Identifying Botnets: Classification and Detection

Rishikesh Sharma, Abha Thakral

***Abstract:**The past few years have witnessed the threats caused by the evolving of botnets. It has been found that the nefarious network consisting of contagious systems called as bots are operated by the botmaster. These botnets have been used for malicious activities. This prevailing threat on the internet has led to spam, Distributed Denial of Service (DDoS) attacks, phishing emails, and other cyber-attacks. The detection of such networks is very important keeping the protocols and features they work upon. The paper talks about the various detection techniques that can be adapted to evade the attacks of bots. The huge amount of traffic created by bots can be studied and distinguished respectively to understand the protocols used by the botmaster; which are further used to detect botnets based on the signature and anomaly patterns. The attacks being done from different locations have made it difficult for a botnet to be caught. It has been mentioned that a few networks provide the bots with a nickname using which the detection can be done. The method has been described thoroughly by also specifying how the bot-names of the same network are similar. Nowadays, the number of botnets has increased with a fewer number of trained bots. These network work upon the protocols like Command and Control (C&C), Internet Relay Chat (IRC), HyperText Transfer Protocol (HTTP) and Peer to Peer(P2P). The detection of such networks is being done classifying the traffic and analyzing the spam e-mails alongside the respected IP address. Even the traps of honeynet are developed which motivate the botmaster to take action and get caught. Such honeynet techniques along with the required steps and the necessary precautions are also mentioned in the paper.*

Index Terms: Botnet, Honeynet, IP Address, Network Traffic Classification, Phishing emails.

I. INTRODUCTION

The botnetshave been addressed as the most severe threat in the field of cybersecurity. Once the network is established, the botmaster can easily operate the contaminated systems whenever and wherever desired. The distinguished protocols are used respectively for the communication. The Internet Relay Chat provides a platform for not just sharing text data but also the images as well which can be transmitted between a variable number of bots. Whereas Peer to Peer (P2P) protocol works on the principle of communication among two nodes. This creates a difficulty for detection and even when any node gets detected, it does not affect the whole network as the communication does not involve many systems in this protocol. These contagious systems remain idle until there is any message received by them from the botmaster. Once they get a notification, it works like a request-response cycle and acts efficiently. The botnets always look for expansion and keep targeting the vulnerable hosts that can be infected via Internet.

Revised Manuscript Received on July 02, 2019.

Rishikesh Sharma,CSE Depertement,Amity University Noida, India
Abha Thakral, CSE Depertement,Amity University, Noida, India

The four phases of a botnet cycle have been mentioned in [1]. It explains the commencement of a botnet i.e. the formation phase to C&C leading to the attack of botnet and post-attacks. Initially, the victim system is targeted to contaminate. Once the connection is established, the system becomes a part of a botnet and can be accessed and controlled by the botmaster anytime desired. The attack phase can be understood when the transformed bot starts executing the commands and perform malicious activities. The final phase includes updating the functionality and improvement of techniques from any attack. It is always the best to bust the bot at the primary attacking phase before it can conduct any harm on the system.

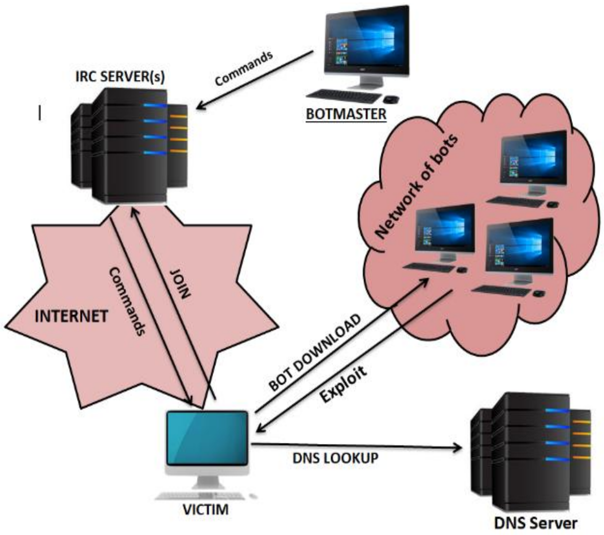

Fig 1Botnet Lifecycle

The main agenda behind the development of a botnet is the financial gain that can be made. This results in the unfair means of making profit leading to fraud. Gameover Zeus (GOZ) is one of the vastest malware attacks on the banks which continued for three years practicing fraudulent activities [2]. The personal information that is collected from the user is shared. The bots also transfer the credentials which are to be very secure. The demand of bots has reached to a certain extent that the bots are now even being rented for a specific purpose like spreading spam emails (eg. Spambot) too which in response collect the data and unsecure the privacy. The evolution of botnets has been exponential which needs to hurdled at the earliest possible with utmost efficiency.

Acknowledging the danger, the researchers have worked upon several detection techniques which are to be discussed in this paper. There has been a lot of contribution to online traffic due to spam/phishing emails. The bot users find it the easiest way for attacking the system and using it for nefarious activities. This can be done in a few different ways but the counterparts for such actions are also been introduced. The detection of bots with the help of IP address has also been proven very efficient.There is also a mechanism using honeynet devices which is a bit risky but provides great results when implemented correctly.

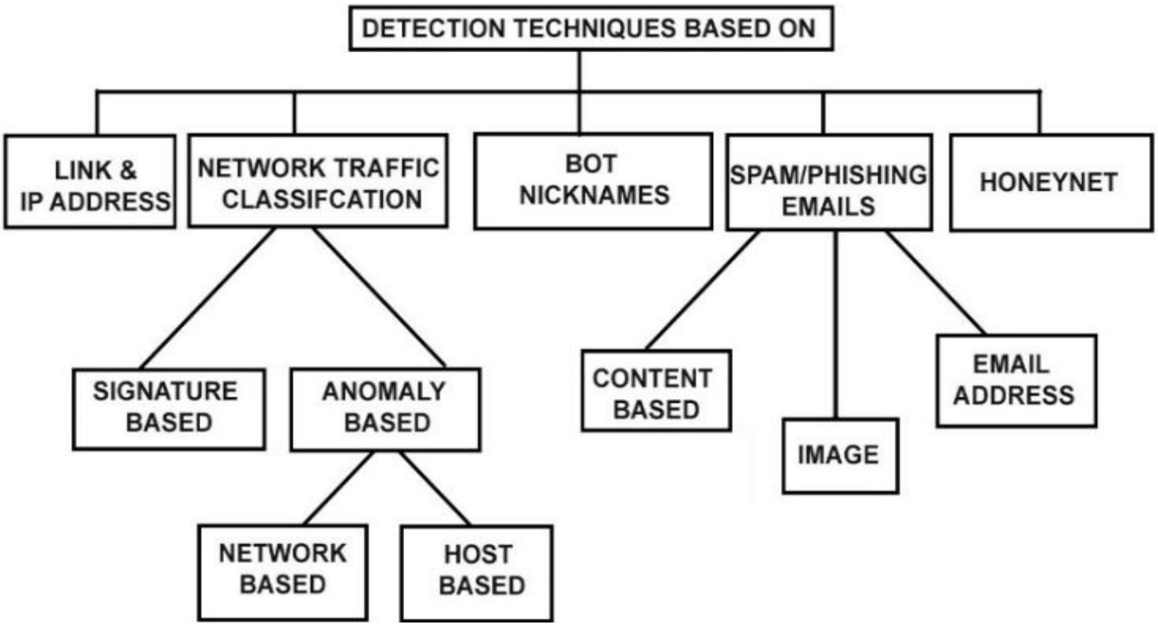

Fig 2 Botnet Detection Techniques

II. LITERATURE REVIEW

There has been a constant concern to the growing threat of botnets. Considering this, a significant amount of work has been done in the past days to resolve the problem by applying different means. Defending against botnets have been a hurdle and a difficulty has been faced limiting the studies intrinsically. The ineffectiveness has still been prevailing in the detection which is explained in [3]. The recent work shows how the IP channel names can be used for the detection while working on the anomaly-based detection [4]. Dagon et al have provided a systematic mechanism for detection based on the responses and query rates on domain names [5]. The unusual nicknames are used to communicate with the bots. Scoring system and n-gram analysis can be taken into consideration for identifying bots in the network traffic. *Rishi* [6] is a software that has implemented it. The identification of nicknames triggers an alarm and even generate a warning e-mail to the system. This can further be used to determine the connected server by which the network traffic can be monitored. The other way for detection instead of verifying the packet contents is to analyze the flow of data through the network. The drawback of such a technique is the absence of payload; a brief survey has been defined in [7]. A thorough guide for the detection of botnets using honeynet has been provided in [8]. The vulnerable parts can be identified providing better scalability. Nepenthes helps in implementing this concept by collecting the spreading malware [9]. This also helps in identifying patterns, trends, and rates of attack which can be further used to analyze the best detection technique to be applied. There have been many algorithms proposed out of which one of them is [10]. It is very efficient in detecting the botnets and characterizing them. Such algorithms can be used to quantify the size of botnets based on the flowing analysis of data on a large scale providing very low false positives.

III. DETECTION BASED ON LINKS AND IP ADDRESSES

Nowadays the command and control of the whole botnet is sold at a very minimal price instead of providing the services and completing the unwilling nefarious jobs individually. The attackers use the concept of redirection to spread malware in the system. The basic idea as discussed in [11] is to detect the flow of traffic and the communication flow which helps in providing an overview of the network system. It has been noticed that they present the spoofed content while indexing and automatically lead to other web pages. The hidden links misdirect the user and one ends up with the contaminated website in the front end which is ready to evolve. A few redirecting embedded links are also sent via scam/phishing emails as discussed. The botmasters benefit from this technique as it offers ease in the field of financial matters, better performance and lesser complications in the management of bots. It is a very quick process divided into several HTTP layers during which the user switches to a different front-end page from the server. The volume and size of the new web page are very low compared to a general one as it is coded just for the redirecting commands.

As the bots in this method are reliable on the constant internet connection, if a single bot goes offline the network can be broken and hence the system can be protected. The involved bots can also be classified with respect to the other attributes like the sequence formed in connectivity, movement of characteristics and DNS comprising record behavior.

By making a list of links and tracking their services, a botnet can be detected. The IP address of such attackers are to be clustered together and their actions are blocked. As there are many layers of redirection and HTTP, the collected IP addresses can lead to the other bots from the pattern noticed. The probing engine is used to detect the trials made for redirection and the NetFlow records are identified. The redirecting URLs are primarily introduced to collect the data and transfer it to the source via the server.

But by applying the logistics for seeking the IP addresses, the links can be blocked and users can be protected. The concepts of machine learning have been efficiently applicable in detecting the botnets via IP flow as well. The research done in [12] shows how the two of the many algorithms completed the detection process successfully.

IV. NETWORK TRAFFIC CLASSIFICATION

There has been a lot of traffic caused by the botnets which needs to be classified for an easy detection. The traffic is created due to the interaction done between the bots for the product of malware. As the threat of botnet has increased rapidly, taking the traffic into consideration the classification of such traffic can be done with the help of machine learning. They can easily function on encrypted network as packet payload is not necessary for the analysis [1]. The activities for extraction of files, developing new processes, evading firewall and antivirus, enhancing the network connection etc lead to such a huge traffic. The contents of packets do not affect the traffic analysis, it is to be analysed between two hosts. There is a huge volume of traffic to be monitored which is not easy to be taken care of. It is very essential to differentiate between the legitimate DNS traffic and botnet DNS traffic after which the unique features of botnet are to be defined. The botmaster uses DNS for rallying as it provides the flexibility but it can facilitate with both invisibility and mobility simultaneously. By using separate algorithms legitimate DNS queries can be differentiated by segregating the important features of Botnet DNS queries [13]. A different way to characterise the traffic is based on packet, burst duration, timing and bandwidth. [14] uses the recent machine learning techniques to distinguish IRC flows and identify C&C traffic using the same.

The detection techniques can be further classified as ‘Signature-Based’ and ‘Anomaly-Based’ and are explained as follows:

A. Signature Based

This technique focuses on the signature used by botnets for detection. It is an easier way which also provides an immediate detection. It also provides a very low rate of false alarm and also unveils more about the kind of attack. It detects the botnet by observing the communication done via command and response between bots. The spatial-temporal correlation indicates the unique key for detection of bots. The patterns formed in traffic also plays a crucial role in detection based on signature. One of the major drawbacks of such a defense mechanism is that it fails to detect the unknown botnets. This method is often used to detect well known bots with an immediate response and negligible false positive results [15]. A major drawback of this technique is that it misses out the detection of botnets with slightly varying signature.

B. Anomaly-Based

This technique focuses on the anomalies found in the network traffic. It functions over the ports of traffic, generated traffic volume, high network latency and unusual system behavior which direct the available bots pertaining in the network. This technique can be further bifurcated into‘network-based’ and ‘host-based’ detection. The anomaly-based technique helps to detect the bots independent of the type of botnet and bots. The base of detection is to differentiate and look for the objects that vary from others. It is very much efficient in finding the unknown botnets as it can identify the anomality present in the network, be it near or far. One of the approaches is to cluster the similar communication and malicious activities pattern. Then the correlation between both is established that is independent of the structure protocols [16]. Abnormally recurring NXDOMAIN replies and the increased rates of query rates also signifies the anomality and can be detected by using different approaches [17]. A problem faced by this detection technique is that IRC networks which are not yet used for attacks may also be acclaimed a botnet. The anomaly-based can further be distributed accordingly:

Network Based

The monitoring of network traffic plays an important role for the detection. By observing the rate of packets injected into the malicious network can help in deciding whether the bot or human is in charge of the session. This works on the fundamentals of command response pattern followed by a cause-effect correlation. The passive monitoring notices the communication between bots that can be suspicious. These monitoring do not contribute in the traffic prevailing in the network.

Host Based

It commences from the user’s anti-viral solutions. Since botnets are so modified that they easily skip the protections and start spreading the malware. Rather than monitoring the external interface, this method strategies by analyzing the internals of a system. The observations are made on the doubted files and processing overheads. If there is any inconsistent or unusual activity observed, the Host notifies the user with the immediate effect. An alert is transmitted when the snapshot of files in the system matches the previous snapshots.

V. DETECTION OF BOTS WITH NICKNAMES

To stop a system from getting any damage, it is vital to stop the attacks of the bot from performing any nefarious activities at the very beginning of a process. Under this heading, we are going to discuss a very unique feature of the bots which can be helpful for tracking one of the most harmful ‘mobile’ bots. One of the interesting attributes about bots is that they have similar names in a single botnet. The botmaster finds it easier to manage communication and collect data in such a way.

This also helps the master to coordinate the big networks and operate them accordingly.

As discussed earlier the bots adapt different ways to communicate and one of the first instructions they receive is addressed by command 'NICK' providing the bot with a nickname with which it will be identified. These nicknames are unique and each bot is designated with a different one. Usually, these names are formed by the concatenation of a few words and numbers together. Botmasters choose them according to the service and destination of the system to be infected. They make sure that there is no redundancy and allot the names with that preference. This can be done by just changing the prefix and suffix to the names for eg. RUS is used for Russia and AUS is used for Australia. By adding different constants to them a new name is formed which is not the same as the previous one. A few networks also consist of the names depending on the operating system they are going to attack like 'XP' and '2K'.

The simplest way to figure out the bots by their nicknames is based on the clustering and listing techniques performed separately under different parameters. The inclusion of special characters, long numbers and abbreviations depict the involvement of bot in the network. The detected bots help in tracking down the source and destination IP addresses used. The detection can be done while the messages are shared by addressing the bots with their respective nicknames. A list is prepared of these nicknames consisting of all the bots. The list is called as the blacklist and helps the system by warning if any threatening activity is observed. On the other hand, a whitelist is prepared as well which consists of the servers that are not detected in the process. These lists keep adding and removing the data in the list as the analysis continues. The nicknames under the data set of the blacklist as well as similar nicknames are blocked by the system to fight against the bots.

The bots with nicknames can also be detected by the tracking of commands which are usually sent by the central server. This can be done as the instructions are conveyed after the primary addressing of the bot by calling it using command 'NICK'. The commands like 'JOIN', 'MODE', 'QUIT' and 'USER' is used to give further instructions. The system receiving these commands are preferred to be infected and can be again used by the botmaster as a host. After the detection is completed, it is important to notify the user about the contaminated system so that the nefarious activities can be terminated then and there. It can either be done by taking charge over the botmaster or generating a mail to the user warning him about the incidents and chances of the system may being used as the host to other bots as well [19].

VI. DETECTION OFSPAM/PHISHING EMAILS

The bots have been used to perform various nefarious activities. Sending spam/phishing e-mails is one of those. It has been found that there has been a huge growth in such e-mails. The spam e-mails have contributed to a large amount of e-mail traffic on the internet. These phishing emails consist of false content and try to manipulate the client. These emails are sent via fake companies whose agenda is to convince the user and try to make the user comfortable so that one provides with his username, account number, password and, other credentials. It is a severe threat to any individual's privacy. The name of big financial business institutes and firms are wrongly used to unveil the personal information and data. The attackers usually prefer this idea to commit fraudulent activities leading to theft. It has also been noticed that a few spams also contain commands hidden within the message which directs to malicious content. The basic idea is that the botmaster wants the emails to be detected and titled as spam. It further results in saving the emails in the spam folder. This provides the attacker with more time and options as spam stays in the folder until it is deleted and helps the master to steal the data and unauthorized information [20]. But yet instead of developing a lot, the methods can be evaded by spreading via indirect ways and performing small activities with a bit more focus. Such recent and updated technologies to wider the network can be stopped by working on the feedback and information applying the defenses other than technical ones [21].

E-mail Traffic generated through spam/phishing mails. The volume of spam has been collected by [22] and the data is used to generate the graphs. The graphs depicted have been drawn on the data collected by 'Statista' (the statistics portal).

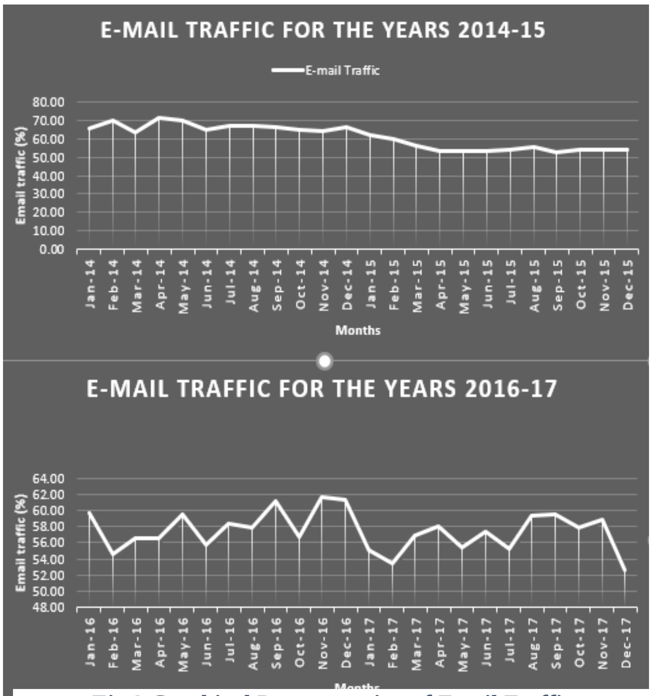

Fig 3 Graphical Representation of Email Traffic

It shows the traffic caused by spam e-mails of all the mails exchanged globally. The data were collected for four years i.e. from 2014 to 2017. It is further divided into a span of two years for further study of the spam e-mails in recent years.

In the years 2014 and 2015, the contribution of emails has beenvarying from 52% to 70%. But in a consecutive couple of i.e. years 2016 and 2017, the contribution of spam in traffic has been constantly compared to the previous years. The maximum contribution of 70% has been reduced to 62% now. The anti-spam techniques have been working efficiently and lowered the traffic but still, more than half of the traffic is being caused due to the spam. The attackers are still moving ahead with their nefarious ideals and the attacks by them have not changed in the last couple of years. But the result shows that the botmasters are facing some difficulties to spread the malware as the botnets are being detected and removed from the server. With the various tracking methods and the continuous study in the field, it is expected to decrease the unwanted activities to the minimal.The data collected by [22] also witnesses that though the decrease has been steady; the number of spam generated recently in the years of 2017 & 2018 are much lesser than the earlier years.

Several anti-spam techniques have been adapted to tackle the unfavorable attacks. But very often legitimate emails are titled as spam and being deleted and are unable to remove bandwidth overload. There have been modifications in such techniques and a few of them are discussed below.

A. Content Based

The analysis of performed surveys ('Zhang et al' in 2004, 'Sahami et al' in 1998, 'Graham' in 2002 and 'Drucker et al' in 1999) [23], [D] depict that there is a particular pattern being followed along with the usage of a few selected words. The embedded text used in the content of body and subject is used very specifically for the spam e-mails. A dataset of keywords is prepared and compared with the content of the message sent in the mail. A database is formed by selecting the preferred terms used in spam e-mail. The punctuations used are recorded along with the size of a standard mail and comprising vocabulary. The datasets of SpamArchive and SpamAssassin provide with one such sample. The attackers place different methods to confuse human understanding by applying several ways. It has been found that bots are taking the help of special characters and symbols like $, #, ~,etc to avoid the unique keywords from getting caught. Even the tags of HTML are used for hiding from the detection. The words are commonly spelt wrong so that they do not match with the list of vocabulary prepared for comparison. These lists keep on modifying and updating with every new text in the mail. The words are not repeated in the list and the matter of redundancy is taken care of. The technique of indexing is adapted by users to decrease the size of vocabulary which forms the set of pointing words. It includes the frequency of words used, the related terms in the file and related attributes.

Tracking such spam/phishing e-mails is a tedious task as it is difficult to prepare a dataset with discriminant attributes. There have been a few sets of data provided by the above mentioned, but no such accurate sample exists which is reliable and can be worked upon. *Nigam & McCallum* in 1998, *Graham* in 2002 and *Sahami et al* in 1998 have applied the theory of Naïve Bayes in their classification of texts. The botmasters have not stopped at this form of phishing mails. Many mails consisting of image attachments have also been unfolded. These are the text embedded images that are sent containing the links and messages. The purpose of such mails is to inculcate the idea of forgery acts as discussed.

B. Images

Treating of text embedded into images for malicious purposes is growing rapidly among the spammers. Such a genre of spam is considered to be one of the most harmful mails. These are sent with the name of big firms and commercials to manipulate the users. The analysis of content can be done in the same way considering that it can be inferred as the message in the body of emails without images. But the attackers have upgraded themselves and now tend to imply various methods that hide such messages by providing the same background colour as the font of the written content or presenting the links in a way that it is confused with the printed caption. The main agenda is to escape from the human readings and provide a platform for them to fall. This kind of trending spam has led to various spam filtering techniques. Firstly, the text from the images needs to be extracted and added in the previously available datasets. Then the often-used keywords and links are segregated and stored in different columns. These databases need to be updated every time any new spam e-mail is detected. Even the pattern and format of the images are found to be similar and recorded for the reference. Each text embedded image is compared with the prepared samples and collected data to detect any phishing activity planned by the attacker. Although this way of tracking bots has not been very efficient as it is able to detect the identical images but not all the similar ones. Even the developed software like 'ABBYY FineReader 7.0 Professional' has failed to provide accurate results of resolution like attributes [25]. The indexing of anti-spam filtering has been done quietly well but still, a lot of study has to be done upon it. [26] has provided a feature set that is able to classify spam and ham emails with a high rate of accuracy exceeding 90% also providing the algorithms for detection within a very short period of time. It is an efficient way that has been proposed while being effective and practical maintaining the overall performance.

C. Email Address

Another way of tracking the spam mails is by observing the sender's address and blacklisting them. It has been found that there are a few big networking sites which lead to most of such communications. By selecting such addresses and making a list out of them can simply block phishing e-mails to even enter the user's inbox. Even the addresses of detected spam by other techniques can be added to the same list and revised time and again. The emails sent from such identities are straight away directed towards the spam container. If a user still wants to refer it, one may do so. The idea of figuring the unwanted emails by this method has been used at a very large scale globally and has succeeded to secure one's information from the botmasters.

VII. HONEYNET DETECTION

A honeynet is a very simple networking system without any authorized services. It is a real network composed of a few computer machines and real servers. Each machine in the network is called a honeypot which appears to be a legitimate system. These honeypots are formed in such a way that they are found to be vulnerable systems and invite the attackers to perform their designated jobs. The honeynet works as a trap or a baiting platform with the primary purpose of inviting hackers to attack the system so that hacking activities can be monitored and their methods and patterns can be studied. The data is captured and controlled in multiple layers with each layer having a specific task as proposed in [27]. The multi-layer protocol is opted to ensure that the data is stored carefully without getting caught. Applying the logic of clustering and segregating the data according to different characteristics, the botnets can be stopped and prevented from performing any threats to the system [27]. The information can be helpful to protect the system from being infected and hence getting protected from the bots. Another important function of honeynets is that they divert the botnets from indulging with the true systems and the valuable resources are conserved from getting wasted. Once a botnet is trapped, the attention of the botmaster can easily be deflected and the further process can take place. Many a time only a single server is used to provide the honeynet structure. If a honeynet gets accessed at any level, it can be assumed that it has been attacked by the bot as there is no authorization provided to the server for any use. Studying their actions, the basic behavior can be analyzed and the botnet can be tracked.

Using honeynet system detection also has a high risk of getting the non-honeynet systems attacked. To gather more information and data the freedom is provided to the bot to perform its actions without any hindrance. But as mentioned in [28] the more freedom provided; the more risk arises. A bot may use the system as a host and attack other computers to gain their control. A honeynet system needs to be agile as well as swift to confront the botnet as soon as the required data is collected. This can be done by adapting a layer of mechanisms together and confuse the bot. It helps to gather as much data as possible and block the actions of bot immediately. The work is done in numerous layers to protect the server from getting any sort of damage.

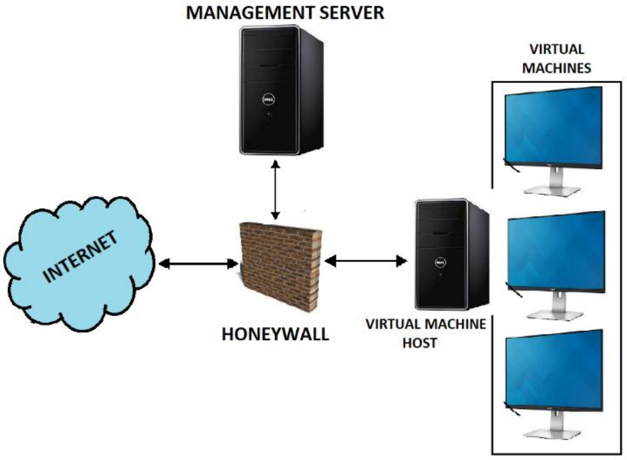

Fig 4 Honeynet

The honeynet structure is designed in such a way that instead of the legitimate systems; the botmaster is made to communicate with the virtual machines. The host of the virtual machine acts as a medium and connects the machine withthe honey-wall. This honey-wall makes the attacker believe that the system is not a trap but a favorable victim. The management server keeps recording all the functionalities and helps in tracking the botmaster. The required information can be gathered in the span of till the connection is established via the internet.

The ‘honeynet’ detecting mechanism provides with the signature of bots that can be used for content-based detection, all the required information can be extracted, the motive behind the attack and the protocol that has been adapted to infiltrate the system. If a honeynet gets revealed and the bot is able to detect it, the bot can very easily escape the bait without giving any sort of data or information. The bot may even provide the pseudo data to mislead the honeynet and puzzle the stored database. If not operated carefully a honeynet can lead to the reverse results which can be highly devastating for the systems as well as the known data.

VIII. CONCLUSION

In this paper, the various ways by which botnet can be detected are discussed. The detection techniques have been classified according to the methods adapted by botnets. Using such techniques help to escape from the bot-attacks and protect our system from being contaminated. There have been a number of methods proposed above to protect the personal data and information. It is very important to be safe from such malicious tasks increasing in the cyber world. A thorough understanding of the concepts has been done to provide the most efficient techniques to tackle the botnet attacks. Henceforth, we have summarized the detection based on the environment and attributes by which the botnet has adapted the growth. On the contrary, the botmastersare developing themselves to launch the attacksvia new methods. The mentioned approaching ways from botmasters have led to various cyber-crimes and are also infiltrating the privacy of a user by collecting the credentials. Hence, each and every user must be updated and ready to tackle such attacks for being secured by taking the required necessary actions. The paper clarifies the phenomenon of a botnet and the various detection techniquesusing which a botnet can be identified.

REFERENCES

[1] Zhao, D., Traore, I., Sayed, B., Lu, W., Saad, S., Ghorbani, A., &Garant, D. (2013). Botnet detection based on traffic behavior analysis and flow intervals. *Computers & Security*, *39*, 2-16.

[2] https://www.knowbe4.com/gameover-zeus

[3] Abu Rajab, M., Zarfoss, J., Monrose, F., & Terzis, A. (2006, October). A multifaceted approach to understanding the botnet phenomenon. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement* (pp. 41-52). ACM

[4] Karasaridis, A., Rexroad, B., &Hoeflin, D. A. (2007). Wide-Scale Botnet Detection and Characterization. *HotBots*, *7*, 7-7.

[5] Dagon, D. (2005, July). Botnet detection and response. In *OARC workshop* (Vol. 2005).

[6] Goebel, J., &Holz, T. (2007). Rishi: Identify Bot Contaminated Hosts by IRC Nickname Evaluation. *HotBots*, *7*, 8-8.

[7] Sperotto, A., Schaffrath, G., Sadre, R., Morariu, C., Pras, A., & Stiller, B. (2010). An overview of IP flow-based intrusion detection. *IEEE communications surveys & tutorials*, *12*(3), 343-356.

[8] Know you Enemy: Tracking Botnets (https://www.honeynet.org/papers/bots)

[9] Baecher, P., Koetter, M., Holz, T., Dornseif, M., &Freiling, F. (2006, September). The nepenthes platform: An efficient approach to collect malware. In *International Workshop on Recent Advances in Intrusion Detection* (pp. 165-184). Springer, Berlin, Heidelberg.

[10] Karasaridis, A., Rexroad, B., &Hoeflin, D. A. (2007). Wide-Scale Botnet Detection and Characterization. *HotBots*, *7*, 7-7.

[11] Sperotto, A., Schaffrath, G., Sadre, R., Morariu, C., Pras, A., & Stiller, B. (2010). An Overview of IP Flow-based Intrusion Detection. *IEEE Communications Surveys and Tutorials*, *12*(3), 343-356.

[12] Haddadi, F., Morgan, J., Gomes Filho, E., &Zincir-Heywood, A. N. (2014, May). Botnet behaviour analysis using ip flows: with http filters using classifiers. In *Advanced Information Networking and Applications Workshops (WAINA), 2014 28th International Conference on* (pp. 7-12). IEEE.

[13] Choi, H., Lee, H., Lee, H., & Kim, H. (2007, October). Botnet detection by monitoring group activities in DNS traffic. In *7th IEEE International Conference on Computer and Information Technology (CIT 2007)* (pp. 715-720). IEEE.

[14] Strayer, W. T., Lapsely, D., Walsh, R., &Livadas, C. (2008). Botnet detection based on network behavior. In *Botnet detection* (pp. 1-24). Springer, Boston, MA.

[15] Zeidanloo, H. R., Shooshtari, M. J. Z., Amoli, P. V., Safari, M., & Zamani, M. (2010, July). A taxonomy of botnet detection techniques. In *2010 3rd International Conference on Computer Science and Information Technology* (Vol. 2, pp. 158-162). IEEE.

[16] Gu, G., Perdisci, R., Zhang, J., & Lee, W. (2008). Botminer: Clustering analysis of network traffic for protocol-and structure-independent botnet detection.

[17] Villamarin-Salomon, R., &Brustoloni, J. C. (2008, January). Identifying botnets using anomaly detection techniques applied to DNS traffic. In *2008 5th IEEE Consumer Communications and Networking Conference* (pp. 476-481). IEEE.

[18] Feily, M., Shahrestani, A., &Ramadass, S. (2009, June). A survey of botnet and botnet detection. In *2009 Third International Conference on Emerging Security Information, Systems and Technologies* (pp. 268-273). IEEE.

[19] Li, C., Jiang, W., & Zou, X. (2009, December). Botnet: Survey and case study. In *innovative computing, information and control (icicic), 2009 fourth international conference on* (pp. 1184-1187). IEEE.

[20] Singh, K., Srivastava, A., Giffin, J., & Lee, W. (2008, June). Evaluating email's feasibility for botnet command and control. In *Dependable Systems and Networks With FTCS and DCC, 2008. DSN 2008. IEEE International Conference on* (pp. 376-385). IEEE.

[21] Dittrich, D., & Dietrich, S. (2008, October). P2P as botnet command and control: a deeper insight. In *2008 3rd International Conference on Malicious and Unwanted Software (MALWARE)* (pp. 41-48). IEEE.

[22] https://www.statista.com/statistics/420391/spam-email-traffic-share/

[23] McCallum, A., & Nigam, K. (1998, July). A comparison of event models for naive bayes text classification. In *AAAI-98 workshop on learning for text categorization* (Vol. 752, No. 1, pp. 41-48).

[24] Sahami, M., Dumais, S., Heckerman, D., & Horvitz, E. (1998, July). A Bayesian approach to filtering junk e-mail. In *Learning for Text Categorization: Papers from the 1998 workshop* (Vol. 62, pp. 98-105).

[25] FineReader, A. B. B. Y. Y. 7.0 Professional

[26] Dredze, M., Gevaryahu, R., & Elias-Bachrach, A. (2007, August). Learning Fast Classifiers for Image Spam. In *CEAS*(pp. 2007-487).

[27] Thonnard, O., &Dacier, M. (2008). A framework for attack patterns' discovery in honeynet data. *digital investigation*, *5*, S128-S139.

[28] Spitzner, L. (2003). The honeynet project: Trapping the hackers. *IEEE Security & Privacy*, *99*(2), 15-23.

AUTHORS PROFILE

Mr. Rishikesh Sharma is currently a student of Computer Science Engineering from Amity University Uttar Pradesh, Noida. His research interests include the field of Computer and Network Security. He is a member of the IEEE student branch.

Ms. Abha Thakral is currently working as Assistant Professor on Grade III with Department of Computer Science and Engineering at Amity University Uttar Pradesh, Noida. She is also a Research Scholar working in the field of Networking.