AI Based Two Stage Real Time Intrusion and Phishing Detection System for Software Defined IoT Networks.

Geethu M Suresh[1], Prof Minu Lalitha Madhav[2]

[1]*PG Scholar, Sree Buddha College of Engineering, Kerala, India*
[2]*Asst.Professor, Sree Buddha College of Engineering, Kerala, India*
[1]*geethusuresh1996@gmail.com,* [2]*minulalitha@gmail.com*

Abstract: *A software defined IoT network is a facility for resource sharing, traffic management and effective utilization of individual IoT devices connected to the network. The data from these IoT Network passing To and Fro through the network are subjected to certain vulnerabilities and also these networks are prone to intrusions. So rather than an evaluation based on traditional Datasets of IoT network an experimental evaluation based on real time data should be carried out. The proposed system captures the real time data from software defined IoT network. The system captures real time packet data and performs an initial filtering over the data. This stage is known as the Deep Packet Inspection. Since the algorithms are performed over real time data any vulnerabilities in the data can be easily detected and reported. A new module called PhishingLimiter is added to the existing system to and then leverage it with Software Defined Networking (SDN) to identify phishing activities.*

Keywords: **Artificial intelligence, IoT network, SDN, Real Time, Intrusion Detection, Phishing**

1. INTRODUCTION

Internet of Things is an emerging technology which allows the interconnection of various devices such that they can interact with each other. This involves communication, resource sharing etc. As the number of devices connected in such a network, there should be a centralized system to control. Software Defined IoT is such a globalized control. There are certain performance characteristics for an IoT network. Scalability, portability, availability, flexibility are some of the constraints.

When the number of devices in such a network increases, the above constraints also needs to be satisfied. The SDN control the efficient management of these constraints. But SDN cannot eliminate the security issues and challenges that may arise in the network data flow. Serious issues and challenges against network data may arise any time in an IoT network. This can be only be addressed by an efficient intelligent algorithm.

Intrusion detection can be generally classified into two types, Misuse Detection and Anomaly Detection. Misuse based systems always learns the properties of the network and based on this historical knowledge any violation from this is considered as an attack. Whereas in Anomaly based also frames a network behavior but also reports novel attacks.so artificial intelligence techniques can be used for both types of detection system since both of them needs the capturing of prior data.

We propose an AI based two stage intrusion detection system which detects vulnerabilities in real time data rather than on a traditional data set. The technique uses BAT and improved Random forest algorithm for feature extraction and Classification. These algorithms are applied on NSL KDD Cup dataset which is assumed to have all the features of an IoT packet captured from a software defined IoT networks. For improving the efficiency of the clustering, a swarm division is performed on the dataset and K-means clustering algorithm is used. All the classification algorithms use features selected by the proposed BA as input. By altering the weights of samples, each tree in the forest can be trained more effectively through picking the samples which are more frequently to be misclassified.

The present algorithms always relies on standard datasets like NSL KDD,so the change in vulnerable data will not be treated with present algorithms. So a real time network data

should be captured to identify the potential intrusions reported. A New module called PhishingLimiter is added to the existing system to and then leverage it with Software-Defined Networking (SDN) to identify phishing activities. A new dynamic phishing detection and mitigation approach using SDN. It utilizes the programmability of SDN to address the dynamics of phishing attacks that cannot be handled in existing system.

The remaining of the paper is organized as follows. Section 2 discusses the background research works. Section 3 introduces the proposed system. Section 4 evaluates the performance of the system. Section 4 concludes the paper and Points out the Future work.

2. RELATED WORKS

In [1] Tang et.al proposed a deep learning network to detect intrusion in software defined IoT networks. The deep neural network consists of three layers, an input layer, three hidden layers and an output layer. The input quantity is six and the output quantity is two. There are twelve, six and three neurons in the hidden layers respectively. The NSL KDD dataset is used to evaluate the results. Each traffic packet has forty one features that are divided into three types of features: basic features, content based features, traffic-based features. Attacks in the dataset are categorized into four categories according to their characteristics.

In [2] An Le et.al proposed a flexible network based intrusion detection system for SDN. OpenFlow switches are used instead of the normal switches, the IPDS sensors and the firewall. The management server receives the information from the OpenFlow switches. In case of intrusion detection, the management server will interact with the controller to deploy new rules over the switches. There are several advantages of this approach. On one hand, total cost of the IPDS system might reduce as the sensors, firewall and the normal switch at the network boundary are no longer needed. On the other hand, the possibility of the management server being down caused by the overwhelming traffic from the sensors is eliminated.

In [3] Sgarciu et.al proposed a wrapper based approach for intrusion detection in IoT networks. Proposed feature selection method is a wrapper based approach and it is implemented for intrusion detection. SI algorithms are structured on two main components: exploration (randomization or diversification of the searched solution) and exploitation (refers to improving the candidate solution). In the case of BA, as the iterations flow the algorithm loses its exploration because the pulse emission rate increases exponentially and condition rand > ri becomes less probable to be satisfied. Also, the algorithm can get trapped into local minima as the new solution is generated near the best solution and with a small variation given by a decreasing loudness.

In [4] Kim et.al implemented the IDS classifier based on LSTM-RNN and evaluated the IDS model. For training phase, they generated a dataset by extracting instances from KDD Cup 1999 dataset. In order to find the proper learning rate and hidden layer size, took an experiment with changing the values. For testing phase, 10 test dataset were made and measured the performance. By comparing it to other IDS classifier, it was found that the attacks are well detected by LSTM-RNN classifier. Because of the highest DR and Accuracy even though the FAR is slightly above the other ones. To improve the FAR, the relation between the dataset and initialized weight values are analysed. The reason is that the performance was changed even though all hyper-parameters are the same. The one variable is the initialized weight values.

In 2005, Stein et. al. [5] propose a hybrid IDS model combining Genetic Algorithm (GA) and Decision Trees. Authors use GA to perform searches for the subset of features evaluated with C4.5 and use the sum of the validation error rates as the

fitness function. Their tests on the KDD99 dataset prove their wrapper feature selection algorithm outperforms the standard C4.5.

In [6] proposed an anomaly network intrusion detection approach using Information Gain for feature selection and Support Vector Machine optimized with Swarm Intelligence for classification. Proposed IDS model has three different phases. First the data set is pre-processed by transforming the symbolic valued attributes to numeric and applying the discretion algorithm. Then, IG is used for feature selection and SVM is used for classification. For SVM, the parameters are selected by the swarm intelligence algorithm (ABC or PSO).

In [7] Santo et.al proposed ATLANTIC, a framework for anomaly detection, classification and mitigation in SDN-based networks. The framework comprises a lightweight phase responsible for monitoring traffic flows and a heavyweight phase responsible for anomaly classification and mitigation. As a result, traffic anomalies can be categorized and the information collected can be used to handle each traffic profile in a specific manner, such as blocking malicious flows.

3. PROPOSED SYSTEM

3.1. Dataset Collection

The collection of raw data packets is the prime step for the proposed system. A small smart office network is built to collect network data. Data is collected for a period of 14 days. The experimental smart home consists of four IoT devices: a camera, motion sensing AC,smart bulb and smart plug. The MAC addresses of the devices are used to split the network traffic into different ARF files corresponding to different IoT devices. These ARF files have timestamps and all the protocol information's. The training data is collected for a period of around 10 days and test data for next 4 days.

3.2 Feature Extraction and Selection

We use BAT algorithm [8] to perform feature extraction. It is a heuristic based algorithm.It absolutely was impressed by the echo sounding behavior of micro-bats with varied pulse rates of emission and loudness. The echolation property of identical bats are considered which moves for their target.
In order to perform the BAT effectively a swarm division is introduced [8].There are two steps in this procedure. In the first stage, general individuals within each subgroup only learn from their present local minima and direct to it through altering their velocity, which is influenced by both the local minima and their historical best positions. In the second level, the local minima with higher fitnesses likewise aim toward the global optima so that global optimal information can be exchanged. In this way, bats hierarchically learn from the superiority and move slightly toward a better position iteratively without being dominated by some specific bats. All the bats will be regrouped after each iteration until the algorithm terminates.
After updating the location of bats at each iteration, the mutation mechanism of Differential Evolution [10] to BA algorithm, which enhances the diversity of the population and the ability of bats to jump out of local optima. It disturbs the target value by using the differences between randomly selected individuals in the swarm. Since the original method can only solve continuous optimization problems, a binary differential evolution algorithm using bit operations based on swarm division was used in the system

3.3 Classification

A. Weight Initialization

The weight of each sample is initialized similarly as (1/N) where N is the total number of the samples in the dataset and the total of the weights is 1. In this way, each sample has anequal probability of being selected. In this each training sample is initialized with a different weight according to the class it belongs to.

B. Weight Updating

After building each tree, the weight of each sample is updated in relation to the result it is classified. The weights of misclassified samples are increased while samples classified correctly are reduced. As a result more concern is on the samples which are misclassified. Those samples with higher weights will be selected and learned in the next tree.

C. Weighted Voting

The classification ability of each tree varies among different classes, the traditional method using the majority votes of all the trees to obtain the final result cannot be used anymore. The weighted voting mechanism to the ensemble trees is introduced

3.4 Phishing Detection

A phishing detection module Called Phishinglimiter [9] is added to the system for detection of the phishing attempts. The system uses two different technologies a store and forward approach and a forward packet method.

Case 1. Within SF, packets are stored in a buffer and detection for malicious activities as denoted is performed. During the examination of pkt through ANN approach, if pkt is phishing, then it is dropped, and forward if it is not as denoted in. sj−1 is updated and compared again for the next pkt.

Case 2. In FI mode pkt is forwarded to the destination and copied for inspection The copied pkt is dropped once PhishLimiter determines if it has malicious intent or not.Processes of updating for sj and comparing sj > sovs for each new incoming packet.

In order to change the direction of F REpresentational State Transfer Application Programming Interfaces (REST API) procedural calls within Floodlight as denoted in where we utilize the Static Entry Pusher module to manipulate flow traffic.

4. PERFORMANCE EVALUATION

The system uses the real time dataset generated from the smart office network which is observed for a period of 14 days. The training set is generated with the first 10 days. The test set was generated with 4 days. This dataset are stored in ARF format. This dataset consist of around 21000 entries. The test set consists of around 14000 entries. The data set consists of all the packet headers of the TCP/UDP protocol. The real time packets were mapped to the packet headers and for each set of devices based on the peculiarities of the devices the ARF files of each devices were created.

As the number of devices increases a time period is assigned to randomly generate the data set. So the Complexity of the process of generation of the dataset depends on the random time only and this will be O(n).

The BAT algorithm and Random Forest also give efficient results over the real time data rather than on the traditional datasets. The phishing algorithm also works well for the randomly generated dataset.

5. CONCLUSION

The proposed system performs intrusion detection on real time data rather than on a traditional data set. This can perform better detection than existing system. The traditional algorithms perform only on the standard datasets so an evaluation based on real time data will bring out more cases of vulnerabilities. Also a phishing detection module is added to the system. This will identify the phishing attempts in the system using two different mechanisms.

Acknowledgments

This research was supported by Dr. S Suresh Babu, the head of our institution. We would also like to show our gratitude to the head our institution, Dr. S V Annlin Jeba for sharing her pearls of wisdom with us during the course of the research. We thank our colleagues from Sree Buddha College of Engineering who provided insight and expertise that greatly assisted us although they may not agree with all of the interpretations and conclusions of the paper.

REFERENCES

11.1. Journal Article

[1] T. A. Tang, L. Mhamdi, D. Mclernon, S. A. R. Zaidi, and M. Ghogho, "Deep learning approach for network intrusion detection in software defined networking," in Proc. Int. Conf. Wireless Netw. Mobile Commun., Fes, Morocco, 2018, pp. 258–263.

[2] A. Le, P. Dinh, H. Le, and N. C. Tran, "Flexible network-based intrusion detection and prevention system on software-defined networks," in Proc. Int. Conf. Adv. Comput. Appl., Ho Chi Minh City, Vietnam, 2015, pp. 106–111.

[3] A.-C. Enache and V. Sgârciu, "A feature selection approach implemented with the binary bat algorithm applied for intrusion detection," in Proc. Int. Conf. Telecommun. Signal Process., 2015, pp. 11–15.

[4] J. Kim, J. Kim, H. L. T. Thu, and H. Kim, "Long short term memory recurrent neural network classifier for intrusion detection," in Proc. Int. Conf. Platform Technol. Service, 2016, pp. 1–5.

[5] H. Saxena and V. Richariya, "Intrusion detection in KDD99 dataset using SVM-PSO and feature reduction with information gain," Int. J. Comput. Appl., vol. 98, no. 6, pp. 25–29, 2014.

[6] N. Cleetus and K. A. Dhanya, "Multi-objective functions in particle swarm optimization for intrusion detection," in Proc. Int. Conf. Adv. Comput. Commun. Informat., New Delhi, India, 2014, pp. 387–392

[7] A. S. D. Silva, J. A. Wickboldt, L. Z. Granville, and A. Schaeffer-Filho, "ATLANTIC: A framework for anomaly traffic detection, classification, and mitigation in SDN," in Proc. IEEE/IFIP Netw. Oper. Manag. Symp., Istanbul, Turkey, 2016, pp. 27–35.

[8] Jiaqi Li , Zhifeng Zhao , Rongpeng Li , and Honggang Zhang "AI-Based Two-Stage Intrusion Detection for Software Defined IoT Networks" in IEEE INTERNET OF THINGS JOURNAL, VOL. 6, NO. 2, APRIL 2019

[9] M. Nobakht, V. Sivaraman, and R. Boreli, "A host-based intrusion detection and mitigation framework for smart home IoT using OpenFlow," in Proc. Int. Conf. Availability Rel. Security, Salzburg, Austria, 2016, pp. 147–156.

[10] A.-C. Enache and V. Sgârciu, "A feature selection approach implemented with the binary bat algorithm applied for intrusion detection," in Proc. Int. Conf. Telecommun. Signal Process., 2015, pp. 11–15.