

A Survey on HTTP BotNet Detection Techniques using Machine Learning

Prof. Chaitali Bhalerao¹, Riddhika Kulkarni², Swapnil Bobade³, Krutika Bharekar⁴, Vaishnavi Zadbuke⁵

¹(Professor, Zeal College of Engineering & Research, Pune, India, Maharashtra)

²³⁴⁵(Students, Zeal College of Engineering & Research, Pune, India, Maharashtra)

Abstract: A botnet is a group of computers linked to the Internet which have been compromised and are being controlled remotely by the botmaster through malicious software called bots.

While substantial research work has been accomplished on botnet analysis and detection, many challenges remain unaddressed, such as the ability to design detectors that can cope with new forms of botnets. So there is a need for an advanced system that can detect traffic behavior accurately. This paper provides the state-of-art techniques used by the researchers for the detection of a botnet attack.

I. Introduction

Distributed denial of service (DDoS) attacks [1][2] occur when a large number of compromised systems flood requests to one or more web servers in the distributed environment. The DDoS attack causes heavy traffic that comes in bursts and prevents the system under attack from providing service to legitimate users. A botnet is a group of compromised computers also called bots or zombies [1] which are controlled by the botmaster's malicious code. Botnets have become one of the most malicious threats over the Internet. A botnet comes from three main elements - the bots, the command and control (C&C) servers, and the botmasters [5]. The botmaster sends orders to all the bots on infected targets and controls the entire botnet through the Internet and the C&C servers. Figure 1 gives the general Structure of a Botnet.

Bots or Zombies [1][14]: Vulnerable computers are compromised by the bot malware, thus becoming zombies within a specific botnet.

Command and Control (C&C) Servers [9][12][14]: It is the main carrier of botnet functionality and the defining characteristic of bot malware. The C&C channel signifies a communication channel recognized between the botmaster and compromised computers. This channel is used by the attacker to issue commands to bots and get information from the compromised machines.

The Botmasters [12]: The botmaster sends instructions to all the bots on infected targets and controls the complete botnet network through the Internet and the C&C servers.

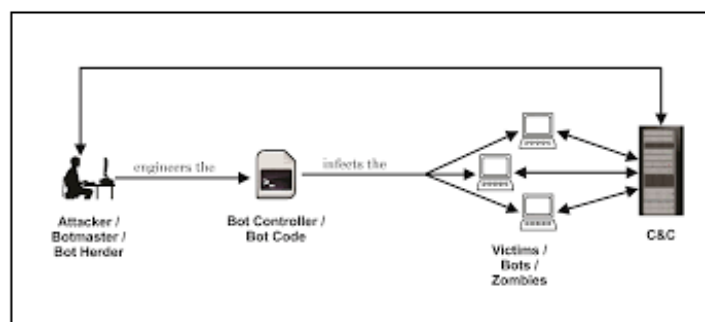


Figure 1: General Structure of a Botnet [14]

Botnets can be designed for spamming, traffic sniffing, key-logging, information gathering, and DDoS attacks. More new types of attacks are invented based on botnets. The detection of the botnet has been a major research topic in recent years. Several approaches have been suggested for the detection and tracking of a botnet.

DDoS attacks are often launched through well organized, remotely controlled, and widely distributed Zombies or Botnet computers of a network, that are continuously or simultaneously sending a huge amount of traffic or service requests to the target system. The attack results in the target system either responds so slowly, unusable, or crashes completely. It is very difficult for the defense mechanisms to identify the original attacker because of the use of spoofed IP addresses by zombies under the control of the attacker with a botnet.

Again the existing solutions for detection of DDoS attack like firewalls, intrusion detection systems are unable to detect the complex DoS and DDoS attacks since most of them filter the normal and attack traffic based on some fixed predefined rules

Sophisticated and automated DDoS attack tools have been developed to assist attackers in implementing all or some steps automatically with the minimal human effort to launch these attacks. The attackers can just configure desired attack parameters for a specified attack and the rest is managed by automated tools.

II. Literature Survey

Several types of research have been done that intend to different botnet detection approaches.

Faisal Hussain et al.[2] tries to enhance the existing solution for the detection of DDoS attacks by integrating it with artificial intelligence (AI). Convolutional neural network (CNN) models can be used to efficiently detect the complex DoS and DDoS by converting the network traffic dataset into images. The network traffic data is converted into image form and trained a state-of-the-art CNN model, i.e., ResNet over the converted data. The system can detect the DoS and DDoS with 99.99% accuracy in the case of binary classification.

Obinna Igbe et al.[3] presented a technique for detecting DoS attacks in a network using the dendritic cell algorithm (DCA) an Artificial Immune System (AIS)-based algorithm. The result evaluation is done using the NSL-KDD dataset. The system can detect DoS/DDoS attacks with a high detection rate and low false-positive rate.

Ryu et al.[4] analyzed the effect of ensembling machine learning algorithms with a neural network for botnet detection. They ensembled decision trees and Naive Bayes classifiers with a neural network and concluded that the given en-sembling technique can detect botnet attacks in network traffic in a better way as compared to individual classifiers.

Some of the existing methods for DDoS detection are limited as they can only perform well for the dataset on which they are trained due to the diversity of attack patterns. This problem is solved in [5] by **Faisal Hussain et al.** a universal features set to better identify the botnet attacks irrespective of the underlying dataset. Four machine learning algorithms like NB classifier, KNN, Ransom Forest (RF), and Linear regression (LR) are used for detecting the botnet attacks across three different datasets that are CICIDS2017 dataset, CTU-13 dataset, and IOT-23 dataset. The Classifier performed best for detecting the botnet attacks in all three datasets. On the other hand, the NB classifier showed the lowest performance for detecting botnet attacks in all three datasets.

Abbas Aboutet al.[6] designed a versatile graph to detect botnet attacks. The authors build a communication graph by representing hosts as nodes and communications between them as vertices. Afterward, they extracted the graphical features and applied different machine learning techniques to better detect the botnets.

Mohit Goyal et al.[7] propose a novel approach based on the behavioral analysis of the botnets to detect the IoT malware. The presence of this malware is detected using supervised machine learning algorithms taking the discovered features as inputs. Various machine-learning techniques are compared to conclude that neural networks outperformed all other methods like Logistic Regression, SVM.

Paulo Angelo Alves Resende et al.,[8] proposed new features to distinguish C&C channels from benign traffic. The detection method uses a random forest classifier implemented over Apache Spark, a Big Data processing framework with more than 99% of accuracy. The proposed features can be extracted before the communication end, which enables an early response.

Sajjad Arshad et al.[9] propose a fully anomaly-based approach that necessitates no a priori information of bot signatures, botnet C&C protocols, and C&C server addresses. One method is implemented that detects bots in the monitored network in real-time along with malicious activities (e.g. scanning, DDoS). The prototype system can be evaluated with real-world network traces including normal traffic and several real-world botnet traces.

Riaz Ullah Khan et al.,[10] presented an effective two-stage traffic classification approach to identify P2P botnet traffic based on both the non-P2P traffic filtering mechanism and machine learning techniques on conversation features. Firstly non-P2P packages are filtered to cut the network traffic with eminent ports, DNS query, and flow counting. Conversation features based on data flow features and flow similarities are extracted and by using Machine Learning Classifiers, botnet detection is done effectively. Three machine learning algorithms i.e Naive Bayes, Decision Tree, and ANN are compared. Where Decision Tree classification is outperformed due to the high-speed network environment.

Lakshya Mathur et al.[11] uses machine learning approaches to train classifiers by a specific network flow dataset. The trained classifiers were applied to the collected data to differentiate the normal traffic and the bot traffic with high accuracy and low false-positive rate. The packet headers are examined instead of the packet content to save both time and resources. By extracting the most relevant subset of features and with use of machine learning techniques like Logistic Regression, MultiClass classifier, Random Committee we compared the performance for botnet detection.

G.Kirubavathi et a.[13] presented a HTTP bot-nets detection system a work based on TCP related features. A Multi-Layer Feed Forward Neural Network training model using Bold Driver Back-propagation learning algorithm is created. This method can efficiently detect Spyeye and Zeus botnets. The given method can outperform as compared to the C4.5Decision Tree, Random Forest and Radial Basis Function.

III. Proposed System Overview

A botnet is several Internet-connected devices, each of which is running one or more bots used to perform Distributed Denial-of-Service (DDoS) attacks. Such attacks are typically attempting to exhaust victim's bandwidth or disrupt legitimate users' access to services. The traditional architecture of the internet is vulnerable to DDoS attacks and it provides an opportunity for an attacker to gain access to a large number of compromised computers by manipulating their vulnerabilities to set up attack networks or Botnets.

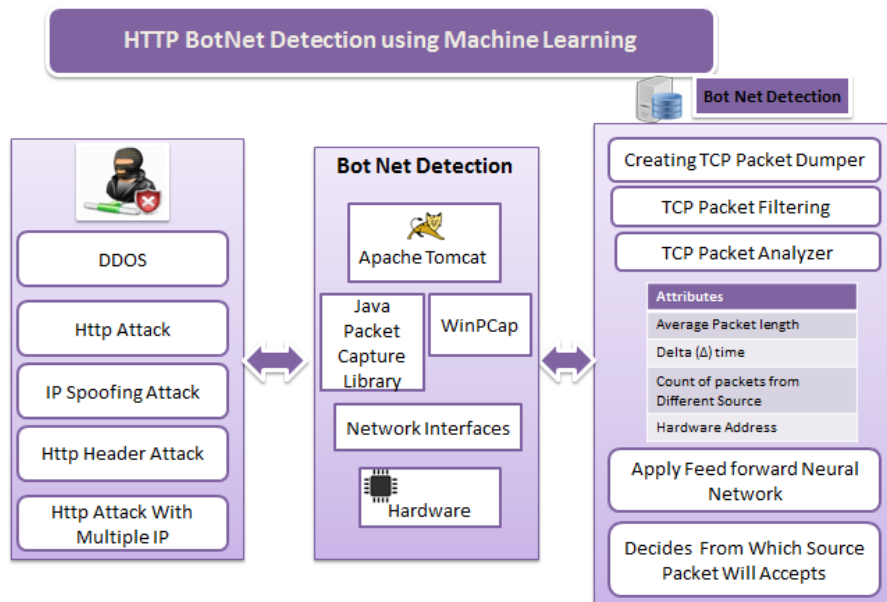


Figure 2: System Architecture

The proposed system helps to detect botnet activity by classifying network traffic behavior using Feed Forward Neural Network. The system makes use of the JPCap library to capture packets and analyze and filter by their type. The system can detect spoofing attacks, Header Attacks, and attacks with multiple IP with by analyzing certain features like Average Packet length, Delta Time, Packet count, and Hardware address. Feed-Forward Neural Network will be used to train the data analytics engine on the TCP Packet Analyzer attributes like count of packets coming from the particular IP, timestamp, and length of the packet, etc. for recognizing the request getting from the bot or not. Based on Feed Forward Neural Network result, the decision is taken to allow the packets from the particular machine or not which can be Bot.

IV. Conclusion

Detection of DDoS attacks is a challenging problem that needs to be addressed. Many researchers contributed towards detecting DDoS attacks with different approaches. This paper presents a review of the up-to-date analysis of existing botnet detection methods for identifying botnet-related traffic and the enormous background available in this area. The survey also helped to examine the botnet defense measures that need to be taken after the detection of botnet activity.

Here an attempt is also made to implement a system that detects botnet activity by classifying network traffic behavior using machine learning with greater accuracy by considering certain features like Average Packet length, Delta Time, Packet count, and Hardware address.

References

- [1] Saurabh P. Chaware , Prof. Sukhada Bhingarkar, "A Survey of HTTP Botnet Detection", International Research Journal of Engineering and Technology (IRJET), Volume: 03 Issue: 01 | Jan-2016.
- [2] Faisal Hussain; Syed Ghazanfar Abbas; Muhammad Husnain; Ubaid U. Fayyaz; Farrukh Shahzad; Ghalib A. Shah, "IoT DoS and DDoS Attack Detection using ResNet", 2020 IEEE 23rd International Multitopic Conference (INMIC).
- [3] Obinna Igbe, Oluwaseyi Ajayi, and Tarek Saadawi, "Denial of Service Attack Detection using Dendritic Cell Algorithm", 2017 IEEE.
- [4] S. Ryu, B. Yanget al., "A comparative study of machine learning algorithms and their ensembles for botnet detection, "Journal of Computer and Communications, vol. 6, no. 05, p. 119, 2018.

- [5] Faisal Hussain, Syed Ghazanfar Abbas, Ubaid U. Fayyaz, Ghalib A. Shah, Abdullah Toqeer, Ahmad Ali, "Towards a Universal Features Set for IoT Botnet Attacks Detection", <https://arxiv.org/pdf/2012.00463.pdf>.
- [6] A. A. Daya, M. A. Salahuddin, N. Limam, and R. Boutaba, "A graph-based machine learning approach for bot detection," in 2019 IFIP/IEEE Symposium on Integrated Network and Service Management (IM). IEEE, 2019, pp. 144–152.
- [7] Mohit Goyal, Ipsit Sahoo, G. Geethakumari, "HTTP Botnet Detection in IOT Devices using Network Traffic Analysis", IEEE 2019.
- [8] Paulo Angelo Alves Resende, André Costa Drummond, "HTTP and contact-based features for Botnet detection", 2018 John Wiley & Sons.
- [9] Sajjad Arshad¹, Maghsoud Abbaspour¹, Mehdi Kharrazi², Hooman Sanatka, "An Anomaly-based Botnet Detection Approach for Identifying Stealthy Botnets".
- [10] Riaz Ullah Khan, Rajesh Kumar, Mamoun Alazab, Xiaosong Zhang, "A Hybrid Technique To Detect Botnets, Based on P2P Traffic Similarity", 2019 Cybersecurity and Cyberforensics Conference (CCC).
- [11] Lakshya Mathur, Mayank Raheja, Prachi Ahlawat, "Botnet Detection via mining of network traffic flow", International Conference on Computational Intelligence and Data Science (ICCIDS 2018).
- [12] Matija Stevanovic and Jens Myrup Pedersen, "Machine learning for identifying botnet network traffic", <https://vbn.aau.dk/ws/portalfiles/portal/75720938/paper.pdf>
- [13] G. Kirubavathi Venkatesh¹, R. Anitha, "HTTP Botnet Detection using Adaptive Learning Rate Multilayer Feed-forward Neural Network", <http://dl.ifip.org/db/conf/wistp/wistp2012/VenkateshN12.pdf>
- [14] Emmanuel C. Ogu^{1,*}, Olusegun A. Ojesanmi², Oludele Awodele¹ and 'Shade Kuyoro, "A Botnets Circumspection: The Current Threat Landscape, and What We Know So Far", Information 2019, 10, 337; doi:10.3390/info10110337oro