IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ICRTTC '23

International Conference on Recent Trends and Technologies in Computing

Department of Computer Science and Engineering, Arasu Engineering College

Vol. 12, Special Issue 1, April 2023

Prevention and Detection of Botnet Attacks using Double layered machine learning Technique

S. Parvathy [1],S. Mounika[2], M. Nihidha [3], M. Sruthi [4]

Assistant Professor, Department of Computer Science and Engineering, Arasu Engineering College, Kumbakonam, TamilNadu, India[1]

U.G. Student, Department of Computer Science and Engineering, Arasu Engineering College, Kumbakonam, TamilNadu, India[2]

U.G. Student, Department of Computer Science and Engineering, Arasu Engineering College, Kumbakonam, TamilNadu, India[3]

U.G. Student, Department of Computer Science and Engineering, Arasu Engineering College, Kumbakonam, TamilNadu, India[4]

Abstract: In multi level botnet attack in prevailing cyber attacks in the IoT environment starts and ends detection activities. In existing detection of botnet attacks compromising the IoT devices initially performs ddos attacks. According to the various performances of existing machine learning botnet detection model is limited to the trained data which are already specified. The consequences towards the datasets according to the diversified attack patterns that performs perfectly will be questionable. In our proposed methodology the generalized scanning of datasets in DDoS attacks generates 33 varieties of detection patterns. Integration of detecting samples of DDoS at tacks with publicly available datasets within the limit of more attacks. Proposed prevention and detection of double layered machine learning techniques helps in training the dataset models. Prior to the attacking stage the IoT botnet attacks identifies from the trained double layered attack identification and detection models. In the next layer efficiency of datasets detection approach with more accuracy and precise training models will be provided.

Keywords: Botnet Attack, Cyber Attacks, Datasets, Double Layered Machine Learning Techniques, Training Datasets.

I. INTRODUCTION

The Internet of Things (IoT) has revolutionized the way real-world devices communicate with each other over the internet, enhancing human life. The adoption of smart IoT devices such as cameras, wearables, toys, bulbs, and others has increased exponentially in recent years, enabling everyday objects to connect and communicate with each other without human intervention. However, most IoT devices have limited or negligible security features and some even have a fixed key or hard-coded default username and password, which cannot be changed by the user. These security weaknesses make it easy for hackers to exploit insecure IoT devices and gain control over them.

Recent trends show that cyber-attacks are increasing rapidly, particularly with the growth of insecure IoT devices. Botnet and Distributed Denial of Service (DDoS) attacks are the most common and frequent types of attacks that have gained importance over the past decade. A botnet attack involves an attacker scanning a network to identify vulnerable IoT devices, then installing a bot program into them via malware. The bot program connects infected devices to a central server or peer network, which sends commands to them to perform malicious activities such as spamming, flooding DDoS, etc., using thousands of infected IoT devices to target a server, website, or other online resource.

Botnet attacks not only pose a serious threat to insecure IoT devices but also to the entire internet. The Mirai botnet attack in 2016 was a significant turning point for IoT botnet attacks. Since the public disclosure of the Mirai botnet source code, many versions and imitators of the botnet have been developed, infecting millions of IoT devices and launching large-scale and catastrophic DDoS attacks on targets such as GitHub and AWS.

Identification of attacks within the insecure IoT devices through online services helps within its double layered machine learning techniques. In technical terms, online search engine services provide a vast amount of information to attackers who seek to exploit insecure IoT devices. By compromising these devices, an attacker can launch a variety of cyber-attacks such as spamming, phishing, DDoS, etc. that can cause significant damage to other resources on the internet. Recent studies have revealed that IoT devices are particularly vulnerable to botnet and DDoS attacks, with a significant number of DDoS attacks being carried out using compromised IoT devices. Gartner, a leading technology research firm, has also predicted that 25% of all cyber-attacks will result from insecure IoT devices.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ICRTTC '23
International Conference on Recent Trends and Technologies in Computing
Department of Computer Science and Engineering, Arasu Engineering College
Vol. 12, Special Issue 1, April 2023

To address this issue, an effective security system must be implemented to detect IoT bots and secure the insecure IoT devices from becoming part of botnets. Existing botnet and DDoS attack detection methods can be broadly classified into two categories: host-based methods and network-based methods. However, due to the resource constraint nature of IoT devices (limited memory, battery, and compute power), host-based solutions are not practical for IoT devices. Therefore, network-based solutions are a better way to protect IoT devices and networks from these devastating cyber-attacks.

There are three types of intrusion detection approaches used to detect and prevent potential attacks in network traffic: signature-based detection, anomaly-based detection, and specification-based detection. Signature-based detection relies on predefined rules in a database to match network traffic and identify potential attacks. Anomaly-based detection analyzes the normal behavior of network traffic and creates a baseline profile of each device communicating within the network. Any deviation from the baseline is considered an anomaly. Anomaly-based detection is further classified into three subtypes: data-based detection, machine learning-based detection, and knowledge-based detection. Specification-based detection performs intrusion detection based on user-defined specifications or policies.

In this paper at first the introduction has been described. In II section previous works existing and its drawbacks will be discussed. In the III section basic knowledge about botnet attacks and explains the detection and prevention of attacks. In Section IV it represents experimental results showing results of proposed double layered machine learning techniques. Finally, Section V concludes the paper.

II. RELATED WORK

In the literature survey, various techniques have been proposed for botnet attack detection. The current detection techniques can be categorized into two types: graph-based techniques and flow-based techniques. In graph-based techniques, communication nodes in a network are analyzed to detect anomalies that communicate differently compared to the neighbor nodes. Conversely, in flow-based techniques, inbound and outbound traffic data are monitored by machine learning algorithms to detect botnet attacks based on traffic pattern resemblance.

Nguyen et al. proposed a graph-based method to detect IoT botnet attacks using PSI graphs. Similarly, Wang et al. proposed a hybrid model named BotMark, which detects botnet attacks based on flow-based and graph-based network traffic behaviors. Yassin et al. proposed a unique approach to detect Mirai attacks using graph visualization and rules generation.

Almutairi et al. proposed a hybrid botnet detection approach that detects new botnets implemented on three levels: host level, network level, and a combination of both. Blaise et al. proposed a bot detection approach named BotFP, which has two versions: BotFP-Clus those groups similar traffic times using clustering algorithms and BotFP-ML that learns from signatures to detect new bots using supervised ML algorithms. Soe et al. developed a machine learning-based IoT botnet attack detection model that includes two stages: a model builder and an attack detector.

Sriram et al. proposed a deep learning-based IoT botnet attack detection framework that considered network traffic flows transformed into feature data passed to the DNN model. Nugraha et al. evaluated the performance of four deep learning models for botnet attack detection and found that CNN-LSTM outperformed all deep learning models for botnet attacks detection Parra et al. proposed an allotted deep learning framework based on cloud computing for detecting phishing and IoT botnet attacks. Their model includes two machine learning models: (1) a distributed CNN (DCNN) for detecting URL-based attacks directed towards a consumer's IoT devices, and (2) a recurrent neural network (RNN) and an LSTM community version for detecting botnet assaults at the backend.

Pektacs and Acarman completed botnet detection using deep learning on network drift traffic. Their proposed deep neural network was deployed to categorize the site visitors as benign or malicious. To improve the overall performance of the model, hidden layers and neurons were investigated. The proposed model achieved 99% accuracy. Similarly, Ahmed et al. proposed a deep learning model for detecting botnet attacks.

Maeda et al. proposed a botnet attack detection method using deep learning on software-defined network (SDN). The authors trained their deep learning model using data collected on flow-based traffic from the botnets and evaluated its detection accuracy. The authors also used a multi-layer perceptron (MLP), a deep learning model, to detect infected IoT devices.

Meidan et al. developed a novel IoT botnet attack detection method using deep auto-encoders. For the malicious network traffic, they infected nine IoT devices with two IoT botnets, Mirai and BASHLITE, and trained deep auto-encoders separately for each IoT device on both benign and attack traffic.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ICRTTC '23
International Conference on Recent Trends and Technologies in Computing
Department of Computer Science and Engineering, Arasu Engineering College
Vol. 12, Special Issue 1, April 2023

Bovenzi et al. proposed a hybrid two-level intrusion detection system (IDS) for the IoT environment. Their proposed method first detects anomalies from the network traffic, and in the second stage, classifies the anomalies into attack categories. The authors used a multi-modal deep auto-encoder for anomalies detection, while using three machine learning classifiers to classify anomalies detected in the first stage.

Mirsky et al. also used auto-encoders and proposed a plug-and-play network IDS called Kitsune to detect anomalies on local network traffic using an unsupervised learning approach. The authors used a self-generated botnet attack dataset and evaluated the performance in both online and offline modes. Their proposed solution achieved good performance similar to offline anomaly detectors.

III. METHODOLOGY

The current botnet detection methods use traffic flow-based machine learning methods for botnet attacks detection. Moreover, most of these solutions do not perform well on different datasets due to the diversity of attack patterns. However, these present methods only detect the botnet attack after the IoT devices get compromised and start performing malicious activities like DDoS attacks. To prevent the IoT devices from being compromised, this study proposes a double layered machine learning approach that detects the attacker's malicious activities at the early stage (i.e., scanning) of a botnet attack. Additionally, the proposed double layered approach has ability to detect and prevent botnet attacks compromised with IoT device which initiates detection of malicious attacking activities.

The proposed methodology describes a standard botnet, which comprises four key components: the bot application, zombie tool, bot-grasp, and command and control (controlling) server. The botnet attack is initiated by the bot-master, which can be an automated program created by the attacker or the attacker themselves. The bot-master scans iot devices connected to the internet and exploits the vulnerable ones by installing a bot program on them. The bot program establishes a connection with the bot-grasp or controlling server to receive instructions for carrying out malicious activities.

The bot program is a malware installed on infected devices that resides within the victim iot device and connects with the controlling server or bot-master to receive instructions for malicious activities, such as spamming or performing flooding attacks. A physical victim device, on which the bot program is installed, is known as a zombie tool, which can include smart cameras, tvs, wearables, and so on.

The bot-grasp or bot header is the primary controller of the botnet, acting as the main operator that issues commands to the controlling server (in client-server architecture) or specific bots (in peer-to-peer architecture). The bot-master is responsible for organizing botnet attacks, and it works as the primary operator of the botnet, issuing commands to the controlling server or individual bots.

The controlling server is the central computer that controls the zombie devices based on the control signals received from the bot-master. However, it is not a mandatory component of every botnet. In a peer-to-peer network, the bot-master directly sends control signals to zombies without the need for a controlling server.

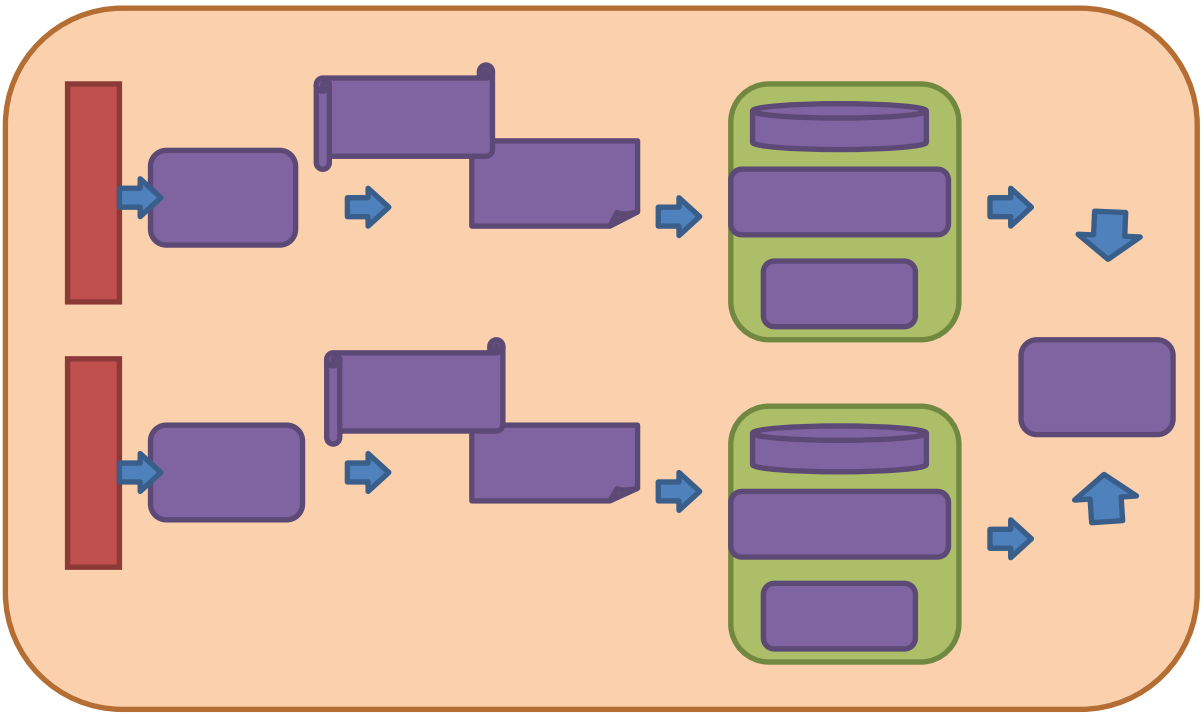
The botnet attack is a complex attack that involves multiple stages, as defined by a botnet life cycle, in order to transform a vulnerable IoT device into a bot capable of carrying out malicious activities such as sending spams or performing DDoS attacks. These stages are scanning, malware injection, botnet connection, command execution, and maintenance & up-gradation. During the scanning stage, the attacker gathers information about the target device, which is then used to exploit its vulnerabilities and inject malware. Once the malware is injected, it establishes a connection with the bot-master and awaits commands. Finally, the bot-master ensures that the infected devices are maintained and upgraded so that they remain undetected and available for future attacks. Updating and maintaining the bot program is crucial for the attacker to ensure the long-term success of the attack.

This methodology presents a novel approach to prevent and detect botnet attacks in IoT networks. The methodology comprises two main folds. In the first fold, a ResNet-18 model is trained to detect scanning activities, which represent the pre-attack stage of a botnet attack. The goal is to prevent an attacker from proceeding to further attack stages by detecting the scanning activity. In the second fold, another ResNet-18 model is trained to detect DDoS attacks, which are carried out after compromising weakly-secured IoT devices. The proposed methodology detects inbound and outbound traffic for both scanning and DDoS attacks.

The proposed methodology involves main stages. In the network traffic is captured in the format and used to train the machine learning models. In the second stage, network flows are extracted from the files, features are extracted from the flows, and the dataset is labeled with 'normal', 'scan', and 'DDoS' labels. In the third stage, a Logistic Regression function selection method is

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940
International Journal of Advanced Research in Computer and Communication Engineering
ICRTTC '23
International Conference on Recent Trends and Technologies in Computing
Department of Computer Science and Engineering, Arasu Engineering College
Vol. 12, Special Issue 1, April 2023

used to optimize the performance of the machine learning model by selecting the most effective features. The LR feature selection method is chosen due to its efficient performance in the existing literature.



IV. EXPERIMENTAL RESULTS

This study analyzed eleven different scanning techniques that attackers often use to gather information about vulnerable IoT devices during the initial stages of an attack. These scanning techniques include SYN scan, FIN scan, ACK scan, NULL scan, SYN-ACK scan, FIN-ACK scan, yuletide scan, UDP scan, TCP Window scan, TCP connect scan, and Banner grabbing. The scans were performed using three widely used scanning tools, namely Nmap, Hping3, and Dmitry, on two lab servers using three different scanning methods, namely horizontal scan, vertical scan, and box scan. A total of 33 types of scanning attacks were conducted to generate and collect scanning traffic. The experiment was performed on an Ubuntu system with a core i7 processor and 8 GB RAM, and the network packets were captured using Wireshark tool and saved as .pcap files. The dataset generated from these experiments is called the ScanLab dataset.

In addition to the self-generated dataset, this study also used three publicly available datasets, namely CICIDS-19, CICIDS-17, and Bot-IoT dataset, to evaluate the performance of the ResNet-18 model for scanning attack detection. The experiment and normal traffic samples from these datasets were obtained for comparison purposes.

The data preprocessing step involved extracting features from the captured .pcap files of the ScanLab dataset using the CICFlowmeter tool, which identified flows based on 5-tuple information, including source IP, destination IP, source port, destination port, and protocol. More than 60 flow features were extracted for each flow, and the resulting data was saved as a .csv file. The data was unlabeled, so the IP addresses used for scanning were labeled as attack, while the remaining network traffic was labeled as normal.

The CICIDS-19, CICIDS-17, and Bot-IoT datasets were also pre-processed and labeled with respect to their descriptions. The experiment attack samples were partially integrated with the ScanLab dataset by randomly selecting 50K samples from these three datasets to provide maximum attack coverage and improve the performance of machine learning algorithms.

After selecting the useful features, we split each dataset into the train, validation, and test set. For this purpose, we randomly selected 60% data for training, 20% data for validation and 20% data for testing to avoid over fitting and for efficiently training the ML model. Both the training set and validation set are used during the training phase. The training set is used to train the machine learning model. In order to efficiently train and better optimize the weights of an ML model, we validate the trained model on the validation set after each epoch based on which the optimizer algorithm updates the weights of the ML model. Finally,

when the ML model completes its training, we test its performance over unseen data, i.e., test set. As mentioned earlier that we used the ResNet-18 model and first trained it over the train set of the ScanLab dataset. Originally, the ResNet-18 model is designed to classify the image processing and computer vision problems which consist of images, i.e., high dimensional arrays. So, before starting the training, we need to convert the data into high dimensional arrays since the ResNet-18 model is prone to overfit at low dimension data.

V. CONCLUSION

In this article, we proposed a double layered machine learning approach for detecting and preventing IoT botnet attacks. The first layer involves training a deep modern version ResNet-18, called ResNetScan-1, for scanning attack detection. The second layer involves training another ResNet-18 version, called ResNetDDoS-1, for detecting DDoS attacks in case the scanning detection model fails to prevent a botnet attack. To validate the performance of our proposed models, we conducted several experiments on three publicly available datasets. The experimental results showed that the proposed ResNetScan-1 and ResNetDDoS-1 models outperformed all other models for detecting test and DDoS attacks. Therefore, our proposed double layered approach is efficient and robust in preventing and detecting IoT botnet attacks with a wide range of attack patterns coverage. However, the current work only covers a limited number of scanning and DDoS attacks, and future work aims to cover more attack techniques to train the proposed framework for more efficient prevention and detection of IoT botnet and DDoS attacks. Moreover, we plan to deploy our proposed double layered approach in IDS to investigate its effectiveness on live network traffic.

REFERENCES

[1] I. Ali, A. I. A. Ahmed, A. Almogren, M. A. Raza, S. A. Shah, A. Khan, and A. Gani, ‘‘Systematic literature review on IoT-based botnet attack,’’ IEEE Access, vol. 8, pp. 212220–212232, 2020.
[2] S. Ghazanfar, F. Hussain, A. U. Rehman, U. U. Fayyaz, F. Shahzad, and G. A. Shah, ‘‘IoT-Flock: An open-source framework for IoT traffic generation,’’ in Proc. Int. Conf. Emerg. Trends Smart Technol. (ICETST), Mar. 2020, pp. 1–6.
[3] M. Safaei Pour, A. Mangino, K. Friday, M. Rathbun, E. Bou-Harb, F. Iqbal, S. Samtani, J. Crichigno, and N. Ghani, ‘‘On data-driven curation, learning, and analysis for inferring evolving Internet-of-Things (IoT) botnets in the wild,’’ Comput. Secur., vol. 91, Apr. 2020, Art. no. 101707.
[4] F. Hussain, S. G. Abbas, M. Husnain, U. U. Fayyaz, F. Shahzad, and G. A. Shah, ‘‘IoT DoS and DDoS attack detection using ResNet,’’ in Proc. IEEE 23rd Int. Multitopic Conf. (INMIC), Nov. 2020, pp. 1–6.
[5] S. Dange and M. Chatterjee, ‘‘IoT botnet: The largest threat to the IoT network,’’ in Data Communication and Networks. Singapore: Springer, 2020, pp. 137–157.
[6] F. Hussain, S. G. Abbas, U. U. Fayyaz, G. A. Shah, A. Toqeer, and A. Ali, ‘‘Towards a universal features set for IoT botnet attacks detection,’’ in Proc. IEEE 23rd Int. Multitopic Conf. (INMIC), Nov. 2020, pp. 1–6.
[7] A. O. Prokofiev, Y. S. Smirnova, and V. A. Surov, ‘‘A method to detect Internet of Things botnets,’’ in Proc. IEEE Conf. Russian Young Res. Electr. Electron. Eng. (EIConRus), Jan. 2018, pp. 105–108.
[8] B. K. Dedeturk and B. Akay, ‘‘Spam filtering using a logistic regression model trained by an artificial bee colony algorithm,’’ Appl. Soft Comput., vol. 91, Jun. 2020, Art. no. 106229.
[9] N. Vlajic and D. Zhou, ‘‘IoT as a land of opportunity for DDoS hackers,’’ Computer, vol. 51, no. 7, pp. 26–34, 2018.
[10] GitHub Survived Biggest DDoS Attack Ever Recorded. Accessed: May 3, 2021. [Online]. Available: https://github.blog/2018-03-01-ddosincident-report/
[11] AWS Said it Mitigated a 2.3 Tbps DDoS Attack, Largest Ever. Accessed: May 3, 2021. [Online]. Available: https://www.zdnet.com/article/awssaid-it-mitigated-a-2-3-tbps-ddos-attack-the-largest-ever/