International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 07 Issue: 04 | April - 2023 Impact Factor: 8.176 ISSN: 2582-3930

Botnet Detection Based on Machine Learning Techniques in P2P Networks

Pavana K P[1], Rohith Adiga H R[2], Shubha M L [3], Vinayaka Patil K G[4], Mohan H G[5]

[1,2,3,4] *Research Scholars, Dept. of CS&E, JNNCE, Shimoga, India*

[5] *Assistant Prof., Dept. of CS&E, JNNCE, Shimoga, India*

***Abstract*— A botnet is a network of computers that are controlled from a botmaster or a command and-control server. Botnet is a major threat on the internet. P2P botnet is a representative of P2P malicious programs. Botmaster gives a command and control (C&C) information via a unique communication channel. It remotely controls the bots that are compromised to initiate malicious activities like distributed denial of service (DDoS) attack, spamming, phishing, and sensitive information stealing. The approaches using Machine learning are used in botnet detection. They are useful to extract unexpected patterns from traffic. In this paper some of the possible technical solutions proposed by researchers are reviewed.**

***Keywords*— Botnet, malicious programs, botnet detection**

I. INTRODUCTION

The words "robot" and "network" together give rise to the word Botnet. Botnet refers to a network of hijacked internet-connected devices that are installed with malicious programs which are known as malware. Each of these infected devices is known as Bots, and a hacker/cybercriminal known as the "Bot herder" remotely controls them. A bot is also called a zombie, and a botnet is referred to as a zombie army. Botnets pose serious threats. They make a volume of infected systems to engage in activities like DDoS attacks, spreading spams. They do this in order to steal user sensitive information. They also carry distributed computing tasks for illegal purposes. Bots are remotely managed through command and control (C&C) channel and target various network topologies. Current botnets target P2P networks for higher connection resilience. The botmaster choses a bot to distribute commands to whole botnet. A P2P botnet is difficult to detect because of its elusive nature. Signature based, anomaly-based, DNS based, and machine learning based are a few categories of detection methods.

II. RELATED WORK

Zainab Alothman et al. [1] describes in detail about the attacks that penetrate and select susceptible IoT devices to build a vast network of "Zombies" that can launch malicious attacks against other internet resources while being distantly controlled by a "Bot master." The Synthetic Minority Over-Sampling Technique (SMOTE) was used. Bot-IoT dataset with instances of main and sub-attack categories. Furthermore, the performance of the top three classifiers—the J48, Random Forest (RF), and Multilayer Perceptron (MLP) networks—was evaluated after testing numerous classifiers. The outcomes demonstrated the supremacy of the R F and J48 classifications over MLP networks and other cutting-edge technologies. The best binary classifier described in this research had an accuracy of 0.999, while the best classifications for the primary assault and subcategories had accuracy values of 0.96 and 0.93, respectively.

Shao-Chien Chen et al. [2] suggested the four different types of botnet identification techniques: machine learning-based signature-based, DNS-based, anomaly-based. The experimental findings indicate that 94.7 % accuracy can be achieved using convolutional features 2.2% false positive rate on the well-known botnets ,CTU P2P and PeerRush. Additionally, it assesses and

ranks the relative effectiveness of the flow-based characteristics used in current botnet identification methods. Rajesh Kalakoti et al. [3] discussed about Channel-based characteristics favoured by selecting features for detection at the communication, post-attack, and control stages, even though host-based features are more useful at distinguishing attacks from bots. In order to enhance prediction results by removing unused features, feature selection aims to isolate the finest groups of features from the incoming data. The RFE starts with a complete collection of features and then eliminates the least significant ones one at a time to determine which features are most essential. Based on a feature-ranking criteria characteristic, this approach is used. A small number of features are chosen using the filters technique, and the models are given the findings. Wrappers methods categorise the group of features using those models. To ensure accuracy, findings from wrapper techniques are validated using the four major machine learning algorithms (Decision Trees, Random Forest, Extra Tree Classifier, and K-Nearest Neighbors). The ideal algorithm and wrapper technique are discovered.

Beny Nugraha et al. [4] discussed about the success of four distinct deep learning models—Convolutional Neural Network (CNN), Long Short-Term Memory (LSTM), CNN-LSTM hybrid, and Multi-Layer Perception—was suggested by the authors. (MLP). They are used for identifying both recognised and unrecognised malware traffic patterns. The CTU-13 dataset, an actual botnet traffic dataset produced in the Czech Republic in 2011 at the CTU University, is used to train, verify, and evaluate the models. Each model's performance is measured using a variety of performance measures, including accuracy, sensitivity, specificity, precision, and F1 score, for the classification of both known and unidentified (zero-day)network traffic patterns.

Mustafa Alshamkhany et al.. [5] One of the algorithms with the greatest promise for identifying P2P botnets is the decision tree classifier. According to the findings, even though the random forest classifier resuted in good accuracy, the random tree classifier was deemed the best as it was both accurate and fast. The algorithms used were Adaboost, NB and Bagging. The results showed excellent performance, with 90% or higher accuracy rates across the board for all classifications and barely perceptible differences between them.

Bin Zhou et al. [6] Based on various traffic statistical characteristics, a two-stage identification technique is suggested to identify P2P bots in monitoring networks. Machine learning techniques are used in the first step to identify hosts that engage in P2P communication within the network being monitored using three characteristics of traffic data. In the second step, two traffic data characteristics are suggested to further identify infected P2P bots among these P2P servers in accordance with the distinctive network behaviour patterns of P2P bots. This technique has a detection rate of 99.7% and an FPR of just 0.3%, and it can identify P2P bots in tracking networks in just 5 minutes.

Belal Ibrahim Hairab et al. [7] trained a model with data which identifies the DoS attack and normal traffic data. Scenarios are used to test the accuracy of the trained model. In 1st Scenario, Testing classifiers are used on DDoS attack data. Testing classifiers on OS Fingerprint attack data is used in 2nd, and Testing classifiers on Service Scan attack data is applied in the 3rd. The results of are compared by evaluating the results of classifiers: LR, NB, AdaBoost, standard CNN, CNN (L1 and L2 regularised).

Ying Xing et al. [8] presented Peerhunter advised host-level community behaviour analysis to spot P2P botnets, but this research did not account for the chance that P2P botnets and legitimate P2P apps might coexist on the same group of hosts. The framework has a 99.2% identification success rate and can spot P2P bots even when there is legitimate P2P data. Peerclean outlined a method that combines dynamic group behaviour analysis and machine learning to detect P2P malwarein network data.

Shamsul Haq et al. [9] adapted low positive rates and high positive rates for evaluating accuracy based on the "correctly and incorrectly instance percentage" using two random dataset partitions. Their sum is equal to the original dataset. They mean accuracy of k-means clustering and j48 classification. The results of classification and clustering are lower in case of classification and higher in case of clustering. At another extreme the results are variable in nature and gives the approximation in both processes.

Di Zhuang et al.. [10] To find P2P botnets, the method Enhanced PeerHunter, which is based on community activity analysis, is suggested. From a specified MCG, this component seeks to identify P2P malware. Uses a community detection technique to group the servers and bots into their respective communities first. Following that it uses a numerical community behaviour research to find botnet groups. Finally, further identifies or validates each potential bot using structural community behaviour analysis. The testing findings demonstrated that Enhanced PeerHunter is highly resilient against the suggested assaults and can achieve high detection rates with few false positives.

Table-1: Summary of survey

AUTHOR	DATASET	MACHINE LEARNING ALGORITHM USED	ACCURACY
Beny Nugraha et.al. [4].	CTU-13	Convolutional Neural Network (CNN), Multi- Layer Perception (MLP), hybrid CNN-LSTM, Long Short-Term Memory (LSTM)	99.001%
Bin Zhou et. at[6].	P2P dataset,Non-P2P dataset , 2P botnet dataset	SVM, Bayesian network, random forest and(J48) decision tree	99.7%
Belal Ibrahim Hairab et al[7].	Bot-IoT dataset	LR, NB, AdaBoost, standard CNN, CNN (L1and L2 regularised).	91%
Rajesh Kalakotiet al.[3].	N-BaIoT, Med-BIoT	Extra tree classifier, Decision tree, k-NN, Random forest ,Pearson's correlation, Mutual Information, Fisher Score, Anova F-Test, Recursive Feature Elimination, Sequential Forward and Backward Selection.	99.57%
Shao-Chien Chenet al.[2].	PeerRush, CTU	Convolutional neural Network (ANN),Decision Tree	94.7%
Zainab Alothmanet al.[1].	Bot-IoT dataset	(J48) decision tree, Random Forest (RF), and MultiLayer Perceptron (MLP) neural network.	96%
Mustafa Alshamkhany et.al[5]	Bot-IoT, (UNSW) datasets	Naïve Bayes, Decision Trees ,K-NearestNeighbour, Support Vector Machine.	99.89%
Di Zhuang et.al[10].	24 hours network traces of (a) 4 popular P2P applications, (b) 5 P2P botnets, And Trans-Pacific backbone between the USA and Japan.	Enhanced PeerHunter	100%
Ying Xing et. al.[8].	Sality, CTU datasets, Kelihos,ZeroAccess.	SAW algorithm combined with PCA	99.8%

Conclusion

A botnet is a network of computers infested by bots controlled by a command-and control server by several channels. The bot is a program that perform tasks as commanded by the bot master Botnet attacks have caused major damages in the form of DDoS attacks the papers from the literature survey contained different solutions to this recurring problem. From the survey, it can be concluded that it is best to create a classifying model with possible optimizations in order to improve performance in Recall, Accuracy, Precision and F score.

REFERENCES

[1] Zainab Alothman, Mouhammd Alkasassbeh, and Sherenaz Al-Haj Baddar., "An efficient approach to detect IoT botnet attacksusing machine learning". Journal of. High Speed Netw. Vol. 26, Issue 3, pp.241–254,2020

[2] Chen, Shao-Chien & Chen, Yi-Ruei & Tzeng, Wen-Guey., "Effective Botnet Detection Through Neural Networks onConvolutional Features", 12th IEEE International Conference on Big Data Science and Engineering, pp. 372-378, 2018.

[3] R. Kalakoti, S. Nõmm and H. Bahsi, "In-Depth Feature Selection for the Statistical Machine Learning-Based Botnet Detectionin IoT Networks," in IEEE Access, vol. 10, pp. 94518-94535, 2022.

[4] B. Nugraha, A. Nambiar and T. Bauschert, "Performance Evaluation of Botnet Detecton using Deep Learning Techniques," 11thInternational Conference on Network of the Future (NoF), pp. 141-149, 2020

[5] M. Alshamkhany, W. Alshamkhany, M. Mansour, M. Khan, S. Dhou and F. Aloul, "Botnet Attack Detection using MachineLearning," 14th International Conference on Innovations in Information Technology (IIT), pp. 203-208, 2020.

[6] Zhou, Bin, Jie He and Ming-Che Tan. "A Two-stage P2P Botnet Detection Method Based on Statistical Features.", IEEE 11thInternational Conference on Software Engineering and Service Science (ICSESS): pp. 497-502,2020.

[7] B. I. Hairab, M. Said Elsayed, A. D. Jurcut and M. A. Azer, "Anomaly Detection Based on CNN and Regularization TechniquesAgainst Zero-Day Attacks in IoT Networks," in IEEE Access, vol. 10, pp. 98427-98440, 2022

[8] Xing, Y., Shu, H., Kang, F., & Zhao. Peertrap: An Unstructured P2P Botnet Detection Framework Based on SAW CommunityDiscovery. Wireless Communications and Mobile Computing, 2022

[9] Haq, Shamsul, and Yashwant Singh. "Botnet detection using machine learning." Fifth International Conference on Parallel,Distributed and Grid Computing (PDGC), pp. 240-245. IEEE, 2018.

[10] D. Zhuang and J. M. Chang, "PeerHunter: Detecting peer-to-peer botnets through community behavior analysis," IEEEConference on Dependable and Secure Computing, pp. 493-500, 2017