International Journal of Scientific Research in Computer Science, Engineering and Information Technology
ISSN : 2456-3307 (www.ijsrcseit.com)
doi : https://doi.org/10.32628/IJSRCSEIT

Two-Fold Machine Learning Approach to Prevent and Detect IoT Botnet Attacks

*[1] P Ganesh Kumar,[2] V. Nikhitha, [3] P. Mounika

[1] Assistant Professor, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India

[2,3] Students, Department of Information Technology, Bhoj Reddy Engineering College for Women, Hyderabad, India

ARTICLEINFO

Article History:

Accepted: 10 April 2023
Published: 30 April 2023

Publication Issue
Volume 9, Issue 2
March-April-2023

Page Number
661-665

ABSTRACT

A botnet is a malware that degrades the functionality as well as access to a healthy computer system through malware programs. Botnet programs perform DDoS attack, Spam, phishing attacks. Botnet attack takes place in two ways which are peer to peer attacks and command and control attack. The peer-to-peer attack takes place to by passing botnet attacks from one system to another in a peer-to-peer network while the command-and-control attack takes place by a botmaster attack on a server which uses various transactions in exchange with systems on the network and those nodes in the networks function as slaves. The report presents a survey of various techniques of botnet detection models built using several types of machine learning techniques. The report gives the review on various methodologies involved in Botnet Detection and to identify the best methods involved to understand various dataset. We also surveyed on how classification, clustering is used in detection of Botnet to improve the accuracy of the model.

Keywords: Command and Control Botnet, Peer to Peer Botnet, Network Security, Machine learning, Network Protocols, Cyberattacks, Clustering, Classification, Deep Learning

I. INTRODUCTION

The Botnet refers to devices that a hacker control remotely. Botnet is a combined term of robot's interaction with network, where there are two important participants the Botmaster and the Bot slave. The Bot slave acts as a slave of the Botmaster and does what Botmaster asks to do. The Botnet's task is to start attacks by giving instructions from the botmasters to the bot clients to function as slaves to the botmaster. Nowadays botnet attack takes place so silently that the Anti Malware software are not at all able to detect it. The botnet attack taking place in peer-to-peer networks have become a challenge as detecting the centre of command is not so easy. But on a general note, tough it is tough to determine botnet command-and-control attacks, it is possible to observe patterns in data to get a complete picture of

network data exchanges and detection of Botmaster is possible. In DDoS attack, the attacker who is the botmaster has high-end computing Systems and servers to run command & control malware programs which instructs the machines in the next layer or level called handlers. These handlers attack the clients called making them the bot slave. Botnet malicious activities is detected using a variety of methods. As known from the above information the malware detection software finds it incredibly difficult to detect these attacks. The typical approach can be by analysis of network traffic data obtained by simulation and of botnet on Virtual machines and obtaining suitable communication and TCP and UDP protocol network exchange data. The Supervised Learning algorithms (ex: Decision trees, Support Vector machines (SVM)) can is effective in classifying normal traffic from botnet traffic. When Unsupervised learning algorithms like the K-means algorithm integrated with classification algorithms, the outcomes get improved. Multilayer deep learning Neural networks is also a better way to approach large volume of network traffic data. It gives us better chances in identifying other different patterns in data apart from machine learning algorithm. General datasets observed to be used for such type of analysis are: CTU-13 Dataset, KDD cup nineteen dataset, UNSW-NB15 Dataset and Bot-IoT dataset. There are many more datasets developed or used depending on the researcher's goal. The selection of the dataset is a particularly important criterion to undergo a good accuracy, stable model to build a better Botnet Detection system. The feature selection process is an important aspect of every Machine Learning model. Selecting proper features according to the information needed depends upon the major agenda of the Researcher that aims to detect and analysed. In the event of Command and Control (C&C) botnet attacks, changing IP address data, various times and techniques of TCP data exchange may be a crucial feature to take hold of the Botmaster. In this way majority of the datasets are processed and features are selected as needed. The major step of botnet detection is selection of suitable techniques in which accuracy of the model is improved and the misclassification of data can be prevented. Many researchers have described various methods to detect various kinds of botnets. The different approaches used in different research work is describe further in this paper in other sections.

II. RELATED WORK

The General view of methods using machine learning regarding detection of Botnet Malicious activity is shown below in the Figure 1.

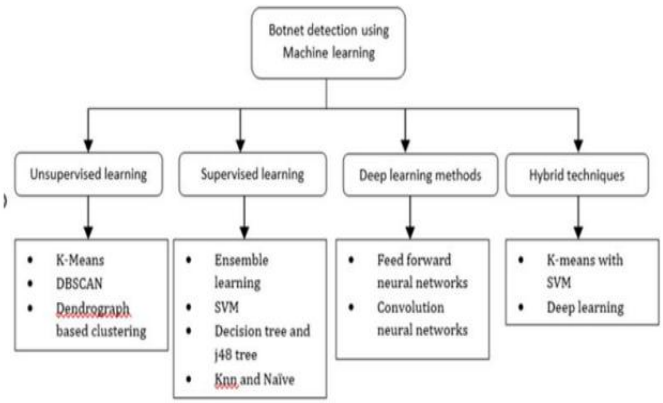

Fig.1 Outline of methodologies used in Botnet Detection Using Machine Learning

The use of BARCA framework [1] has been seen to use feedback-based approach than using historical data, considering metrics correlations, and it is designed to adapt to changes in the behaviour of the system using an incremental and iterative process. BARCA [1] has three main components which are Behaviour extractor (BE), Behaviour Identifier (BI) and a Feedback Provider (FP). The Behaviour extractor (BE) extracts and collects the performance information of the System periodically and creates a Behaviour Instance whenever new periodic data gets collected. Behaviour Instances are representations of system behaviour obtained timely. The Behaviour

Identifier (BI) uses a Behaviour model to classify each Behaviour Instance as normal or abnormal. When the data is classified abnormal, the Behaviour Identifier feedbacks to the Feedback Provider (FP) [1]. The Feedback Provider uses the information from the Behaviour Model to alert the administrator of the system in scan or the user, which acts accordingly to prevent the activity and consistently reports current states of system when alerts are received. The Behaviour Identifier (BI) uses the combination of single class classifier and multiple binary classifiers to be able to reach the main goals which are: the one class classifier detects the set of anomalies while the various binary classifiers are used to detect anomalies that are known. The Behaviour model is built using Gradient Descent and Support Vector Machines. BARCA [1] uses the user supplied feedback to build its behavioural model. BARCA's advantage is that it does not use Historical Data and hence can dynamically change the system's behaviour accordingly. The next method found was Twofold approach to build Botnet detection models [2]. The Command-and-Control botnet network (C&C) data capture is the basic method used. Botnet detection is conducted using a new two-fold method The figure 2 shows the Two-fold approach for botnet detection.

III. PROPOSED SYSTEM

In this research, CHRISP-DM was used. The data mining research methodology is mainly used for achieving the research objectives. However, it will be very beneficial to be used in any machine learning project with some very logical steps that could cover and assist almost any project without any regard to its nature. Data mining research methodology is the abbreviation of the Cross-industry standard process for data mining which refers to the process model that gives a system to the carrying-out data mining project. The data mining research methodology is meant to do large mining projects, more reliable, less expensive, more repeatable, quicker, and more achievable. The development model that is utilized for this research is the data mining research methodology reference model. This model was picked in light of the fact that it outlines the project life cycle. Fig. 2 shows the research methodology, consisting of four main steps: collection, preparation, feeding the data to the classifier, and evaluating the results.

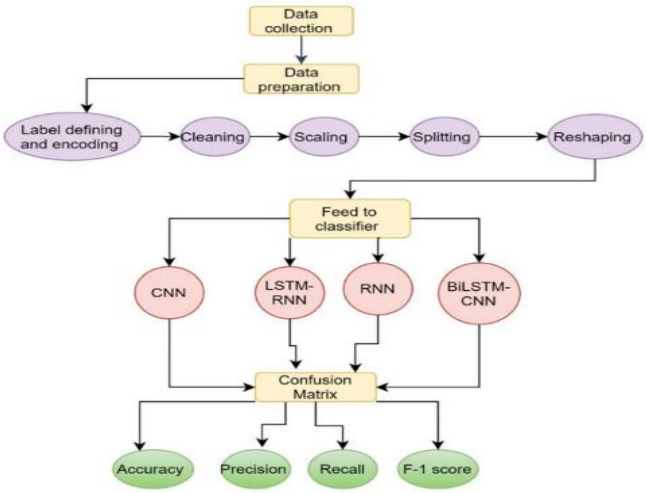

Fig. 2 Research Methodology

IV. CONCLUSION

As a result of the attackers' use of spoofing methods, DDoS assaults pose a serious danger to everyone utilizing the Internet. Based on DDoS history, we observe that "Mirai" and "Bashlite" will not be the last and most powerful botnets as they evolved and bypassed the old methods for mitigation like adding more bandwidth or doing traffic extraction. Cloud computing is still the future. IoT-based DDoS attacks are threatening it. As IoT devices are fundamentally insecure yet, some of them show high resistance, but the rest do not. So, we cannot just leave for the goodwill of the manufacturer. The deep learning approach we chose has proven very worthy, giving an average accuracy of 0.896975, which indicates thatit is a solid way to deal with DDoS attacks no matter what way they use. As observed from fig 5, BiLSTM-CNN has proven to be a great combination acquiring the

highest accuracy, recall, precision, and F-1 score, and the lowest error rate followed by RNN has achieved an accuracy of 0.8977, and the error rate amounted to 0.1576. LSTM follows it with an accuracy of 0.8971. However, LSTM has achieved the highest error rate compared to the other classifiers. Although CNN has achieved the lowest accuracy rate of 0.895, it did achieve the second-lowest error rate of 0.1685. The usage of such a huge dataset and the excellent outcomes that we have obtained is an obviously great demonstration of the capacity to eliminate DDoS threats using deep learning. With more work and more innovative methods to use AI, we believe that we could have a solid opportunity to halt DDoS attacks once and for all.

V. REFERENCES

[1]. J. A. Cid-Fuentes, C. Szabo, and K. Falkner, "Adaptive performance anomaly detection in distributed systems using online SVMs," IEEE Trans. Dependable Secure Computer., vol. 17, no. 5, pp. 928–941, Sep./Oct. 2018

[2]. F. Hussain et al., "A Two-Fold Machine Learning Approach to Prevent and Detect IoT Botnet Attacks," in IEEE Access, vol. 9, pp. 163412-163430, 2021, doi: 10.1109/ACCESS.2021.3131014.

[3]. Y. Zhang, X. Chen, D. Guo, M. Song, Y. Teng, and X. Wang, "PCCN: Parallel cross convolutional neural network for abnormal network traffic flows detection in multiclass imbalanced network traffic flows," IEEE Access, vol. 7, pp. 119904–119916, 2019.

[4]. A. Esfahan and D. L. Bhaskari, "Intrusion detection using random forests classifier with SMOTE and feature reduction," in Proc. Int. Conf. Cloud Ubiquitous Computer. Emerg. Technol., Nov. 2013, pp. 127–13

[5]. T. Trajanovski and N. Zhang, "An Automated and Comprehensive Framework for IoT Botnet Detection and Analysis (IoT-BDA)," in IEEE Access, vol. 9, pp. 124360- 124383, 2021, doi: 10.1109/ACCESS.2021.3110188.

[6]. K. Alieyan, A. ALmomani, A. Manasrah, and M. M. Kadhum, "A survey of botnet detection based on DNS," Neural Comput. Appl., vol. 28, no. 7, pp.1541–1558, Jul. 2017

[7]. S. Haq and Y. Singh, "Botnet Detection using Machine Learning," 2018 Fifth International Conference on Parallel, Distributed and Grid Computing (PDGC), 2018, pp. 240-245, doi: 10.1109/PDGC.2018.8745912

[8]. D. Zhuang and J. M. Chang, "Enhanced PeerHunter: Detecting peer-to-peer botnets through network-flow level community behaviour analysis," IEEE Trans. Inf. Forensics Security, vol. 14, no. 6, pp. 1485–1500, Jun. 2019

[9]. X. D. Hoang, "Botnet detection based on machine learning techniques using DNS query data," Future Internet, vol. 10, no. 5, pp. 1–11, 2018.

[10]. R. Chen, W. Niu, X. Zhang, Z. Zhuo, and F. Lv, "An effective conversation-based botnet detection method," Math. Problems Eng., vol. 2017, pp. 1–9, Apr. 2017.

[11]. D. Zhao, I. Traore, B. Sayed, W. Lu, S. Saad, A. Ghorbani, and D. Garant, "Botnet detection based on traffic behavior analysis and flow intervals," Comput. Secur., vol. 39, pp. 2–16, Nov. 2013.

[12]. Z. Bazrafshan, H. Hashemi, S. M. H. Fard, and A. Hamzeh, "A survey on heuristic malware detection techniques," in Proc. 5th Conf. Inf. Knowl. Technol., May 2013, pp. 113–120.

[13]. J. A. Caicedo-Muñoz, A. L. Espino, J. C. Corrales, and A. Rendón, "QoSclassifier for VPN and non-VPN traffic based on time-related features," Comput.Netw., vol. 144, pp. 271–279, Oct. 2018.

[14]. R. Rapuzzi and M. Repetto, "Building situational awareness for network threats in fog/edge computing: Emerging paradigms beyond the security perimeter model," Future Gener. Comput. Syst., vol. 85, pp. 235–249, Aug. 2018

[15]. P. Sun, J. Li, M. Z. A. Bhuiyan, L. Wang, and B. Li, "Modeling and clustering attacker activities in IoT through machine learning techniques," Inf. Sci., vol. 479, pp. 456–471, Apr. 2019.

Cite this article as :

P Ganesh Kumar, V. Nikitha, P. Mounika, "Two-Fold Machine Learning Approach to Prevent and Detect IoT Botnet Attacks", International Journal of Scientific Research in Computer Science, Engineering and

Information Technology (IJSRCSEIT), ISSN : 2456-3307,
Volume 9, Issue 2, pp.689-692, March-April-2023.