Detection of botnet by analyzing network traffic flow characteristics using open source tools

K.Shanthi
Dept. of Computer Science and Engineering
K.S. Rangasamy College of Technology
Tiruchengode, Namakkal District, India
shanthikannappan@gmail.com

D.Seenivasan
Dept. of Computer Science and Engineering
K.S.Rangasamy College of Technology
Tiruchengode, Namakkal District, India
bitseenu@gmail.com

***Abstract*—Botnets are emerging as the most serious cyber threat among different forms of malware. Today botnets have been facilitating to launch many cybercriminal activities like DDoS, click fraud, phishing attacks etc. The main purpose of botnet is to perform massive financial threat. Many large organizations, banks and social networks became the target of bot masters. Botnets can also be leased to motivate the cybercriminal activities. Recently several researches and many efforts have been carried out to detect bot, C&C channels and bot masters. Ultimately bot maters also strengthen their activities through sophisticated techniques. Many botnet detection techniques are based on payload analysis. Most of these techniques are inefficient for encrypted C&C channels. In this paper we explore different categories of botnet and propose a detection methodology to classify bot host from the normal host by analyzing traffic flow characteristics based on time intervals instead of payload inspection. Due to that it is possible to detect botnet activity even encrypted C&C channels are used.**

***Index Terms*— Bot, Bot master, Botnet, Botnet cloud, Mobile Botnet.**

I. INTRODUCTION

Botmaster exploits the vulnerabilities of individual computers connected in the internet to form the Botnet. Each vulnerable host infected by botmaster is termed to be a bot. Botmaster controls all the bots through Command and Control(C&C) Channel. Based on the command received from the attacker the bot performs. This activity is similar to robot function. So it is termed as Botnet. Bot itself is not a malware but can able to perform malicious actions by using simple scripts. Botnet can be available for lease in the market can ease any novice attacker to launch a large attack with little effort. Cyber terrorism activities can also be motivated become a threat to national security. Through strict laws and regulations many developed countries like US and UK can able to reduce the botnet activities but not completely overcome it.

The structure of botnet can be either centralized or decentralized. In the centralized botnet structure botmaster use legitimate servers for Communication and Control namely C&C server. Botmasters use legitimate protocols and ports to hide its presence. In these circumstances it is very difficult to differentiate bot host from the normal host. The existing detection techniques are not feasible for changing trends of botmasters.

A botnet lifecycle consists of four phases: initial formation, C&C, attack and post-attack. The initial formation phase is the first phase in which botmaster exploits the vulnerabilities of a target system and infects the victim machine by injecting the scripts and makes the compromised computer as a bot which enables the bot to execute malicious code in the third attack phase. The bot will establish a connection to the command and control channel then joining the botnet authoritatively in the second phase. In the attack phase, after launching a connection with C&C channel, bot will request for the command from the botmaster to begin the malicious activities. Typically in the post-attack phase, updation of the scripts which is loaded in the initial phase can be done to defend against detection by the intrusion detection mechanisms.

The main difference between Botnet and other kind of malwares is the existence of Command-and-Control (C&C) infrastructure. The C&C allows Bots to receive commands and malicious capabilities, as devoted by BotMaster. BotMaster must ensure that their C&C infrastructure is sufficiently robust to manage thousands of distributed Bots across the globe, as well as resisting any attempts to shutdown the Botnets. Recently, attackers are also continually improving their approaches to protect their Botnets. The first generation of Botnets utilized the IRC (Internet Relay Chat) channels as their Common-and-Control (C&C) centers. The centralized C&C mechanism of such Botnet has made them vulnerable to being detected and disabled. Therefore, new generation of Botnet which can hide their C&C communication have emerged, Peer-to-Peer (P2P) based Botnets. The P2P Botnets do not suffer from a single point of failure, because they do not have centralized C&C servers. Attackers have accordingly developed a range of strategies and techniques to protect their C&C infrastructure.

India ranks third among botnet infected developing countries as per Symantec's Internet Security Threat Report. Regardless of infecting computers, mobile phones especially Smartphones and andriod phones have become the target of

978-1-4799-6480-2/15/$31.00 ©2015 IEEE

attackers. Botmasters can utilize the vulnerable features of Bluetooth and SMS facilities to form the mobile botnet in order to theft personal details for financial purpose.

Recently botmasters can use Cloud services to construct botnet and it is the most dangerous because it is difficult to detect the botnet in a highly lively and diverse environment.

The most popular botnet detection mechanisms namely BotHunter and BotMiner are mainly based on complete payload inspection and these techniques can cause time and space complexity and it also incurs more money. Also these techniques are not feasible to detect encrypted C&C and some recent trends of botnet such as mobile botnet and bot clouds.

The rest of the paper is designed as follows. Section II describes about related work of botnet detection methodologies, section III explains about types of botnet, section IV describes the system overview, conclusion are drawn in the last session along with the future work.

II. Related work

Many researches have been carried out to detect the botnet and most of these works are based on payload inspection and not considering the network flow behavior. Due to these works IRC Botnet is mostly controlled. But novel bot techniques evade these intrusion detection systems.

BotHunter [1] is a botnet detection system designed with snort an open source tool which produces alarms for bot activities. It is the first kind of Bot infection profile analysis tool. BotHunter captures network activity and performs complete traffic payload analysis. It considers the botnet life cycle phases to detect bot activities. In this system detection activity is mainly based on Snort rules. The system consists of a correlation engine to generate a score based on probability of bot infected the network. BotHunter cannot detect encrypted command and control channels due to payload analysis. Payload inspection is very slow because of presence of large amount of data and user privacy is also affected. Recently Botmasters use stealthy scanning techniques to evade BotHunter.

BotMiner [2] is designed based on the group behavior of individual bots within a botnet. The system clusters similar behavior performed by different machines on a network. BotMiner performs 'C-plane'(C&C Traffic) clustering and 'A-Plane' (Activity Traffic) clustering techniques based on unique behavior of bots in the network via snort. After examining both the C-Plane and A-Plane, BotMiner associates hosts which show both similar network behavior as well as malicious activity to detect the botnet. It is also vulnerable to encrypted C&C channels and recent trends of botmasters such as mobie botnet and bot cloud. It gives very low false positive rate nearly 1% in the detection of IRC Botnet,P2P Botnet and HTTP Botnet.

Chunyong Yin, Lei Yang and Jin Wang [3] proposed a new general Botnet detection strategy. The proposed technique was based on degree distributions and anomaly net flows combined with data mining technology. In this scheme, accurate traffic profiles are constructed based on packet behavioral mode, and then dialog flow is introduced to draw traffic profile of node. Finally degree distributions of node and group are created and applied the degree distributions of node as input of data mining, which were then classified and distinguished to obtain reliable results with acceptable accuracy.

BotFinder [4], a novel system that detects infected hosts in a network using only high-level properties of the bot's network traffic. BotFinder does not rely on content analysis. Instead, it used machine learning to identify the key features of command-and-control communication, based on observing traffic that bots produce in a controlled environment.

Hossein and Azizah [5] framed a detection framework which focused on P2P based and IRC based Botnets. The proposed framework is designed to capture network traffic and efficient filtering is used to improve efficiency.Application classifier is used to classify applications such as IRC, P2P and HTTP. Finally clustering similar communication and malicious activity patterns within the same Botnet. In this approach there is no need for prior knowledge of Botnets such as Botnet signature.

HosseinRouhaniZeidanloo, et al [6] proposed a new detection framework which focused on P2P based botnets. Bots which are present belongs to the same botnet show unique communication behavior. In the proposed detection framework, group hosts which shows similar communication pattern in one stage and also group hosts which performs similar malicious activities in another step. Finally determining common hosts on them.

III. Types of botnet

Botnets are classified into five major categories based on the C&C Channels and resources used by the botmaster are shown in Fig. 1. They are,

- IRC-based Botnet
- Web-based or HTTP based Botnet
- P2P-based Botnet
- Mobile Botnet (SMS Based and Bluetooth Based)
- Bot Cloud (Cloud based Botnet)

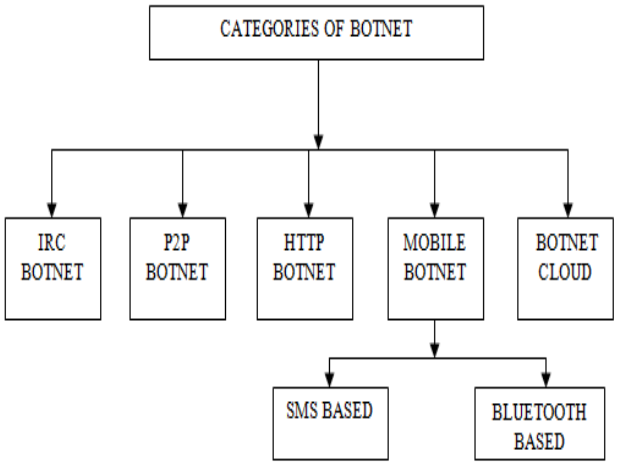

Fig. 1. Categories of Botnet

A. IRC-based Botnet

IRC is an on-line text-based instant messaging protocol in the Internet. It has client/server architecture with default channels to communicate between servers. IRC can connect hundreds of clients via multiple servers. Using IRC channels as handlers, attackers can use legitimate IRC ports to send commands to the bots making it much more difficult to track the command and control structure. Furthermore, an attacker can easily hide its presence because of the large volume of traffic that IRC server usually has. Examples of IRC Bots are GTbot, SDbot, Agobot ,Spybot etc.,

B. P2P Based Botnet

In P2P botnet, each bot can act as both C&C server and client. It is the decentralized network of nodes. There is no C&C server.The C&C communications are based on P2P protocols like WASTE and some are self-defined. Every bot in the network send keep alive message with one another and search the command file based on search key specified by botmaster. P2P botnet are more robust due to high resilient network. It requires higher latency for communication and it is the disadvantage of P2P Botnet. Examples of P2P Botnet are Nugache ,Storm, etc.,

C. HTTP – Based Botnet

HTTP based botnets are new technique of command and control used by botmasters to control their compromised computers (bots) and use HTML to communicate. Botmasters use HTTP protocol to hide their activities among the normal web traffic and easily escaped from current detection methods such as firewalls and other Intrusion Detection Systems.

The bots use specific URL or IP address defined by the Botmaster, to connect to a specific web server, which act as the Command and Control Server at regular intervals defined by Botmasters. Instead of remaining in connected mode after establishing a connection, the HTTP bots periodically visit HTTP C&C server (certain web server) to get updates or new commands. This model is called the PULL style. HTTP based botnets are mainly used to send spams.

D. Mobile Botnet

Mobile botnet can be formed by exploiting the SMS and Bluetooth features of the mobile phones. Botmaster can use both the Bluetooth and SMS as the C&C channel through which it is possible to control the bots.

1) SMS Based C&C

The first and foremost SMS mobile botnet is iKeeB[7]. Botmaster utilize SMS (Short Messaging Service) on smartphones and other mobile devices with such facility to transmit the malicious commands/scripts without the knowledge of the user. SMS-based C&C have some benefits that the bot master can easily communicate the root node, because the communication is depicted in a tree topology. On the other hand, SMS-based C&C have the disadvantage that SMS-based C&C need a list of nodes to be operated on infected phones. In the botnet, SMS based C&C channel is considered to be the weakest point and make the bot master hard to manage all the bots of the botnet.

2) Bluetooth Based C&C

Bot master can also utilize the vulnerabilities of Bluetooth technology and make it as the Command and Control (C&C) channel [8]. A Bluetooth based C&C is responsible for command transmission between the Bluetooth-enabled devices. It enables the faster communication by simplifying the authentication and authorization process. Example Bluetooth based Botnet is ZitMo.

E. Bot Cloud

Botmasters can use Cloud services to build botnet and it is the most extreme challenge faced by the cloud service providers because it is hard to detect the botnet in a highly dynamic and heterogeneous environment [9]. Such Botnets are known to be Bot clouds or cloud based botnet. The main advantage of the botcloud over the traditional botnet is that it can be possible for the botmaster to fully utilize the resources without any interruption. The resources are always available online and ready to use. A botcloud can be constructed and functional in a few minutes while a traditional botnet requires a lot of time for construction, waiting for days or months to take the control of the bots without the knowledge of the ordinary user. Finally a bot cloud can also be free from the risk of detection by security mechanisms.

IV. PROPOSED SYSTEM OVERVIEW

The architecture design of the proposed botnet detection is shown in Fig. 2.The proposed System consists of four phases. They are Network Traffic Capturing, Filtering, Attribute Selection and Classification. The System Overview is shown in Fig. 3. The proposed System finally produces white list, blacklist of IP address and Summary report.

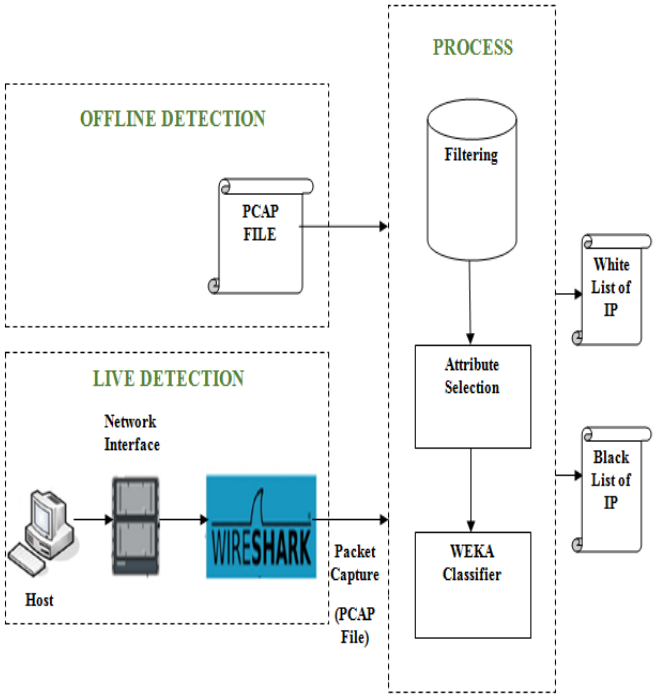

Fig. 2. Architecture Design

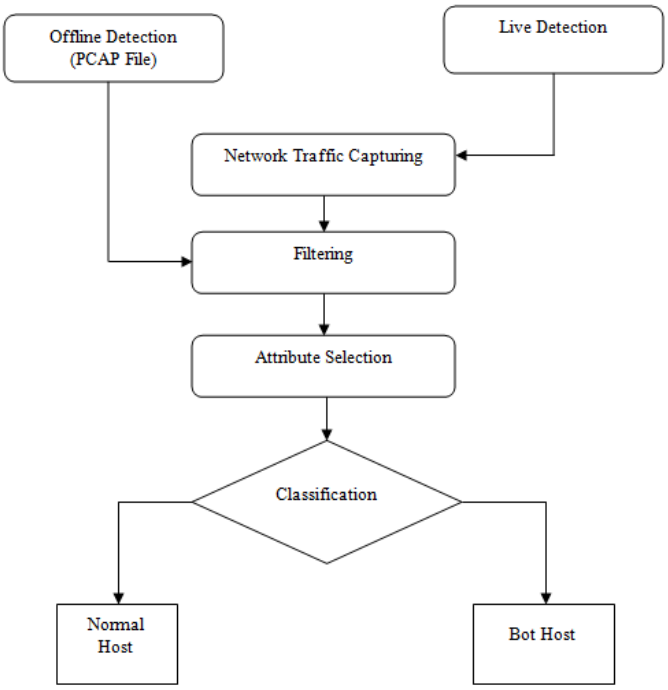

Fig. 3. System Overview

A. Network Traffic Monitoring

The WIRESHARK tool can be used to capture the network traffic based on the duration mentioned by the user. The network traffic caused by the wireless connection and local area connections are captured. The first phase facilitates for both offline detection and Live detection. In the offline detection pcap files captured by any network traffic capturing tool can be used for analyzing. The proposed system also analyze pcap file containing Mobile botnet traffic. Live detection can be done by capturing the traffic using WIRESHARK tool based on duration and network interface specified by the end user.

B. Filtering of Legitimate IP Address

The main aim of Filtering is to reduce the data for analyzing to improve the system performance efficiently. First, filtering out those traffics which targets destination IP address which are not possibly host Botnet C&C servers. For this purpose the top 100 websites from the web (http://www.alexa.com/topsites) can be used and it is hosted by Amazon. Flows having Single packet have also been removed. Similarly, flows representing NetBIOS services, broadcasting, SSDP, MDNS, NBNS, LLMNR, BROWSER and DHCP are removed.

C. Attribute Selection

An attribute is defined as the characteristic of a flow in a given time T which is a numeric value specified by the end user. First five attributes are taken directly from WIRESHARK pcap file and others have to be calculated. The attributes selected for further process are given in the following Table I.

TABLE I. LIST OF ATTRIBUTES

Attribute	Description
SrcIp	Flow source IP address
SrcPort	Flow source port address
DstIp	Flow destination IP address
DstPort	Flow destination port address
Protocol	All Transport and Application layer protocols
PX	Number of packets exchanged for time interval
PPS	Number of packets exchanged per second in time interval T
FPS	The size of the first packet in the flow
NDU	The number of ICMP Destination Unreachable
NR	The number of TCP retransmission
FPSec	Number of flows from this address over the total number of flows generated per Second

D. Classification

The classification is based on the selected attributes using WEKA decision tree J48 classifier without reduced error pruning with 10 fold cross validation and Naïve Bayes. This classification is to be done because botnet traffic exhibits unique behavior and more number of packets exchanged in a particular time frame rather than the normal traffic flow. Comparison of two classifications is given in terms of True and False Positive rates are shown in Table II. In the proposed system it is proven that Decision tree classifier is more efficient than Naïve Bayes.

TABLE II. COMPARISON OF CLASSIFIERS

Classifier	True Positive Rate	False Positive Rate
J48 Decision Tree	86.6953 %	13.3047 %
Naïve Bayes	78.5408 %	21.4592 %

E. Detection of Botnet

Based on the result of classification phase bot host and the legitimate host can be classified. Finally it is possible to detect the most dangerous botnet before it beginning the malicious actions. The proposed system produce the white list and blacklist of IP address and summary report.

V. CONCLUSION

Botnet represent currently one of the most predominant threats on the Internet. Bots act as sleeper cells based on the requirement of the botmasters. The most serious thing about botnet is it motivates cyber terrorism activities. Effective online botnet detection is challenging because of the complexity and ever changing technology underlying botnets. Early botnet detection poses some challenges because processing the entire payload of traffic is very difficult. Botmasters encrypt the C&C channels and randomize the packet size to evade from the current intrusion detection systems.

The need for novelty detection is to reduce the space and time complexity. Furthermore, the network traffic data volume and characteristics change rapidly and poses many challenges to the current intrusion detection mechanisms. The novel detection technique is based on analyzing the network traffic flow characteristics instead of payload inspection.

The proposed model allows detecting bot activity in both Offline mode and online mode based on the observation of its network flow characteristics for time intervals specified by the user. The concept of time intervals is used to limit the duration to detect the botnet. Finally the proposed system produce white of legitimate IP address and Blacklist of malicious IP address and Summary report of captured traffic.

The proposed work can be extended in the future to detect the growing trends of botmasters such as Mobile botnet and bot cloud.

REFERENCES

[1] GuofeiGu, Phillip Porras, VinodYegneswaran, Martin Fong and Wenke Lee, ”BotHunter: detecting malware infection through IDS-driven dialog correlation,” Proc. 16th USENIX security symposium, pp. 167-182, 2007.

[2] GuofeiGu, Roberto Perdisci, Junjie Zhang, and Wenke Lee, ”BotMiner: clustering analysis of network traffic for protocol- and structure-independent botnet detection,” Proc. 17th USENIX security symposium, pp.139-154,2008.

[3] Chunyong Yin, Lei Yang and Jin Wang, ”Botnet Detection Based on Degree Distributions of Node Using Data Mining Scheme,” International Journal of Future Generation Communication and Networking, Vol.6, No.6, pp.81-90,2013.

[4] Florian Tegeler, Xiaoming Fu, Giovanni Vigna and Christopher Kruegel,” BotFinder: finding bots in network traffic without deep packet inspection,”Proc. ACM CoNEXT '12 8th international conference on Emerging networking experiments and technologies, pp. 349-360,2012.

[5] HosseinRouhaniZeidanloo and AzizahBt Abdul Manaf, ”Botnet Detection by Monitoring Similar Communication Patterns,” International Journal of Computer Science and Information Security, Vol. 7, No. 3, pp.36-45,2010.

[6] HosseinRouhaniZeidanloo, AzizahBt Abdul Manaf, RabiahBt Ahmad, MazdakZamani and SamanShojaeChaeikar,”A Proposed Framework for P2P Botnet Detection,” IACSIT International Journal of Engineering and Technology, Vol.2, No.2,pp.161-168,2010.

[7] Abdullah J. Alzahrani and Ali A. Ghorbani, “SMS mobile botnet detection using a multi-agent system: research in progress,”Proc. ACM ACySE1st International Workshop on Agents and CyberSecurity,2014.

[8] Heloise Pieterse and Martin S. Olivier,” Bluetooth Command and Control channel,”Elsevier Computers & Security, vol. 45, pp. 75-83,September 2014.

[9] Jerome Francois, Shaonan Wang, Walter Bronzi, Radu State and Thomas Engel, “BotCloud:Detecting botnets using MapReduce,” Proc. IEEE International Workshop on Information Forensicsand Security, pp. 1–6,2011.