International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8, Issue-1, May 2019

Botnet Detection Techniques – An Analysis

Jwala Sharma, Samarjeet Borah

Abstract: A botnet is a network of computers that has been compromised under the influence of malware code and being controlled by botmaster remotely. Every single day botnets develop new evasion techniques to make their presence undetectable. Knowing the consequences of botnets attacks; security analyzer must develop more robust detection methods. In this paper, an overview of botnets along with some real example of the same i.e. rustock, waledac, zeus, conficker are given. Signature based detection technique focuses on pattern, such as network traffic and then search for the known malicious pattern only. Therefore, to detect unknown attacks a robust detection technique is required. Anomaly-based detection techniques are used in such cases. Analysis of anomaly-based detection techniques are probed in this paper.

Keywords : Botnet, Security, Botmaster, Signature Based Detection, Anomaly Based Detection

I. INTRODUCTION

The recent internet landscape has reached the era which is much more than just human and machine interaction, the trending such one is "botnets", bot being derived from word robot and net from network, bot being able to perform repetitive task and can establish a connection among networking users and having an ability to mimic human behavior are rapidly growing all over the internet.

Number of factors contributes to the growth of digital data, which has resulted in enormous data generation and sharing of this data worldwide throughout internet users, be it social networking sites, email service, websites, botnets are becoming a severe threat to network security. E-mail services widely being used to exchange information, personal information, cooperate information has resulted with the emerging of unsolicited message, usually known as spam. Botnets are used in wide variety of application which are responsible for performing malicious activities like Distributed-denial-of-service (DDOS), mass spamming, phishing, identity theft, DNS spoofing, click frauds, adware installation etc. [1], [3]. It is estimated that 269 billion emails [2] are sent every day, so with the dependency on use of email services has created a lot many problems as it is very annoying for the users who do not wish to get flooded their mailbox with advertisement link, fraudulent links, and for innocent users who fall for such scams. In this paper, study of different real botnets and their evasion techniques has been discussed. Contents of the paper are organized in the following way: Section II discusses the overview of botnets along with architecture, Section III discusses some real botnets, section IV discusses about botnet detection techniques and some of the related works, section V discusses about the comparative analysis about different detection methods and section VI is about the conclusion.

Revised Manuscript Received on 30 May 2019.
***** Correspondence Author
Jwala Sharma*, Department of Information Technology, DDE, Sikkim Manipal University, Sikkim, India
Samarjeet Borah, 2Department of Computer Applications, Sikkim Manipal Institute of Technology, Sikkim, India

II. BOTNETS

A botnet is a network of computers that has been compromised under the influence of malware (bot) code and being controlled by botmaster remotely. It works on command and control mechanism, once the system has been compromised, after bot code or malware bring installed on it, the system becomes a bot or a Zombie. These bots distribute themselves over the internet by finding vulnerable or unprotected computer system that they can infect and further create a network of Zombies [3]. There are different botnets available across the internet, despite being different, their activities are common. The activity refers to the botnet actions they perform during their life time, which is referred as botnet life cycle: spread and infection, hiding and securing, command & control, launch, attack etc. Therefore, to address arising problem of many botnet detection techniques have been proposed in the past years; meanwhile spammers continue to upgrade their evasion techniques to make their presence undetectable.

A. Common Components of a Botnet

- Command and control server: C&C are the centralized server that issues commands to bots and receives reports or information from bots. The two main are C&C mechanisms used by botnets are centralized and decentralized.
- Peer to peer botnet: P2P operate without having any structure, and it can include C&C server too
- Botmaster: Also known as controller or bot herder. It remotely controls the botnet using C&C.
- Bot: A host that has been compromised by malware code or bot code. It can be IOT, smart phone or computer.

B. Botnet Architecture

Use of command and control infrastructure as one of the important characteristics of botnet makes them different from any other type of malware [4]. Botnet works on command and control mechanism and it has mainly two types: centralized and distributed or decentralized.

1. Centralized Architecture

Botmaster uses central server to give commands to selected set of bot, they make use of protocols such as internet relay chat or hypertext transfer protocol to operate their operations.

Internet relay chat (IRC) [5] is based on instant messaging protocol over internet and works like client server architecture. Botmaster creates IRC channel on command and control server and sends command to it, which on return relays the command to the client through Zombie is created. After the successful installation of bot on the victim host, client executes the commands and reports their result back to the botmaster. Centralized architecture often suffers single point failure as they can be easily traced and blocked by the administrators using firewall. Because of the restriction on IRC, spammers began to use HTTP protocol to hide their activities. Using HTTP was quite advantageous for the spammers as HTTP traffic are used in wide range for the services it provides, further they can avoid being detected by the firewalls. Still using HTTP in centralized architecture suffers from a single point of failure, because the working of the network and distribution of commands from the botmaster heavily relies on the central server.

2. Distributed Architecture

[5] Absence of centralized server, where each bot is a peer acting as a server and as a client at the same time. It is based on P2P protocol, in this architecture botmaster sends commands to more than one bot and they deliver the commands to its neighbors and they form a formation by connecting to each other. It doesn't suffer from single point of failure as in this architecture if one bot is detected by IDS, only it is, and its neighbor is affected not the entire network. The spontaneity of framing a network of bot in this architecture is often slow in convergence, having suffering form greater response time, it becomes difficult to manage the network and lacks in scalability in the whole discourse.

C.Botnet Features

Understanding the features of botnet is very important for the detection of them. Botnet features vary through a large category [22]; **propagation** methods can be active or passive. In the active propagation method bots disseminate through scanning open ports and passive propagation is derived by drive by downloads, removable media, and social engineering.

Another way of understanding the botnets is to know their **purpose**, information gathering, distributed computing, DDOS attacks, cyber fraud, spreading malware, identity theft is some common.

One of the popular techniques, honeynet is very useful to understand the **topology** of the botnets in the network, the centralized architecture, where the use of central server is made to give all the commands to the bots in the network, is quite a simple topology, which lacks robustness and can be easily bought down. To make the network more robust, decentralized or P2P topology is implemented, where each bot acts as a server and a client at the same time. Even, more sophisticated topology is implemented, by combing the both, that is hybrid topology.

In order to make their presence imperceptible bots uses **evasion techniques.** When traffic is monitored for malware detection, bot uses binary obfuscation, to confuse the detector, similar anti-analysis, security suppression; rootkit technology is used by bots as an evasion technique.

Botmaster uses C&C to remotely control the botnets. Therefore, if detection technique can find out this C&C server, whole botnet can be brought down. This can be performed by tracing the IP generated by the C&C server. To make their presence undetectable, botmaster implements IP Flux, Domain Flux, Rouge DNS server techniques. Many other evasion techniques used by C&C includes encryption of the communication traffic, protocol and traffic manipulation.

III. Review of Botnets

In this segment generic characteristic of real-world botnets are explained. Botnets use various method to propagate across the network and infect the machine, over the time as many security researches are coming up with new detection techniques, meanwhile spammers and botnet developers are keeping pace to come up with anti-detection techniques.

As per the research the oldest internet bot can be traced back to 1988, with the appearance of IRC, internet relay chat. And so the era of botnets emerged with other bots like web crawlers for the first search engines to index the web pages in 1994 and with this came up other malwares like Trojan, worm etc.6

The popularity of botnets continued to grow and created chaos in the history of internet and is still on the run. So here are some the most dominant botnets that has created an emergency over internet in the past and still are in run to elude their presence.

- *Bagle:* It is a mass mailing computer worm first sighted in early 2004. It used SMTP engine to mass mail to the infected computers. It made a copy of itself in Windows system and opened a backdoor on TCP. It was estimated that it infected 150,000 to 230,000 systems before it was brought down at January 18 2004.6.
- *Rustock:* It was dominant in the year 2006 until it was brought down in 2011. It consists of computer system running Microsoft windows as their operating system. It was estimated that it was able to send 25,000 spams per hour and at an average 192 per minute. It gained its popularity at that era because of the Rootkit technology it implemented to keep their presence undetectable. Rootkit is a software program designed to enable access to computer system being unauthorized user, it simply means that attacker obtained administrator /root access and it can install any malicious code it wanted to.6,[11]
- *Storm:* It was created in 2004 and first identified as botnet on September 2007. It distributed email with different subjects once such subject was "230 dead as storm batters Europe ".

It comprises of computer running in Microsoft windows, once the computer was infected with malicious code, it becomes a bot and it starts to perform automated tasks like gathering data, attacking websites, and forwarding emails. It used P2P architecture to create its network and the method used by storm to propagate were websites offering free music. On September 25 2007 Microsoft update Windows malicious software removal tool which helped to reduce the botnet size and on 2008 its storm declined until 2010 Stormbot2 was back with the absence of P2P architecture, which it used earlier.6,9

- *Torpig:* it began its development 2005. At its initial stage it spread through phishing emails. The infection method torpig used was drive-by-downloads, where users are not required to click on any ad, as malicious ad recognizes old software version in system and download will commence without any indications. It gained its popularity by taking the advantage of outdated version of java, flash. After spreading and infecting the system, torpig during its main stage, collects and report information like financial data, credit card numbers, banking account credentials, windows password etc. it was estimated that 70 GB data was stolen and redirected.[6]
- *Conficker:* It was first detected on 2008; it was a computer worm mainly targeting Microsoft windows operating system. It propagates by exploiting MS08-67 vulnerability in server service which then gives complete control of the infected machine remotely to attacker. It infected millions of government business, and over 9 to 15 million systems were infected by conficker.11,6
- *Zeus:* It used spam emails and drive-by downloads as its infection methods. It was used to steal banking information. It was early identified on July 2007. It was said that it was the largest botnet on the internet as it was very difficult to detect. It mainly targeted business to capture password. Later it was said that it stopped after the possible retirement of the creator.6,11.

IV. BOTNET DETECTION TECHNIQUES

Botnet detection is one of the imminent tasks of any cyber security. In the past years as the botnets are emerging with new techniques and architecture to make themselves undetectable and cyber security and researcher are also coming up with countermeasures to stop the malicious activities. Spammers are using IP address flux, domain flux, encrypted communication, rootkit techniques to elude tracing. IP address fluxing is used in waledac and storm as an evasion technique. Domain flux is another technique used by conficker, torpig, zeus where in botmaster changes the domain name of C&C server periodically in order to elude tracing and shutdown attempts. 6 Rootkit techniques is another way in which a bot gets access to computer system as an administration and installs the malicious activity and also keeps its presence undetectable. Botnets are coming up with anti-detection techniques and creating havoc across the internet, so for the secure and reliable communication cyber security needs to build up strong defense mechanism. One of the most popular techniques is implementation of intrusion detection system (IDS); it is a software application that monitors a network or system for malicious activity. It is further classified as Network IDS and Host IDS. HIDS is a system that monitors operating system files. NIDS is a system that analyzes incoming network traffic and is placed at some point within a network to monitor traffic to and from all the devices.

In this section, a review of botnet detection mechanism has been carried out. The study is arranged in the following manner:

A. Categories of Detection Methods

1. Signature Based Techniques

It is a detection technique that looks for the specific pattern, such as network traffic and look for the well-known malicious pattern in the payload. It basically generates spam signature and characterize the botnets that perform malicious activities by examining the context and its traffic properties. This technique is widely used for detecting the known attacks, i.e. the pattern or signature is already identified and kept in the log. So basically, it suffers from one limitation and that is it cannot detect new or unknown attacks [5].

2. Anomaly Based Detection

It was later introduced with the purpose of detecting unknown attacks. In this technique a Machine learning approach has been proposed where a model is created, and it is trained to recognize all the activities of a system and network. At initial a model is created and this model goes to two or more phases one of which is training phase during which a profile is created by feeding the model with different traffic pattern and combination of malicious activities and later compared with new behavior against this model in second phase called as testing phase, so bots that belong to same botnet is likely to have similar behavior during their life time, so using this technique we can find behavior that similar to host and try to detect bots by correlating similar behavior. Anomaly based techniques enables detection of previously unknown attack or pattern that has not been registered but it may suffer from false positive, it may result in error while reporting the activity test result may indicate presence of any condition which is not present.

Anomaly detection techniques have been categorized as unsupervised anomaly detection, supervised anomaly detection and semi-supervised anomaly detection [24]. We have a wide variety of application of anomaly detection techniques such as Intrusion detection, fraud detection, system health monitoring, removing anomalous data from data set. In this study we focus on anomaly-based botnet detection approach, the study has been done on the basis of different examples taken from the literature and which is employed using either one of three approaches: static, dynamic or hybrid, which specifies the way in which the information is gathered.

3. Dynamic Anomaly Based Detection

As stated earlier this approach had two main phases namely training phase, also referred as learning phase in which the detector/model learns about the regular behavior of the system or the host and during detection phases also known as monitoring phase, it monitors the program during its execution and will check for the any irregularity or any kind of contradiction with the pattern that has been learned during training phase. During dynamic approach the model tries to detect or determine the malicious activities during program execution or after program execution.

B. Related Works

In this section, different techniques from literature for detection of botnets are presented along with an analysis.

Payload based anomaly detector, *Ke Wang & Salvatore* presents PYAL [12] a anomaly based detection techniques in which the normal application payload of a network traffic is modeled, under unsupervised data and the model is deployed in environment with high bandwidth like firewall, network appliance or target host. PYAL is used for calculating the expected payload for each port or service on the system. Network payload is basically a stream of bytes which does not have a fixed format. To model payload stream of bytes are divided into small clusters based on some criteria and later associated with similar stream. Network service provides port numbers that are fixed integer number ranging from 0-65,535 like 20 used for FTP data, 21 FTP command, 22 SSH, 23 TELNET, 80 HTTP and so on. Each application on a system has its own port number or a service number. Payload model is computed for different length range for each port and service which is a characteristic of a normal payload.

Once the characteristics of normal payload has been identified, system is made to learn about the profile that created during the first phase that is learning phase and it is assumed that, it is a normal behavior or expected payload of the traffic during normal functioning of a system. Anomaly detection phase begins during a learning phase when a Centroid model is computed, (grouping of similar set of objects, in which object in same group is called as cluster, and cluster is represented by central vector, where number of cluster is K which is fixed), which analysis the network traffic and captures incoming payload and test the payload for any inconsistencies from the centroid model using Mahalanobis distance metric (it is distance metric used to compare the similarity between the new payload and the previously computed model). If the incoming payload is too far from the centroid model, which means the Mahalanabis distance value is large, then the payload is malicious, or we can say if any new payload is found to be too far from the expected payload then it is malicious, and an alert is generated and accordingly a decision is taken.

Ke Wang & Salvatore implemented this technique with 1999 DARPA IDS data set on MIT Lincoln lab. The network traffic was evaluated and logged in *tcpdump* format, after 3 weeks of training data and two weeks of test data. Out of 201 different attacks in Lincoln lab, out of which 97 should have been detected but the author's technique could detect only 57 attacks. This technique has been applied only to focus TCP network traffic, so PAYL suffers from certain limitations, as it is unable to detect attacks that have been done using other protocol other than TCP, like UDP, ICMP, ARP and it cannot detect attack that do not contain payload as its model is computed using payload part.

Data mining is the process of knowledge discovery from large amount of data, finding the interesting patterns, analyzing the data and presenting the useful information for the heterogeneous data set. It has a wide variety of applications around the world medical field, market analysis, fraud detection, risk analysis and management and may others.

In the field of intrusion detection, *lee and stolfo* has purposed data mining technique [13], using association rules and frequent episodes. Association rule is considered to satisfy both the minimum threshold and minimum confidence threshold, so its major goal is to derive multiple feature correlation from database table and form a rule set. For the effective intrusion detection enough, training data is required to derive the knowledge. Each record in a database table is a set of items, association rule expression $x=>y$, confidence, support, where x and y are subset of item in record.

Rule set are created that consists different aspect of the target system, normally the knowledge about how system behaves in normal state. *Lee and stolfo* in their work have constructed base classifier that models different aspect of target system. Meta detection agents use information from multiple base detection agents and this information is audit data, which are generally data streams that have been processed for detection purpose, in this case authors have used tcpdump output. Mining based detection techniques is now used to execute rule set on audit data (tcpdump) and after doing the manual analysis of normal behavior and abnormal behavior anomalous data were detected.

Wei Lu and Ali a Ghorbani proposed a network signal modeling technique for detecting network anomalies [14]. Wavelet having specific properties that can be used in signal processing, for conveying information about the behavior or attributes of the sound. Wavelet analysis technique has been widely used for anomaly detection in past, the technique differs in the approach used. The architecture proposed by Weilu and Ali A. Ghorbani consists of some major components as such, feature analysis, normal network traffic modeling based on wavelet approximation and prediction by ARX (Auto Regression and Exogenous) model and intrusion detection. During the process fifteen attributes of network traffic behavior has been identified, which are different from the current existing wavelet-based network anomaly detection technique as it has a smaller number of attributes to define the traffic behavior. Normal daily traffic is modeled based on the attributes of network traffic that have been identified and is represented by set of wavelet approximation coefficients which is predicted by ARX model. The major focus on this paper is identifying and choosing fifteen network flow-based features which characterize the network traffic input function.

Anomaly in a network can be caused by several reasons, as such discussed by *M H Bhuyan* in his paper focuses on security related anomalies is caused by activity of intruder who hijack the bandwidth by intentionally flooding the network by unnecessary traffic [15].

ANIDS, anomaly network intrusion detection system consists of components such as *detection engine*, which is a heart of NIDS, attempts to detect occurrence of any intrusion either online or offline. Anomaly based approach is used to detect unknown attacks based on *matching mechanism.* By continues monitoring of network, a pattern or profile can be built, matching mechanism determines whether the new instance belongs to a known class defined (profile built) or not. *The reference data* another component of ANIDS stores information about known intrusion signature of profile of normal behavior. The storage space is required to store intermediate results, known as *configuration data* (partially created intrusion signature). Any kind of indication received by detection engine is raised by *alarm component* of NIDS architecture, a human analyst is required to interpret and analyze the alarm raised and take the necessary action. A profile needs to be updated by *security manager* as per the decision taken by analyst which is known as pre-processing activity. *Traffic capturing* is the major component of NIDS; the traffic is captured at packet and flow levels, preprocessed and send to detection engine for the process.

Guofei Gu et al.,[16] proposed network-based anomaly detection approach to identify botnet C&C channels. A prototype system botsniffer is made to capture spatial-temporal correlation in network traffic and utilize statistical algorithm to detect botnet using many real-world network traces. In the centralized architecture botmaster uses either push or pull to propagate to reach bots. In push style C&C commands are pushed or sent to bots e.g. IRC based C&C, botmaster has real-time control over the botnet. C&C is loose in pull style because there is a delay in between the time when botmaster issues a command and the time when bot gets the command, e.g. HTTP based C&C. By observing the spatial-temporal correlation and similarity nature of these botnet, a set of heuristics is derived that differentiate C&C traffic and normal traffic. An anomaly-based detection algorithm is implemented to identify both IRC and HTTP based C&C. Botsniffer is implemented as plug and play for open source snort.

Since this technique is IRC and HTTP based, it is restricted to botnet that uses IRC and HTTP protocols. Botsniffers is unable to detect protocol and structure independent botnets.

Botnet group activity detector is based on group activity model and metric. [17] Hyunsang Choi et. al. has proposed a DNS based BotGAD model which consists of four parts:

a) Data Collector,

b) Group classifier,

c) Similarity analyzer

d) Botnet reporter.

IRC, HTTP, P2P are widely used botnet protocols for communication. Incoming and outgoing traffic is monitored at network gateway by data collector, group classifier makes groups from traffic using predefined group size threshold, the similarity analyzer estimates group similarities and botnet reporter summarizes /reports results.

BotGAD implemented in TCP/UDP traffic generates many false positive. Therefore, only DNS-based botnet can be evaluated using this technique, which makes this technique structure and protocol dependent.

An anomaly-based approach is proposed that do not require any previous knowledge of bot signatures, protocols of botnet or any C&C server address that refer to bots properties. Sajjad Arshad et al. [18] has proposed a prototype system to evaluate a real world network traces along with normal traffic to find the presence of bots. The main approach is to cluster bots of similar netflows and find the behavior similarities of host in different properties such as netflow information through a predefined time window and to detect bots by correlating these similar behaviors between different time window.

The architecture for botnet detection approach consists of nine interconnected components that analyze traffic online.

The components work in the following manner

- Traffic dispatching component delivers traffic to Domain-IP Mapping, Netflow Generating and Alert Generating components.
- Domain-IP Mapping component maps the DNS domains to corresponded IPs for filtering purposes.
- Netflow generating component generates TCP netflows between hosts.
- Alert Generating component reports the malicious activities of the hosts like scanning.

The other five components, Alert filtering component filters useless alerts generated by Alert generating component and Netflow Filtering filters the netflows generated by Netflow generating component, at the end of each time window.

The proposed technique is based on the perception that all bots respond to command and perform malicious activates in a similar fashion, but as we know the trending botnets are structure independent and uses different evasion techniques to hide. The major drawback of this technique is that it is unable to detect botnets which refer to different structure.

The result also reflects that the efficiency of netflow clustering algorithm is low.

A method to detect, track and characterize botnets on a large-scale network is found [19]. The method is passive and is invisible to operators. A. Karasaridis, et al., has developed an anomaly-based passive analysis algorithm that detect IRC botnet controllers. It works on different phase, data collection process, a botnet controller, which is evaluated by botnet detection algorithm.

The detection technique is implemented on a large-scale network, which has low false positive rate and can also detect encrypted communication.

The trending botnets are structure and protocol independent which makes them powerful and can easily evade the current detection techniques, therefor the current techniques proposed is structure dependent and evolution of botnets cannot be accommodated.

Botnet detection method analyzes the social relationship between nodes.

In this paper *Jing wang and Ioannis ch* has proposed a method that do not focus on C&C channels of botnets, instead botnet is detected by analyzing the social relationship, modeled as graphs of nodes.

For detection of bots analyzing the social relationship nodes are required, which are modeled as graph of nodes.

For this purpose, Jing *Wang* et.al has proposed two social graphs in his paper:

- Social interaction graphs (SIG) in which two nodes are connected if their interaction between them.
- Social correlation graph (SCG) in which two nodes are connected if their behavior is correlated.

The technique proposed has two main stages, in network anomaly detection stage, SGI model is constructed as a reference model and threshold vale is set. The model is then monitored for nay abnormal SIG. Another stage is botnet discovery, is triggered whenever the threshold value is higher than the earlier set threshold value. Such highly interactive node is referred as *Pivotal nodes*. Botmaster and target frequently interact with bots, which leads to categorize them as *Pivotal nodes.* The interactions between bots and *Pivotal nodes* should be correlated; therefor to make this correlation SGC is constructed. The community that has high interaction with *Pivotal nodes* in the SGC model is likely to be detected as bots.

This technique is applied to the real-world network traffic which resulted in detection with high accuracy. The problem of this method lies in grouping of the botnets based on their interaction and some common behavior, where some botnets groups were misclassified because of its heterogeneous property.

A P2P method for bot detection based on adaptive multilayer feed forward neural network in cooperation with decision tree is proposed by Mohammad Alauthaman et al. [21]. The proposed method works on two fundamental concepts. Firstly, it passively monitors network traffic secondly it utilizes the fact that during propagation phase bot shows frequent communication behaviors with its C&C server.

To increase the performance of the framework they have used network traffic reduction approach. A payload independent, connection-based detection method is used, and the information is extracted from the TCP network traffic.

At the final approach, classification and decision tree is used to select the important features to reduce the size and dimension of dataset and finally the P2P botnet are detected from the extracted dataset by differentiation the botnet traffic from legitimate traffic.

The procession capability of this technique is very good, and accuracy of detection is also achieved as proposed in the paper.

A limitation of the proposed research is that it considers TCP traffic only to detect botnets, so if in case botnets uses UDP packets for communication then this approach may not be able detect the botnets.

A real time botnet detection mechanism which basically is a proactive measure has been proposed in this paper [23]. A payload independent botnet detection approach is used to prepare a dataset from the traffic and based on the classification and clustering algorithm essential features from the traffic is extracted. Implementing expectation-maximization clustering algorithm, a real-time botnet is detected.

A clustering-based detection technique proposed by *Pijus* has a strong underlying theoretical support and inferred from the real world.

As we are aware of the botnets constantly evolving nature and they come up with new techniques of avoiding of avoiding detection, as a result use of this technique may result in false negative.

Real world data consists of sensitive information, generating real world network traces is another drawback in this approach.

V. ANALYSIS

Botnet is a network of computers which works on command and control mechanism, there are different types of botnets available across the internet, and despite being different the activities of botnets are common.

Signature based technique is one of the popular methods for detecting the known attacks, like for those the pattern or signature are already identified and kept in the log file, so this technique was limited within knowing known attacks only. Later anomaly based detection technique was introduced with the purpose of detecting unknown attacks where Machine learning approach is used where initially a model is created and that model is trained to recognize all the activities of a system and network, the dataset that have been used in different methods are different like for the PYAL technique TCP network traffic was analyzed , so this technique was limited to TCP network traffic only as it could botnet detect the attacks that have been done using other protocol other than TCP, similarly the major drawback of anomaly based detection technique lies in the use of such dataset to test the model in such approach, the evaluation of the network traffic also gets limited. Nature of the anomalies that keeps changing is another major challenge for the security personnel to combat the botnets activity. The need of construction of realistic and comprehensive dataset is required for the evaluation of the network traffic in real time, which do not limit under the use of one protocol so that robustness and high performance can be ensured.

Table 1. Summary of Botnets Detection Techniques

Author/Approach	Techniques	Research Gap
Ke Wang et al. [12]	• Clustering • Learning phase • centroid model using Mahalanobis distance metric	• Focuses only on TCP network traffic, • Unable to detect non-TCP based
Sajjad Arshad et al. [18]	• Anomaly based • Net flow clustering algorithm	• Based on intuition that bots respond to command and perform malicious activities in a similar fashion • But the trending botnets are structure independent • Unable to detect botnets which refer to different structure • Efficiency of netflow clustering algorithm is low
Lee and Stolfo [13]	• association rules • frequent episodes	• Restricted to area of application • Cannot be implemented to analyze botnet that targets smart phone.
Hyunsang Choi et al. [17]	• BotGAD	• Based on DNS traffic as a case study. • Limited to DNS traffic • Unable to detect protocol and structure independent botnets.
Wei Lu and Ali [14]	• Auto Regression and Exogenous) model • Intrusion detection.	• Low detection rate
A. Karasaridis, et al. [19]	• An anomaly based passive analysis algorithm •	• Unable to detect protocol and structure independent botnets. • lacks real time botnet detection
Jing wang et. al. [20]	• social interaction graphs • Social correlation graph	• Grouping leads to misclassification of botnets • Suffers false positive
Alauthaman et al. [21]	• neural network • decision tree	• Considers only TCP traffic • Unable to detect botnets that used UDP traffic for communication
G. Gu, J. Zhang, et al. [16]	• Anomaly-based detection algorithms • botsniffer	• IRC and HTTP based C&C • Unable to detect protocol and structure independent botnets. • lacks real time botnet detection
Bhuyan et al. [15]	• ANIDS	• Most NIDSs depends on the environment. • A system or method should be independent of the environment. • The nature of anomalies keeps changing over time • Intruders adapt their network attacks to evade existing.
Pijush Barthakur [23]	• Machine-learning • Classification • Clustering	• Based on theoretical backing • Real-world data are difficult to extract • The nature of anomalies keeps changing over time • May further result in false negative

VI. CONCLUSION

In the recent years, the world has witnessed the massive growth of internet traffic and this traffic is not always flooded by human activity, half of it is botnet activities. It has become very important task for researchers and security personnel to combat these botnets activities. In the past year many techniques have been implemented to stop the vulnerability of botnets.

A very important issue is differentiation of legitimate users and spammers, as such one technique IP blacklisting, many legitimate users were also blacklisted in the process as the bot user and legitimate user may share a common computer. The real challenge is to differentiate a bot user and human and to detect the bot. Bot uses many evasion techniques to make their presence undetectable. Future scope in this area lies on creating a robust technique that does not limit evaluation of the network traffic on one protocol or use of one dataset.

REFERENCES

1. Zhaosheng Zhu,Guohan Lu, Yan Chen, Zhi Judy Fu, Phil Robets, Keesook Han, Botnets Research Survey, 2008 32nd Annual IEEE International Computer Software and Applications Conference, DOI: 10.1109/COMPSAC.2008.205
2. Number of sent and received e-mails per day worldwide from 2017 to 2023 (in billions), URL: https://www.statista.com/statistics/456500/daily-number-of-e-mails-worldwide/ (accessed on 08/05/2019)
3. Meisam Eslahi, Rosli Salleh,Nor Bdrul anuar, Bots and Botnets, An overview of characteristics , detection and challenges,2012 IEEE International conference on control system, computing and engineering. DOI:10.1109/Iccsce.2012.6487169
4. Somayeh Soltani, Seyed Amin Hosseini seno, Maryam Nezhadkamali and Rahmat Budirato, A survey on Real world Botnets and detection Mechanism, International Journal of Information & Network Security (IJINS) Vol.3, No.2, April 2014, pp. 116~127 ISSN: 2089-3299
5. Wazir Zada Khan, Muhammad Khurran Khan, Fahad T.Bin Muhaya, Mohammed Y Aalsalem, A comprehensive Study of Email Spam Botnet Detection, DOI:10.11.09/COMST.2015.2459015
6. Somayeh Soltani, Seyed Amin Hosseini Seno, Maryam Nezhadkamali and Rahmat Budirato, A survey on real world botnets and detection mechanisms, International Journal of Information & Network Security (IJINS) Vol.3, No.2, April 2014, pp. 116~127 ISSN: 2089-3299
7. Bagel(computer worm) URL: https://www.revolvy.com/page/Bagle-%28computer-worm%29
8. An analysis of conficker's logic and rendezvous points URL: http://www.csl.sri.com/users/vinod/papers/Conficker/
9. Storm worm DDOS attack URL:https://www.secureworks.com/research/storm-worm
10. Top 5 Scariest Zombie Botnets
11. URL:https://www.welivesecurity.com/2014/10/23/top-5-scariest-zombie-botnets/
12. Studying spamming botnets using Botlab URL:https://www.usenix.org/legacy/event/nsdi09/tech/full_papers/john/john_html/
13. Ke Wang, Salvator J. Stolfo, Anamalous Payload-Based Network Intrusion Detection, Conference: Recent Advances in Intrusion Detection: 7th International Symposium, RAID 2004, Sophia Antipolis, France, September 15-17, 2004. Proceedings DOI: 10.1007/978-3-540-30143-1_11.
14. Wenkee Lee, Salvator J. Stolfo, Data Mining approaches for Intrusion Detection DOI: 10.1109/SECPRI.1999.766909 ISSN: 1081-6011
15. Wei Lu, Ali A. Ghorbani, Mahbod Tavallee, detecting network anomalies using different Wavlet Basis Functions, communication networks and service research conferences ,2008, DOI:10.1109/CNSR.2008.75
16. Guofei Monowar H. Bhuyan, D. K. Bhattacharyya, and J. K. Kalita, Network anomaly detection, methods Systems and Tools, IEEE Communications Surveys & Tutorials, Vol. 16, No. 1, First Quarter 2014 ISSN: 1553-877X DOI: 10.1109/SURV.2013.052213.00046
17. Gu ,Junjie Zhang BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic 2008
18. Hyunsang Choi, Heejo Lee, and Hyogon Kim, BotGAD: Detecting Botnets by Capturing Group Activities in Network Traffic DOI10.1145/1621890.1621893
19. An anomaly-based botnet detection approach for identifying stealthy botnets. Sajjad Arshad, Maghsoud Abbaspour ISBN: 978-1-4577-2058-1, DOI: 10.1109/ICCAIE.2011.6162198
20. Anestis Karasaridis, Brian Rexroad, David Hoeflin., Wide-scale botnet detection and characterization, 2007.
21. Jing wang, Ioannis ch paschalidis Botnet detection using social graph analysis DOI: 10.1109/ALLERTON.2014.7028482 ,
22. Mohammad Alauthaman,, Nauman Aslam ,Li Zhang, Rafe Alasem,, M. A. Hossain : A P2P Botnet detection scheme based on decision tree and adaptive multilayer neural networks, Volume 29, Issue 11.
23. Sheharbano Khattak, Naurin Rasheed Ramay, Kamran Riaz Khan, Affan A. Syed, and Syed Ali Khayam: A Taxonomy of Botnet Behavior, Detection, and Defense DOI=10.1.1.721.7203
24. Pijush Barthakur, Development of a Real-Time Machine-Learning based Botnet Detection Mechanism, PhD Thesis, Sikkim Manipal University, 2016
25. Salima Omar, Asri Ngadi, Hamid H. Jebur, Machine Learning Techniques for Anomaly Detection: An Overview, International Journal of Computer Applications, Vol. 79 (2), 2013, ISSN: 0975–8887, pp. 33-41

AUTHORS PROFILE

Ms. Jwala Sharma is currently working as Assistant Professor in the Department of Information Technology, DDE, Sikkim Manipal University. Her areas of interest include Computer Network & Security, Data Mining etc.

Dr. Samarjeet Borah is currently working as Professor in the Department of Computer Applications, SMIT, Sikkim Manipal University. His areas of interest include Data Mining, Computer Security, NLP etc.