A SURVEY ON CYBER ATTACK: BOTNET

Sindhu Arumugam[1], Nandhini Selvam[2], V.Vanitha[3], V.P.Sumathi[4]

[1,2]PG Student, [3]Professor, [4]Assistant Professor

Computer Science and Engineering, Kumaraguru College of Engineering, (India)

ABSTRACT

Botnets represent one of the most serious cyber security threats faced by organizations today. Botnet is a network of compromised machines (Bots) used for malicious purposes. Unlike other types of malware, botnets have a central authority (the botmaster), use the infected computer to carry out malicious activities widely. This paper presents the classification of botnet based on their architecture and different types of bot family. The survey discuss about various methods for detecting botnet.

Keywords: Botnet, Bot, C&C Server, Centralized, DNS, HTTP, IRC, P2P

I. INTRODUCTION

Cyber security is also referred as information security which deals with protecting computers, networks, programs and data from misuse by unauthorized person. It is important because corporations, hospitals, governments, financial institutions, military and other businesses process which stores a great deal of confidential information on computers and transmit that data across networks. With the growing threats and attacks through malwares like botnet, hence it is necessary to protect sensitive business data, personal information, and to safeguard national security. Malwares are the malicious software programs that make some harmful attacks on the data that enters and leaves the internet.

Botnet is a distributed malware; consist of a network of compromised machines. It consist a leader named botmaster who control the whole network. The botmaster sends commands to the system (bots) in network. The systems perform actions without the knowledge of the user. The commands are sent through a centralized server or by using peer to peer communications.

Bot's life cycle consists of five stages which are depicted in the Fig 1. A host must undergo these stages in order to become an active bot.

The first stage is the injection stage wherein the host gets infected by downloading malicious software from websites or infected files attached to emails unknowingly. Then the infected host runs a program that searches for malware binaries in a given network database and hence become a real bot.

The second stage is the connection stage, the host contacts the server by using the IP addresses which is encoded directly as a list of static IP addresses or a list of domain names, which can be static or dynamic. While this makes it more difficult to take down or block a specific C&C server.

After establishing the command and control channel, the bot waits for commands to perform malicious activities. This stage is known as the waiting stage.

In the execution stage, the bot perform the malicious activities such as information theft, DDoS attacks, spreading malware, extortion, stealing computer resources, monitoring network traffic, spamming, phishing, etc. The last stage of the bot life cycle is the maintenance and updating of the malware. Maintenance is necessary if the botmaster wants to keep his army of zombies or to update codes for many reasons like adding new features, moving to another C&C server or escaping from detection techniques.

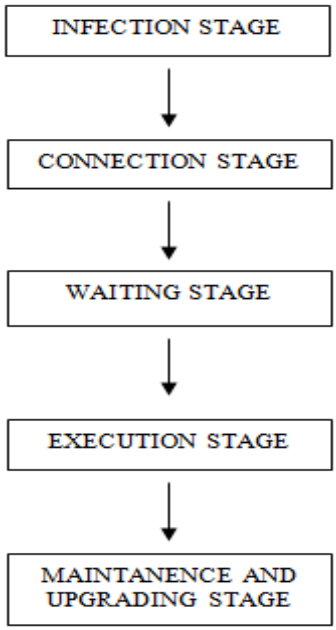

Fig. 1 Botnet Life Cycle

Botnet phenomenon supports a wide range of malware activities, including phishing, DDoS attacks, spam emails, click fraud, malware distribution, etc. Botnets have caused over $9bn in losses to US victims and over $110bn in losses globally. The botnet infected systems in India was 25,915 in 2007 which peaked to about 6.5 million in 2012. In 2010, McAfee Labs has detected more botnet infections almost 1.5 million was in India than in any other country. Approximately 500 million computers are infected every year globally, translating into 18 victims per second. Nowadays the effects of botnet spread into the new platforms, including smart phones, tablets and other mobile devices. In 2014, McAfee lab reported that more than 300 new threats evolves every minute with mobile malwares raising by 16% and overall malwares increasing by 76% every year. Therefore, in future need to have a powerful security model that should be compatible to every computing device.

II. ARCHITECTURE OF BOTNET

According to previous review [1], the architecture of botnet is based on how they communicate with the bots. It is classified into three types, they are centralized, decentralized and hybrid.

2.1 Centralized

In a centralized topology (Fig 2), all the bots report to and receive commands from a single C&C server. It provides good coordination and quick responses. But when the server gets fault then there won't be any future communication. Examples for centralized architecture are IRC and HTTP bots.

Star: It is the basic centralized topology. There will be only one server and the bots are connected to that server. By hijacking that server it will be easy for the detector to find the botmaster and bots connected to that server.

Hierarchical: Botnets incorporate one or more layers of proxies between the bots and botmaster. The proxies themselves are compromised machines serving the botmaster. If one of the proxies gets fault there won't be any change in the operations. This kind of architecture is very difficult to detect.

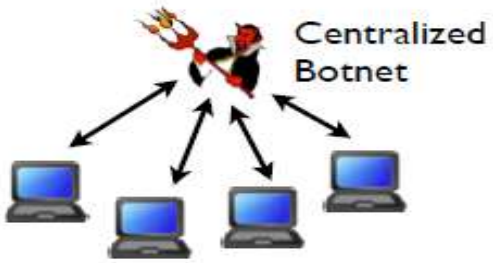

Fig. 2 Centralized Architecture

2.2 Decentralized

Botnets that have decentralized architecture (Fig 3) are very difficult to separate because there is no central server to find and to disable it. There are large numbers of bots present in this architecture; the loss of one of the bot does not spoil the whole network communication. Distributed: In a distributed topology, multiple servers control a subset of the bot family i.e. the servers are able to communicate directly with each other. If one server fails the nearby server take the control of the bots. There is no centralized point for failure. This architecture performs master and slave actions. Random: In this architecture there are no master slave actions. The commands can be sent and receive by any bots. This topology is robust and difficult to perform operations.

Fig. 3 Decentralized Architecture

2.3 Hybrid

The hybrid architecture (Fig 4) contains the features of both centralized and decentralized topologies. This method does not exist in the real time though it is very complex.

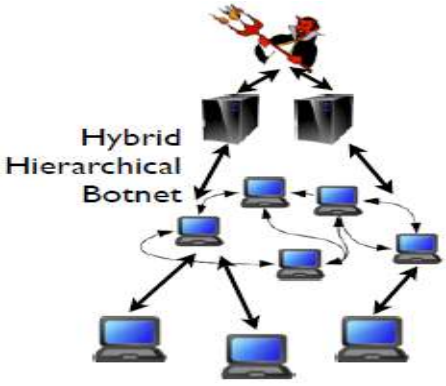

Fig. 4 Hybrid Architecture

III. TYPES OF BOTNET

The command and control channel (C&C), the means by which individual bots form a botnet, may be classified according to its specific architecture and operational modes. Generally bots are classified into four types which are given in the Fig. 5

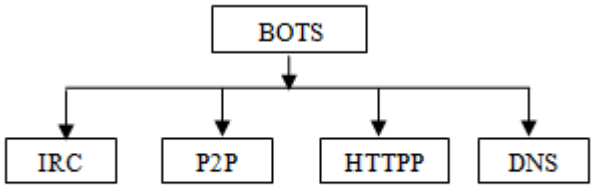

Fig. 5 Types of Bots

3.1 IRC BOT

The first generations of botnet use the Internet relay chat or IRC and the relevant channels to establish a central command and control mechanism. The IRC bots shown in Fig. 6, follow the PUSH approach as they connect to selected channels and remain in the connect state till it receives command from botmaster. The IRC botnet are simple to use, control and manage, but they are suffer from a central point of failure.

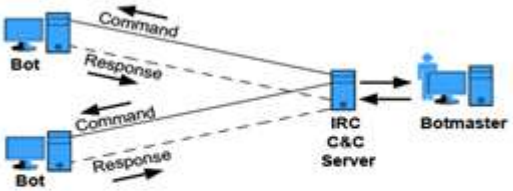

Fig. 6 IRC Bot

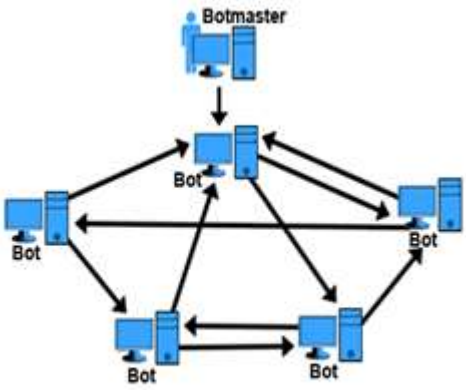

Fig. 7 P2P Bot

3.2 P2P BOT

To overcome this issue, the peer to peer architecture is used in the second generation of botnet where instead of having a central C&C server, the botmaster sends a command to one or more bots, and they deliver it to their neighbors which is depicted in the Fig. 7

3.3 HTTP BOT

Since the botmaster commands are distributed by other bots in the bot network, it is not able to monitor the delivery status of the commands. Furthermore, the implementation of a P2P botnet is difficult and complex. Hence, botmaster have begun to use the C&C again, in which the HTTP protocol is used to send the commands via web servers. Instead of remaining in connected mode, the HTTP bots periodically visit certain web servers to get updates or new commands. This model is known as PULL style and continues at a regular period which is specified by the botmaster. The diagrammatic representation of HTTP botnet is shown in the Fig 8.

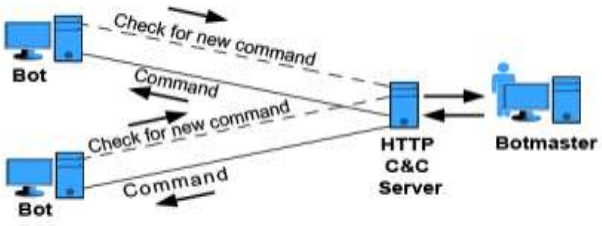

Fig. 8 HTTP Bot

Detection of the HTTP botnet with low rate of false alarms (e.g. false negative and false positive) has become a notable challenge. The detection of HTTP Botnet gets even worse where the Botmasters use the legitimate websites (e.g. hacked servers) or normal services (e.g. social bots) to establish their command and controls.

3.4 DNS BOT

DNS is a technique that a cybercriminal can use to prevent identification of his key host server's IP address. By misusing the way the domain name system works, he can develop a botnet with hosts that join and leave the network faster.

IV. LITERATURE REVIEW

The EggDrop bot is one of the earliest popular IRC bot was developed in 1993 to control the interactions in Internet Relay Chat (IRC) chats room. Later bots designed for performing malicious activities. Bots attack emerging technologies like social media, cloud computing, smart phone technology, and critical infrastructure. Hence it is important to detect and prevent the bots in order to avoid the malicious behavior both in the host and network.

Zhao et al. (2013) [2] proposed a data mining approach for botnet detection based on decision tree classification algorithm, achieving detection rates of 90% with false positive rate of 5%. In their work the authors capture network flows, splitting into multiple time windows and extracting set of twelve attributes, were used to classify malicious and non malicious flows. In detection technique, there were two phases namely training phase and detection phase. In training phase, using the existing datasets collected by honeypot, the algorithm will trained. In the detection phase, the botnet will be detected. Using existing datasets, authors experimentally proved that it is possible to detect the presence of unknown botnet based on time interval.

Lu et al. (2011) [3] proposed a clustering algorithm to detect the botnet. Here a payload analysis methodology was implemented. This methodology will analyze the characteristics of bit strings in the packet message. The main disadvantage is when the message are encrypted this method is not applicable. Authors design a clustering framework for botnet detection which consists of three components, namely feature analysis, clustering and botnet cluster labeling. In feature analysis, two observations were made. First, the response time of bots though it will be an instant reply. Next, behavior of botmaster commands though it will be automated. In clustering algorithm, the botnet instances and the normal instances were separated into two clusters. A high detection rate with low false positive rate was obtained.

Dietrich et al. (2013) [4] proposed the machine learning approaches for detecting the C&C botnet. Initially the known malware traces are gathered. The authors have chosen three features for traffic analysis process. They are protocol of the C&C protocol, length of the packet and finally the number of distinct byte values in the query section of the HTTP request URI. By using hierarchical clustering, the centroids for the known malwares were calculated. Thus classification was done for the unknown flows where the training is based on the cluster's centroids and the bot flows can be detected.

A new framework is proposed by Choi et al. (2012) [5] for detecting DNS botnet namely BotGAD (Botnet Group Activity Detector). The BotGAD framework consist of five main components namely data collector, data mapper, correlated domain extractor, matrix generator and similarity analyzer. By calculating the similarity score for the matrixes, the botnet were detected. The BotGAD can be used to detect the botnet in real time and also unknown bots can be detected here. But the botnet using DNS protocol were only be detected by this framework.

In this paper P2P botnet were detected by analyzing the communicational behavior of the command and control request packets. Zhang et al. (2014) [6], initially identifies the P2P clients in the network. Then statistical fingerprints of the P2P communications were analyzed for the detection purpose. A flow clustering method is used for detecting the P2P botnet. The evaluation results shows 100 % detection rate for P2P botnet.

Meng et al. (2014) [7] developed an adaptive blacklist-based packet filter which is a statistic-based approach aiming to improve the performance of a signature-based NIDS. The filter consists of two parts namely blacklist packet filter and a monitor engine. The blacklist packet filter is the component that filters out network packets by comparing packet payloads with stored signatures. The actions of the monitor engine are monitoring the NIDS, collecting the statistical data and calculating the confidences of IP addresses. It is responsible for updating the blacklist in a fixed time. Advantages of this Adaptive blacklist-based packet filter algorithm are Adaptive to the real context, reduction of the processing time, defend against IP spoofing.

Wang et al. (2011) [8] proposed a behavior-based botnet detection system based on fuzzy pattern recognition techniques. This system is deployed to identify bot-relevant domain names and IP addresses by analysing network traces. After identified Domain names and IP addresses used by botnets then that information can be further used to prevent protected hosts from becoming one member of a botnet. The system consists of two phases namely the DNS phase and the network flow phase. In the DNS phase, detection of bots based on DNS features is done. In the network flow phase, detection of bots based on network flow features is carried out. Advantage of this algorithm is the ability to find inactive bots.

Botnets often use IRC (Internet Relay Chat) as a communication channel through which the botmaster can control the bots to perform attacks or launch more infections. Chen et al. (2014) [9], proposed anomaly score based botnet detection in order to identify the botnet activities by analysing the similarity measurement and the periodic characteristics of botnets. To improve the detection rate, the proposed system employs two-level correlation. This method can differentiate the malicious network traffic generated by infected hosts from the normal IRC clients. Advantage of this detection system is ability to find the IRC bots at the communication stage.

Network traffic monitoring and analysis-related research has struggled to scale for huge amounts of data in real time. Kamaldeep et al. (2014) [10], designed a scalable implementation of real time intrusion detection system using the open source tools like Hadoop, Hive and Mahout. This implementation is used to detect Peer-to-Peer

Botnet attacks using Random Forest machine learning algorithm. The technologies underlying this framework are Libpcap, Hadoop, MapReduce and Mahout. The Random Forest Algorithm was chosen because the problem of Botnet detection has the requirements of high accuracy of prediction and has ability to handle diverse bots.

Huang et al. (2013) [11], proposed an effective solution to detect bot hosts within a monitored local network by tracking failures generated by a single host for a short period. The proposed system can be divided into two parts namely, the training phase and the detection phase. In the training phase, collect numerous benign traces, peer-to-peer application traces, and bot traces and then filter out non-failures, extract features from failure flows after which build the classification model using the C4.5 algorithm. In the detection phase, the process is similar to the training phase but it make use of the knowledge gained in training phase and analyse the data using those knowledge. The performance of the proposed solution depends on both training traces and selected features.

V. BOTNET DETECTION TECHNIQUE

Botnet detection technique can be classified based on their structure, protocols and behavior. Here the detection techniques are classified into three types based on tracking the components of botnet. They are bot detection, C&C detection and botmaster detection.

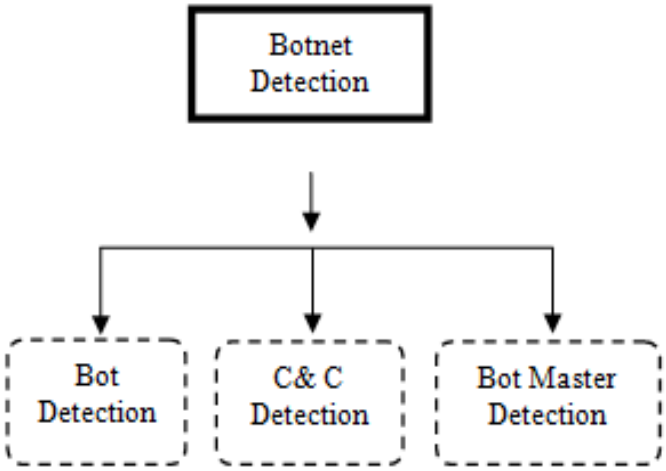

Fig. 9 Taxonomy of Botnet Detection

5.1 Bot detection

Detection does not depend on the bot family. Find whether the machine gets infect by botnet.

5.2 C&C detection

Detection of the C&C server or channel through which a central authority (i.e., the botmaster) may issue commands to his army of zombie machines and essentially take full control over the infected machines.

5.3 Botmaster detection

Most of the current botnet detection approaches work only on specific botnet command and control (C&C) protocols (e.g., IRC) and structures (e.g., centralized), and can become powerless when botnets change their approaches. Hence it is essential to detect Botmaster who giving commands to army of zombie machines. These three botnet detection are again classified into two types, namely passive detection and active detection.

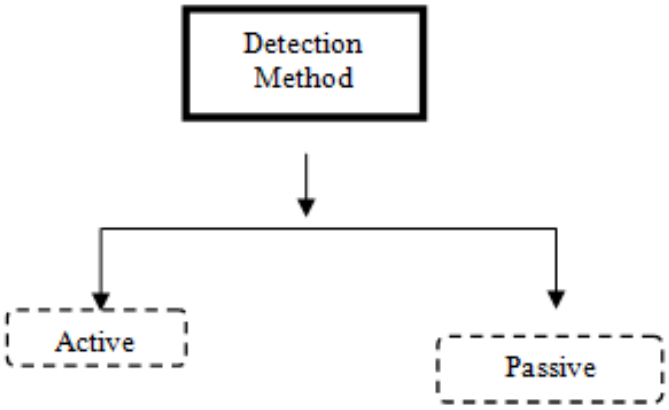

Fig. 10 Classification of Detection Method

In active detection methods, packets are injected into the network to measure network response whereas passive detection techniques observe data traffic in the network and look for presence of any suspicious communications. Active and Passive detection can also be future classified and it is shown in the Fig. 11 and Fig. 12

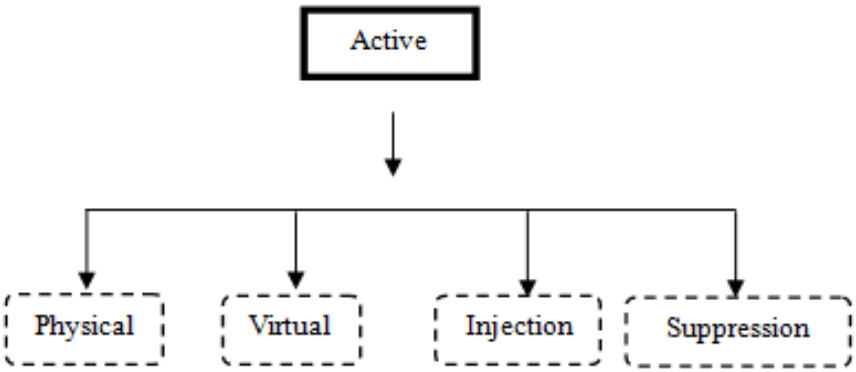

Fig. 11 Active Detection Techniques

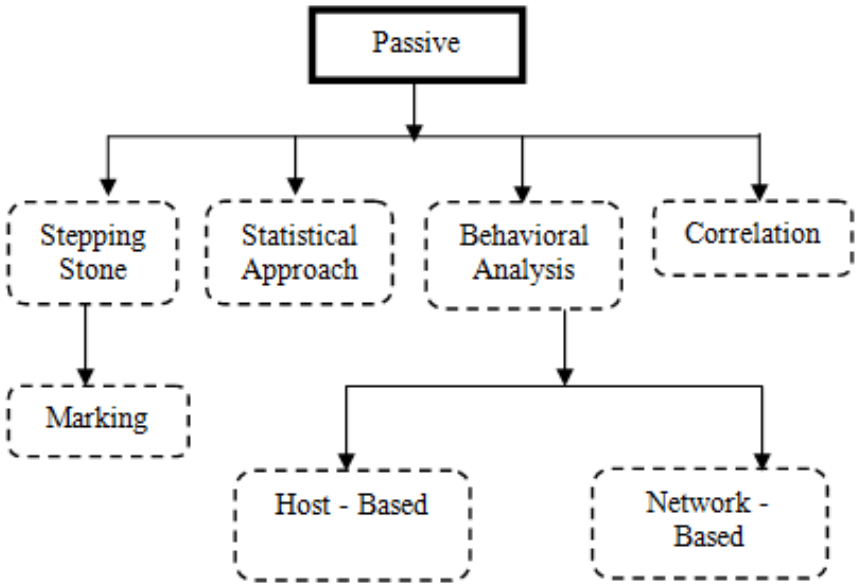

Fig. 12 Passive Detection Techniques

VI. EFFECTS OF SOME BOTS

- Festi - is also known as a king of spam is one of the most powerful spam and DDoS attackers since 2009
- GOZeus - short for Gameover Zeus is a peer-to-peer variant of the Zeus family of malware, designed to steal bank log-in credentials (2011)
- Zitmo - short for Zeus in the mobile designed for mobile devices has stolen information, from more than 30,000 banking customers(2012)
- Brobot - powerful botnet, used to launch distributed denial-of-service (DDoS) attacks in almost 50 banks in America(2013)
- Malwares taking control of smart appliances such as Television, Air conditioners, Refrigerators etc.
- Smart refrigerator act as bot when connecting to the internet and used to send thousands of malicious emails.

Table 1: History of Botnet

Year	Botnet Name	Description	Size	Company Taken Down
1993	EggDrop	C&C	<1,000	-
:	:	:	:	:
2009	Waledac	1.5 Billion spam mail/day	80,000	Microsoft
2008	Srizbi	One of the world's largest botnet	40% of all the spam	McColo
2011	TDL-4	P2P	4.5 millions	Kaspersky
2012	Grum	18000 Spam mail /day World's 3rd largest	560,000–840,000	Fireeye
2012	Carna	Internet census of 2012	420,000	-
2013	Zeus	Bank malware $70 million	163,812	Microsoft
June 2013	Citadel	Steal more than half a billion dollars	More than 5 million	Microsoft
February 2013	Bamital	-	More than 8 million computers	Microsoft
2013	Zeroaccess	$2.7million/month Bit coin miner	2 million victims	Microsoft

Table 2: Comparative Study of Detection Techniques

PP. No.	Botnet Name	Algorithm used	Advantage	Result	
				TP	FP
[2]	Weasel Black energy	Decision tree classification Reduced error pruning algorithm	Detect during C&C, attack phase	90%	5%
[3]	IRC	Merged K-Means clustering	Training the unlabelled dataset	84.6%	10%
[4]	Virut, Palevo Hlux, Miner	Agglomerative hierarchical clustering	Recognize the C&C Channels	88%	0.1%
[5]	DNS	Botnet Group Activity Framework	Early detection	95%	0.4%
[6]	Waledec, Storm, Skype, Zeus	K-Means Clustering Hierarchical clustering	Identify Stealthy P2P Provides Scalability	100%	0.2%
[7]	-	Adaptive blacklist based packet filtering algorithm	Adaptive to the real context Reduction of the processing time	-	
[8]	IRC, HTTP,P2P bots, 44 bot sample, zeus bot, etc	Fuzzy pattern recognition based filtering algorithm.	Can find inactive bots	95%	0-0.3%
[9]	IRC bot	Anomaly score based detection technique	Identifies the existence of botnets in the communication stage	90%	<7%
[10]	P2P	Random forest decision tree algorithm	Scalable, Disributed	99%	2%
[11]	-	C4.5 decision tree Algorithm	Ability to find new bots	99%	0-0.5%

VII. CONCLUSION

From the previous studies and research work it can be concluded that Botnet connection has specific network flow characteristics in which frequent communication happens between C&C and infected machine. To detect botnet, network flow analysis is the best approach. Data mining algorithms and machine learning approaches are easily applicable on network flow information.

Classification, Clustering, Fuzzy Logic and Artificial Intelligence related approach already been implemented for botnet detection. Most of the botnet detection methods identified the bots by analyzing the offline data and hence, need an approach to implement detection techniques in real time. As new bots evolving every day, it is important to enhance the detection techniques.

REFERENCES

[1] Sérgio S.C. Silva, Rodrigo M.P. Silva, Raquel C.G. Pinto, Ronaldo M. Salles,"Botnets: A survey", Computer Networks 57, **2013,** pp **378–403**.

[2] David Zhao, Issa Traore, Bassam Sayed, Wei Lu, Sherif Saad, Ali horbani and Dan Garant, "Botnet detection based on traffic behavior analysis and flow intervals", computers & secrity 39, **2013,** pp **2-16.**

[3] Wei Lu, Goaletsa Rammidi and Ali A. Ghorbani, "Clustering botnet communication traffic based on n-gram feature selection", Computer Communications 34, **2011,** pp **502–514.**

[4] Christian J. Dietrich, Christian Rossow and Norbert Pohlmann, "CoCoSpot: Clustering and recognizing botnet command and control channels using traffic analysis" Computer Networks 57, **2013,** pp **475–486.**

[5] Hyunsang Choi and Heejo Lee, "Identifying botnets by capturing group activities in DNS traffic", Computer Networks 56, **2012**, pp **20–33.**

[6] Junjie Zhang, Roberto Perdisci, Wenke Lee, Xiapu Luo, and Unum Sarfraz, "Building a Scalable System for Stealthy P2P-Botnet Detection", IEEE transactions on information forensics and security, vol. 9, no. 1, **2014.**

[7] Yuxin Meng, Lam-ForKwok, "Adaptive blacklist-based packet filter with a statistic-based approach in network intrusion detection", Journal of Network and Computer Applications 39, **2014**, pp. **83–92.**

[8] Kuochen Wang, Chun-Ying Huang, Shang-Jyh Lin, Ying-Dar Lin, "A fuzzy pattern-based filtering algorithm for botnet detection", Computer Networks 55, **2011,** pp. **3275–3286**.

[9] Chia-Mei Chen, Hsiao-Chung Lin, "Detecting botnet by anomalous traffic", journal of information security and applications, **2014** pp. **1–10.**

[10] Kamaldeep Singh, Sharath Chandra Guntuku, Abhishek Thakur, Chittaranjan Hota, "Big Data Analytics framework for Peer-to-Peer Botnet detection using Random Forests", Information Sciences 278, **2014**, pp. **488–497**.

[11] Chun-Ying Huang, "Effective bot host detection based on network failure models", Computer Networks 57, **2013,** pp. **514–525**.