, as results of the useless flow built by the controller for the attack packets, the extent of the switch flow table becomes full, leading the network performance to decline to a critical threshold. The challenge lies in defining the set of rules on the SDN controller to dam malicious network connections. Historical network attack data are often wont to automatically identify and block the malicious connections. In this review paper, we are going to propose using ML algorithms, tested on collected network attack data, to get the potential malicious connections and potential attack destinations. We use four machine learning algorithms: C4.5, Bayesian Network (BayesNet), multidimensional language (DT), and Naive-Bayes to predict the host which will be attacked to support the historical data. DDoS attacks in Software Defined Network were detected by using ML-based models. Some key features were obtained from SDN for the dataset in normal conditions and under DDoS attack traffic.

KEYWORDS: *Prediction, SDN, DDoS, Machine Learning, Algorithms*

How to cite this paper*: Dr.* C. Umarani | Gopalshree Kushwaha "Review Paper on Predicting Network Attack Patterns in SDN using ML" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-6, October 2020, pp.1635-1638, URL: www.ijtsrd.com/papers/ijtsrd35732.pdf

IJTSRD35732

INTRODUCTION

The extreme increase in the number of devices connected to the Internet has resulted in a number of useful solutions in different fields. Such an enlarge in demand for connectivity has challenged the traditional network architectures. To account for the challenges, Software Defined Network (SDN) architecture was preferred to disassociate the conventional user and control plane. An attacker can execute attacks, like Secure Shell brute force attack, on the Software Defined Network (SDN) controller which will end in important security threats. Even if the network administrator detects a possible attack and attacker, it may not be possible to efficiently account for simultaneous attacks in real time. Therefore, there is a need for certain security rules that are usually performed on the SDN controller close to firewall rules. Malicious users have certain advantages that can be used to transform between the attackers and authorized users. Patterns such as correlate attacks and sharing of password dictionaries are similar among attackers. Different techniques, including machine learning, can be used to detect such patterns. Machine learning based schemes have shown remarkable potential in classification of the users. Nowadays with the advent of 4G networks and economic smart devices there is a huge growth in the usage of the internet that has become a part of daily life.

A wide range of services provided over the internet in various application areas such as business, entertainment, and education, etc. made it an essential part in framing various business models. These factors made security over wireless networks as the most key factor while using the internet from insecure connections. Various security algorithms and frameworks are advances to qualify the protection from Internet based attacks while conceive high performance Intrusion detection systems (IDS), which behave as a defensive wall while resisting the attacks, over internet based network devices. Distributed framework based computing environments like cloud computing and IOT are more liable towards DDoS attacks in which multiple devices are coordinated to launch attacks over distributed targets. DDOS attacks are primarily launched within the context of exhausting the connectivity and therefore the processing of the target server resources during which it enables the access constraints to the authorized users to use the services provided by the target server that points towards the partial or total unavailability of the services. The phenomenon of distributed computing is predicated on the one-to-many dimension during which these sorts of attacks. It's observed from the previous research studies that the damage capacity, also because the disrupting nature of the DDoS attacks, is gradually increasing with the speed of internet usage. Traditional network infrastructures are unable to deal with certain requirements like high bandwidth, accessibility, high connection speed, dynamic management, cloud computing, and virtualization. Thus, Software Defined Networking (SDN), which features a flexible, programmable, and dynamic architecture, has emerged as an alternative to traditional networks. SDN model comprises control, data, and application planes.

Devices, like switches and routers, are placed on the info plane. The control plane is liable for the management of transmission devices placed on the info plane. The controller, which performs because the brain of the network, is found on this plane. Devices on the info plane perform packet transmission consistent with the principles set by the controller. The appliance plane communicates with the devices located on the network infrastructure via the controller.

The controller is exposed to DDoS attacks through the communication line between the controller and therefore the data plane. DDoS attacks direct an outsized amount of traffic to the OpenFlow turn on the info plane. If packets arriving at the OpenFlow switch don't match with the flow input within the flow table (miss flow), packets are taken into the flow buffer. Then, they're transmitted to the controller with the Packet-In message to write down a replacement rule. During this phase, the sources of the controller remain incapable and therefore the network becomes inoperative. The bandwidth of the communication line between the controllers that's exposed to attack traffic. Therefore, network performance severely declines. The info plane is exposed to DDoS attacks through the flow table located within the network devices. In DDoS attacks, data packets are sent to the switch from unknown sources. The controller mentions a rule for these arriving data packets and forwards them to the flow table of the switch. After a few times, the capacity of the switch flow table becomes full. Consequently, new rules can't be written to the flow table and packets can't be forwarded. The incoming packets fall automatically because the buffer capacity becomes full with the attack.

Thus, decreasing DDoS attacks in SDN is harder in comparison to traditional networks. ML based approaches can provide more dynamic, more efficient, and smarter solutions for SDN management, security. So as to make sure network security, the detection of DDoS attacks is extremely important for taking the required measures during a timely manner. Processing SDN flow data with the ML-based DDoS attack detection systems integrated into SDN structures can cause a self determining network that's ready to learn and react. Moreover, SDN having an integrated machine learning application could also be a reference model in forming a secure structure in studies providing for the mixing of 5G networks.

Related Work

Literature Review

Machine learning techniques that would be won't to handle intrusions and DDoS attacks in SDN (Software Defined Networks). Neural networks, Bayesian networks, symbolic logic, genetic algorithms and support vector machines and their application in SDN anomaly detection. Unlike anomaly detection approaches in SDN like kNN(k-Nearest Neighbors), Bayesian Networks, Support Vector Machines, and Expectation Maximization. Machine learning scheme offers, C5.0 classifier, to assort the traffic in SDN, and gathers ground fact data using crowd sourcing mechanism to integrate with the centralized control of SDNs data reporting system.

DDoS Attacks against the Controller

Control functions are taken from the switch and given to the controller, which is the brain of the network in SDN architecture. Parent level rules are easily applied to the network with the assistance of the controller. The controller can add new rules to the transmission devices and alter the prevailing rules. It can do these changes by sending data with transmission devices via a secure channel through the Open-Flow protocol. Continuity and therefore the unity of knowledge traffic are ensured through this channel. If this secure channel interrupts, the connection connecting the controller and transmission devices fails. SDN architecture is the target for DDoS attacks. While the cracker is cracking the SDN network, it's three main targets, as shown in Figure 1.1: to overwhelm the sources of the controller, to take over the bandwidth of the channel between the controller and therefore the switch, and to fill the flow tables within the switch with unnecessary flows. In DDoS, attacks over the controller, the attacker transmits a massive number of data packets to the Open Flow switch via zombie users. It's tough for the controller to differentiate between traffic sent by the attacker and legal traffic.

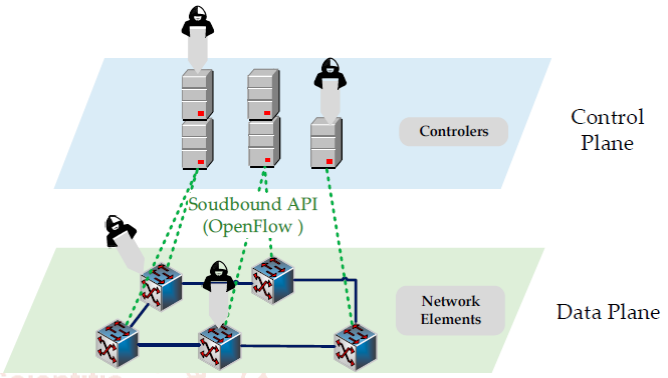

Fig 1.1 Targets of DDoS Attacks in SDN [1]

The Open Flow switch seeks a match within the flow input by checking the packet header (source port, target port, source IP address, target IP address, etc.). If there's no match, the packet is forwarded to the controller by encapsulating the packet header within the flow request with the Open Flow protocol (OFPT) PACKET_IN message. The controller answers with the OFPT FLOW_MOD message. This message involves the method to be administered on the packet and therefore the timeout of the flow within the flow table assigned for the packet. Because the number of packets forwarded to the controller increases, the sources of the controller are consumed (bandwidth, memory, and CPU), new flow input for the new legal packets arriving at the network can't be processed, and therefore the SDN architecture collapses.

The procedure of conducting a scientific literature review (SLR). The most intention of conducting SLR is to execute a well-planned literature study within the context of answering the Research Questions that are framed at the initial stage of the study. SLR enables the researchers to get, evaluate and amalgamate the research studies conducted by various network security researchers. The event of SLR includes the subsequent process:

- Development of a review protocol that has complete procedure involved in conducting SLR of predicting DDoS attacks using the machine and deep learning applications.
- Primary and secondary selection strategy to filter the articles addressing the research questions.
- Synthesize the chosen studies to answer the research questions.

ML Algorithms

We apply four different machine learning algorithms:

1. C4.5 Decision Tree
2. BayesNet (BN)
3. Decision Table (DT)
4. Naive-Bayes (NB)

Machine learning algorithms have been widely used for a number of classification and prediction problems and have provided accurate results. We describe some of the most used ML algorithms.

1. C4.5 Decision Tree:

C4.5 Decision tree is used for inductive speculation. In C4.5, the discrete-valued functions are approximated and a decision tree is employed for representing the learned function. In C4.5, the data is partitioned into sub-parts periodically. C4.5 is vigorous to very noisy data and is favored for learning various disconnective expressions. As discussed in, the main learning steps of C4.5 DT are:

- An attribute is opted, relay on which a logical test is formulated.
- The operation is run recursively on all the child nodes.
- A leaf is reveal as a node
- based on certain termination rule

2. Bayesian Network:

Bayesian Network encrypts probabilistic relationships among different variables of interest. It consists of a variety of variables and a set of edges between the variables, leading to an acyclic graph. Every node within the graph represents the variant and a directed edge from one variable to a different. Every variable within the Bayesian network is independent of the no descendants. Bayes Net has been used as a classifier and if trained properly, may result in highly accurate classifications.

3. Naive-Bayes:

Naive-Bayes uses Bayesian theory that predicts the sort of the unknown samples supported prior probability using the training samples. The Bayesian classification model depends on statistical analysis. Bayesian theory that comprises Bayesian learning. Bayesian learning uses the prior and posterior probability together and uses it to seek out the posterior probability as per the supplied information and data samples. The Naive-Bayesian algorithm operates by segregating the training set into an attribute vector and a choice variable. The algorithm also assumes that each member of the attribute vector independently acts on the choice variables.

4. Decision Table:

Decision Table (DT) is employed to arrange and document logic during a way that assists in easy inspection. DT helps in representing the machine learning output because the input, and involves selecting a number of the info attributes. They also aid in assessing different sets of rules for obscurity and redundancy.

Methodology

We use machine learning (ML) algorithms to predict potential target host attacks supporting the historical network attack data for SDN. We have four algorithms: C4:5, BayesNet, DT, and Naive-Bayes for predicting the system that might be attacked, and compare their performance in terms of accuracy. Figure 1.2 shows our ML-based architecture for outlining security rules on SDN controllers. Historical data is employed to coach a model. The trained model is then handed-down for predicting the probable attacks on different systems using real time network data, and certain rules are passed to the SDN controller to dam the potential attacker.

Some principles of our proposed scheme are:

1. Use historical data to train the ML-based models
2. Use the trained model to identify potentially vulnerable host

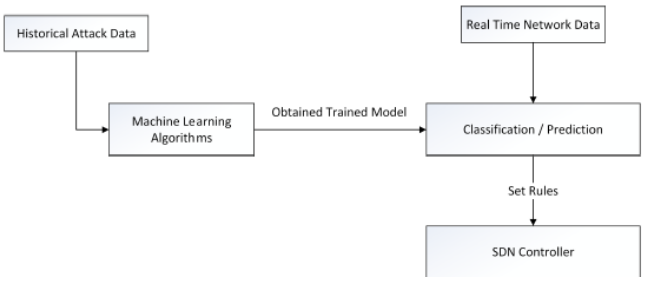

Fig. 1.2: Machine learning based architecture for defining security rules on SDN controller [2]

1. Use historical dataset to train the ML-based models

To obtain accurate classifiers to spot potential vulnerable hosts, historical data is required to coach the ML-based models. The training helps the model learn and acquire better results. Our goal is to use the historical attack pattern to spot which hosts are going to be attacked by a possible attacker. Supported attacker IP, we predict the potential host which will be attacked. These predictions are often wont to define security rules on SDN controllers to determine the network security. Instead of restricting the access of one IP, we also propose blocking the whole subnet to stop the longer term attacks from an equivalent attacker, attacking through a special IP within the same subnet.

2. Use the trained model to detect potentially vulnerable host

Once the model is trained, it is often won't to identify potential hosts which will be attacked by an IP. During the testing phase, we used our trained model to predict the attacked host supported the attacker's IP. If the attacker actually attacked a number as predicted by the ML algorithm, it means the model is accurate.

Conclusion

In this paper, we revived a machine learning algorithm to predict the vulnerable host in an SDN network that is highly likely to be attacked. The prediction output of ML algorithms, the safety rules for SDN controllers are often defined which will prevent malicious users from accessing the network. Experimental results showed that ML algorithms can help in defining security policies for SDN controllers by predicting the potential vulnerable system. Security rules for SDN controllers are often defined which will prevent malicious users from accessing the network. The Bayesian network was up to accurately predict 251,321 attacks.

References

[1] Detecting DDoS Attacks in Software-Defined Networks Through Feature Selection Methods and Machine Learning Models, Huseyin Polat, Onur Polat

and Aydin Cetin

[2] Predicting Network Attack Patterns in SDN using Machine Learning Approach, Saurav Nanda, Faheem Zafari, Casimer DeCusatisy, Eric Wedaaz and Baijian Yang

[3] Prediction of DDoS Attacks using Machine Learning and Deep Learning Algorithms, Saritha, B. RamaSubba Reddy, A Suresh Babu

[4] A taxonomy of DDoS attack and DDoS defense mechanisms, Mirkovic, Jelena and Peter Reiher

[5] Analyzing Distributed Denial of Service Tools: The Shaft Case, Dietrich, Sven, Neil Long, and David Dittrich

[6] Worldwide ISP Security Report, Arbor Networks Lee, Wenke, and Salvatore J. Stolfo.

[7] A systematic review of systematic review process research in software engineering, Barbara Kitchenham and Pearl Brereton

[8] Data Mining: Practical machine learning tools and techniques, I. H. Witten and E. Frank, Morgan Kaufmann, 2005.

[9] The power of decision tables in Machine Learning, R. Kohavi, 1995.

[10] Supervised machine learning: A review of classification techniques, S. B. Kotsiantis, I. Zaharakis, and P. Pintelas, 2007