RESEARCH ARTICLE OPEN ACCESS

Towards botnet detection through features using network traffic classification

Harpinder Kaur[1] , Sanjeev Kumar[2]
[1]*M.E Research Scholar, Department of Computer Science and Engineering,*
[1]*Chandigarh University, Mohali India*
[2]*Cyber Security Technology Division, CDAC, Mohali Punjab*

ABSTRACT
Botnet are becoming the most significant threat to the internet world. Botnet is the automated process of attackers that interacts with network traffic and its services. Botnet are automatically updated into the compromised system to collect the authenticated information. In this paper, we present a model to extract some features which are helpful to analyze the behaviour of bot members present in the particular network traffic. On the other hand, various superior methods are evaluated to extract weather network traffic contain bot or not. In particularly, our evaluation shows that the particular traffic contain any bot member in their communication.
Keywords: botnet, botnet traffic behaviour, network traffic classification, C&C server etc

I. INTRODUCTION

Botnet are the most serious threat to the computer network. Botnet are organized their network of compromised system by the command and control channel. Compromised machines contain bot codes which are running in bot network. Through the command and control channel, the botmaster can control all the compromised machines in their botnet by applying bot command. The network of the botnet will carry number of malicious activities which can infect our system and will make a part of bot network. These malicious activities are like sending number of spam emails, phishing private information, Ddos attack, stealing personal information or banking information, click-fraud activities etc. Through this botnet, it can cause vast damage on computer network [5].

By the lack of comprehensive evaluation of features sets, the number of existing studies resort to a limited number of botnet traces in their tested datasets. These approaches mostly compromised high detection and low false positive rates obtained.

In this paper, we aim to address the problems are mainly focus on the proper selection of accurate features for detection of botnet network. By reviewing the network flow based features commonly employed in the botnet detection system and examines their applicability. By properly analyzing flow based features, we design a dataset that contain both benign and malicious traffic through which we can compare dataset with the help of features [2].

To the best of our knowledge, our study attempts to systematically evaluate the large dataset with the given features of botnet traffic.

The rest of the paper is structured as fellows Section 2 discuss and summarizes the related work on botnet detection using network flow. Section 3 introduced our proposed model and our methodology. Section 4 outlines the evaluation and experiment. Section 5 shows the results. Section 6 makes some conclusion remark and outline with our future work.

II. RELATED WORK

These days botnet detection system are the most popular in research world. Number of methods has been proposed for botnet detection. Many researches use the observation to identify botnet network traffic different from normal network traffic, much botnet detection are categorized as host- based, network based detection and the hybrid of two. In the case of host based detection have been done on particular host. Host based basically focus on behaviour based and signature based approaches. On the other hand, in network based, detecting the network traffic flow. Finally in hybrid, detecting both the cases [14].

Girorie et al [16] is based on anomaly that bots must create on command and control server regularly, for the purpose to receive commands from the central authority. By this method, the author made one white list host through detection and kept rest of them in malicious state. Bot sniffer [17] it exploits the special and temporal similarities of botnet activities to differentiate them botnet network traffic from benign traffic. They can categorized the bot response by message responses,

such as sending message to IRC , activities responses by scanning the network traffic as sending spam emails.

Abdullah et al [18] evaluated the flow based features in previous botnet detection studies by employing a diverse dataset to effectively result features for accurate botnet detection. Bekerman [4] network classified to classify traffic is malicious or benign and attribute malicious activity to a malware family both for new or known malware. They show results by analyzing network traffic through feature selection.

Data collection for feature analyses: In this section, we describe the process of data collection. The data should be collected through the honeypot device. We will train are classifier for setting a good network traces of legitimate traffic even captured botnet traffic.

Normal traffic: To collect the legitimate network traffic, we used honeypot to capture the data from the network. And visualize that traffic onto wire shark. In normal traffic capturing processor, not any malicious activity should take place [3].

Botnet traffic: To collect the botnet traffic is the most challenging task. To analyse the behaviour of bot traces, first it should capture in the isolated environment so that, they cannot damage any other system, and system itself privacy. After executing the bot binaries, bot traces should be analyzed by their bot activities on the system. After analyzing its feature we collect some features which are possibly differentiate between benign and malicious traffic. Some features like total length, number of hosts, command and control channel types, source port, destination port, TCP connections etc [4].

Feature selection: In this proposed model, number of features were selected to make a detection model. Most of the features were selected from the TCP flow, because maximum conversation between attacker and victim system were done through TCP flow. But for this proposed model, features subset was selected to analyse the traffic contain botnet traffic. To calculate those traffic some features are selected which distinguish the normal traffic with botnet traffic, like first packet is unique when attacker attacks on the system, 3-way communication, packet sizes etc through which we analyse the traffic contain botnet traffic or not.

III. METHODOLOGY

The goal of this research is to be able to detect the presence of botnet in the network traffic, as well as trying to find the family of bot from which they belong. The bot is using IRC, HTTP or P2P protocol to infect the compromised system. Our detection system is independent to the content of the C&C message i.e. they don't examine the payload of the packet. Their only ability is to detect the botnet network traffic through selected features and differentiate the benign and non benign traffic. Figure 1 shows a graphical representation of the steps taken to create a instance to classify botnet traffic or benign traffic

- Collect all network traffic through honeypot
- Visualize network traffic through wire shark.
- Split the network traffic into conversation.
- Extract feature value for particular instance.
- Filter some impersistance instance from the traffic.
- Pass instance from random forest classifier.
- Find botnet high detection false positive.

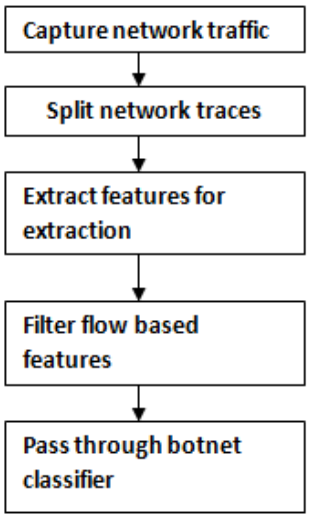

Figure 1: Generated steps of instance

IV. EVALUATION

We used Weka Learning Machine Java library, to build our classifiers. During the period of training, all selected features are labeled and indicating with source IP, destination IP, source port, destination port, protocol, total length, packet sizes, and resolved IP TCP connection. This labeling is prior to our knowledge for the botnet which are used for generated network traffic. Through this classifier we examine the effectiveness of Random Committee classifier. In the data processing stage it separates the individual conservations from the merged traces.

The best features set for the Random Committee Classifier was found by sensitively performed analysis on our feature set to reduce subset of features. These features are given below

that give best overall activity. Mean of bytes sent, bytes received, packet captured, TCP packet transmitted.

V. RESULTS

In the results, we choose Random Committee Classifier to classify the features of network traces of entire conversation through this we can find the accuracy rate of our detection method for classifying the entire network traces. In accuracy they defined the number of features classified divided by total number of features. And through this classifier, we can find the false positive rates which defined the number of features of legitimate traffic classified as botnet traffic divided by the total number of features of legitimate traffic. By using this, the classifier reduced feature sets.

Whereas the Naïve Bayes approach found a total 118656 instances, while in this approach instances associated with persistent conversations generated from legitimate traffic while calculating - false positive, it doesn't make sense to count any additional instances. By using Random Committee classifier the accuracy is 99.85% with 0.145% false positive rates and the accuracy of Naïve bayes is 80.21% with 0.975% false positive rates.

Random Committee classifier

Conversation	Total	Correct
Normal	54793	54791
Tbot	37723	37710
Kelihos	76	20
IRC bot	49	38
Blackhole	32134	32130
Zero access	544	538

Naïve bayes classifier

Conversation	Total	Correct
Normal	54793	54626
Tbot	37723	377664
Kelihos	76	66
IRC bot	49	28
Blackhole	32134	29429
Zero access	544	504

As shown in the above tables, the features values calculated for those traffic, in which something malicious or bot activity occurred. Through this features calculation we can find the false positive rates of network traffic. We can shown in figure 2 the example of bot activity in TCP conversation.

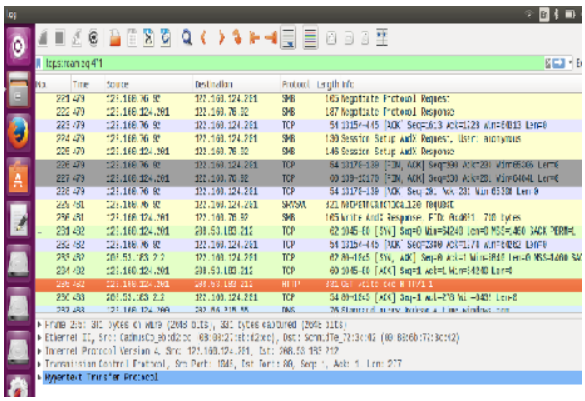

Figure2: TCP-3 way handshake with suspicious activity

VI. CONCLUSION

Botnet are the most serious threat to the internet world. There are number of malicious activities are interfaced through botnet into the computer, these activities are like sending email spams, stealing personal information like account information, phishing information and many more. In this paper, we detect botnet traffic through some features of the network flow and differentiating the traffic is malicious or not. In our proposed model we use some unique features to identify the traffic containing bot or the traffic is benign traffic. These methods are becoming the most popular method to identify the network traffic. Our detection method can achieve 99.85% accuracy of Random Committee classifier with 0.145% false positive rates and the accuracy of Naïve bayes is 80.21% with 0.975% false positive rates.

In near future, we would like to extend our work by adding more features to our model to identify the exact family of bot from which they belong.

ACKNOWLEDGEMENT

We would like to thank Cyber Security Technology Division lab of C-DAC, Mohali for providing us capturing live data from honeypot, and sharing their knowledge and time with us.

REFERENCES

[1]. Punit Sharma, Sanjay Tiwari, et. Al "Analyzing bot family behaviour and its detection" at international journal of engineering trends and technology

(IJETT) – volume 9 number 2, march 2014.

[2]. Claudio Mazzariello "IRC traffic analysis for botnet detection" at the fourth international conference on information assurance and security, 2008.

[3]. Basam Sayad, Weilu, et. Al "Botnet detection based on traffic behaviour analysis and flow intervals" computer and security technology, Nov 2013.

[4]. M.Stevanovic and J.M.Pederson, "Machine learning for identifying botnet network traffic", Networking and security, Department of electronic Systems, Aalborg university, tech. Rep. 2013.

[5]. "Tcpdump and libpcap", http://www.tcpdump.org.retrieved july 23, 2013.

[6]. L.Bilge, D.B Alzarotto, W.R Obbertson, E.Kirda "Disclosure: Detecting botnet command and control servers through large scale net flow analysis" in Proceeding of the 28th Annual Computer Security Application Conference. ACM, 2012, PP, 129-128.

[7]. Peter Ekstrand Berg, Katrin Franke, et. Al "Generic feature selection measure for botnet malware detection" computer science and media technology, Gjovik University, Norway, 2013.

[8]. Matija Stevanovic and Jens Myrup Pedersen "An analysis of network traffic classification for botnet traffic" Aalborg university, 2013.

[9]. Yao Zhao, Fang Yu, et. Al "Bot graph: Large scale spamming botnet detection" Microsoft research Silicon Valley, northwestern university, 2013.

[10]. Constantinn Musca, Emu Mrica, Razvan Deaconescu, et. al "Detecting and analysing zero-day attacks using honeypots" on 19th international conference on control system and computer science, 2013.

[11]. W.T.Strayer, R.Walish, C.Livadas, et al. "Detecting botnet with tight command and control, Local computer networks" ,31st IEEE Conference, pages195-202, Nov 2006.

[12]. Gregory Fedynyshyn, Moa Choo Chuah , et. Al "Detection and classification of different botnet C&C channels" at Lehigh university. Bethlehem, PA 18015, USA, 2010.

[13]. Yao Zhao, Yinglian Xie, et. Al "Bot graph: Large scale spamming botnet detection" at northwest university, Microsoft research Silicon Valley, Microsoft corporation, 2010.

[14]. Xiaonan Zang, Athichart Tangpong, et.al "Botnet Detection through Fine Flow Classification", The Pennsylvania State University, Jan 2011.

[15]. Wireshark, http://www.wireshark.org.

[16]. Frederic Giroiree, et al "Exploiting Temporal Persistance to Detect Covert Botnet Channels", Recent advances in Intrusion detection, 2009.

[17]. Guofei Gu, et at," Botsniffer: Detecting botnet command and control channel in network traffic, network and distributed system security",2007.

[18]. Raihana Syahirah Abdullah, Mohd Feisal Abdollah, et. Al "Preliminary study of host and network analysis on P2P botnet detection" on international conference on technology, informatics, management, engineering & environment, banding, Indonesia, June 23-26, 2016.