A COMPREHENSIVE STUDY ON BOTNET AND ITS DETECTION TECHNIQUES

P.Panimalar[1], Dr.K.Rameshkumar[2]

Abstract- A botnet is a group of compromised computers also called bots or zombies which are controlled by the botmaster's malware code. Botnets have become one of the most malicious threats over the Internet. Any unprotected computers in any locations in the world may become bots as long as they are connecting with Internet and have been infected by botnet malware code. The master computer communicates with its bots by a command and control (C&C) channel, which passes commands from the botmaster to bots, and transmits stolen information from bots to their master. The attacked bots can also infect other computers enabling them to be botnet members. This survey paper provides useful information about botnet and its various detection techniques.

Keywords : Active technique , Botnet, Passive technique , Spam

I. INTRODUCTION

A Botnet, or the army of bots (zombies), is comprised of more than thousands or tens of thousands of compromised computers. Bot detection systems can be classified into two categories, i.e., passive techniques and active techniques. The working scenario of a botnet can be classified into two phases. One is the infection phase and another is the attack phase. In the infection phase, a botherder or botmaster tries to expand the size of its army of bots. The bot herder commands existing bots to compromise more user's computers. It is common that the traffic generated by bots are mixed with regular network traffic. To improve botnet detection efficiency, existing researches often reduce the amount of input traffic by filtering out bot-irrelevant traffic. Hence, a bot detection algorithm is able to concentrate only on bot traffic. A good traffic reduction algorithm may improve the overall system performance. However, if not well designed, it could increase false negative rates and/or false positive rates [1]. In the attack phase, a botherder sends commands to compromised hosts, i.e., the bots. On receipt of the commands, each bot launches various tasks based on the instructions embedded in the commands.A botherder is therefore able to ask bots to collect valuable information, report botnet status, and launch attacks to target hosts. The prevalence of botnets, which is defined as a group of infected machines, have become the predominant factor among all the internet malicious attacks such as DDoS, Spam, and Click fraud. The number of botnets is steadily increasing, and the characteristic C&C channels have evolved from IRC to HTTP, FTP, and DNS, etc., and from the centralized structure to P2P and Fast Flux Network Services.The following fig :1 shows the general life cycle about botnet:

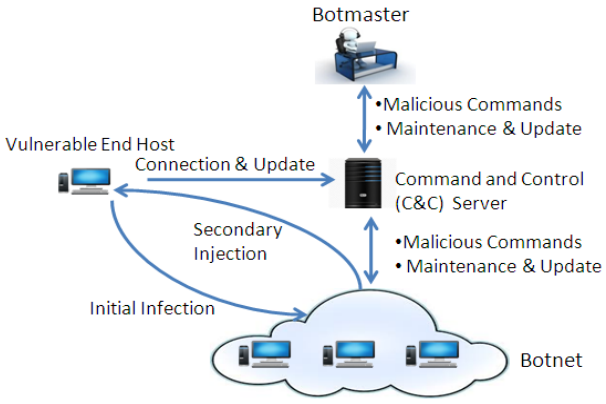

Fig: 1 A General Botnet Life Cycle

II. BOTNET AND ITS ORIGIN

In order to discuss more about botnet it is required to know about bot, its origin [2], and effect of botnet in real world situation. It is also required to know that how an attacker changed the trend to use botnet
in searching or in providing better service to users and used bot as malicious activity more. First bot is created in 1988 but it was not as a malicious activity. First malicious bot is created in 1998 and after 1998; much more type of botnet is traced as a dangerous activity to online user.

[1] *Research Scholar, Research and Development Centre, Bharathiar University,Coimbatore-41046,inidia,*

[2] *Research Supervisor, Research and Development Centre, Bharathiar University, Coimbatore-641046, India,*

International Journal of Computer, Information Technology & Bioinformatics (IJCITB) ISSN: Volume-1, Issue-112 measured the botnet based infections as follows:

1. Mariposa Botnet infected 10.3 million computers all over the world.
2. Zeus Botnet has the report of infecting over 3.6 million computers in the United States.

A. Botnet Inventions Made So Far

1. The first bot (non malicious bot) IRC was invented in August of 1988 by Jarkko Oikarinen from University of Oulu, Finland.

2. In 1989, Greg Lindahl, an IRC server operator, created the benevolent bot called GM which would play a game of hunting the Wumpus with IRC users.

3. Eggdrop (non malicious bot) created by Jeff Fisher for assisting IRC channel management in 1993 has significance.

4. The first malicious bot, GT-BOT, found in April 1998 and now there are at least a hundred variant of GT-BOT are available which include IRC client, mirc .exe as part of bot.
5. Pretty party worm was the first worm which emerged to make use of IRC as a means of remote control in June 1999.

6. In April 2002, Agobot's source code was published on many websites and slapper was the first worm with P2P communications. So many more different types of Botnets are measured after 2002, on various communication techniques. And now recently Zeus Botnet [2009], spy eye [2010], Mariposa [2009], Asprox(a P2P Botnet) [2009].

7. The following bots (Bobax, Torping, Trojan, Donbot, Mega-D, Grum, Maazben, Onewordsub, Cheg, Wopla, Xarvester, Spamthru, Rambot, Internbot, Akbot, Gumblar, Social bot and Decbot) are all belonging to the same category of Agobot.

B.Botnet Key Terms and Definitions

1. *Attacker: It is the one that configures the bot and* comprises a machine to install a malicious bot, controls and directs the bots once it joins the designated IRC channel.

2. Bot*:* It is typically an executable file, capable of performing a set of functions, each of which could be triggered by a specific command. A bot when installed on victim machine copies itself into configurable install directory and changes system configuration each time system boots.

3. *Control channel:* It is a private IRC channel created by the attacker as rendezvous point for all the bots to join once they are installed on infected machine which are online. This control channel comprises a channel name and a password key to authenticate.

4. Victim machine: It is the compromised internet host on which the malicious bot is installed after the attacker has exploited an application or operating system vulnerability or has duped the user into executing a malicious program. Once infected the target host are also referred to as Zombies.

5. IRC server: It is a server providing IRC services, this could be a legitimate public service provider like DALNET etc. or another attacker's compromised machine to perform attack.

III. BOTNET DETECTION USING PASSIVE TECHNIQUES

This group of passive measurement techniques consists of those where data is gathered solely through observation [3]. Through monitoring, activities can be tracked without interfering with the environment or tampering with the evidence. This makes these approaches transparent and, in many cases, their application can be hidden from botmasters.

A.Packet Inspection

A popular concept for increasing a network's security is to inspect the network data packets. The basic idea is to match various protocol fields, or the payload of a packet, against pre-defined patterns of abnormal or suspicious content. This may, for instance be a packet containing shell-code sequences used to spread malware, communication with an Internet address that is a proven host of malicious content, or a file server that suddenly begins to communicate via the chat protocol, IRC. These patterns are also called 'detection signatures'.

B. Analysis of Flow Records

Analysis of flow records can be considered as a technique for tracing network traffic at an abstract level. Instead of inspecting individual packets, as described in the previous section, communication streams are considered in an aggregated form.

In this context, a flow record consists of several properties that describe a network data stream. Typical attributes are the source and destination address, the related port numbers and also the protocol used inside the packets, the duration of the session and the cumulative size and number of transmitted packets.

C. DNS Based Approaches

When a host has been compromised by a botnet, communication has to be established to either a commanding server or other infected hosts, depending on the botnet infrastructure. This requires the integration of a communication protocol into the malware. Two ways of specifying a firm contact point are available for this purpose:

- Fixed IP addresses can be integrated into the bot,executable upon distribution.
- A domain name (or set of domain names) can be defined that will be contacted when the host system is compromised.

Use of a domain name offers flexibility in various ways. First of all, one domain name may be associated with multiple IP addresses, helping to create a redundant architecture that is more robust against machine takedowns. These IP addresses do not have to be static but can be changed dynamically on demand.

D. Honeypots

A 'honeypot' is an intentionally vulnerable resource deployed in a network with the aim of soliciting attacks or even compromise by a malicious entity. Although this definition seems rather wide, it is necessary to capture the versatility of honeypots and how they fit with the various ways this concept has been realised.

Two features are primarily used to distinguish between the two categories of honeypots: client and server honeypots and low interaction honeypots. The main reason for researching and developing honeypots is to discover new information about the practices and strategies used by creators of malware and hackers. In general, two kinds of information can be gathered by honeypots:

- Types of attack vectors in operating systems and software used for attacks, as well as the actual exploit code which corresponds to them.
- Actions performed on an exploited machine. These can be recorded, while malware loaded onto the system can be preserved for further investigation.

IV. BOTNET DETECTION USING ACTIVE TECHNIQUES

The group of active measurement techniques contains approaches that involve interaction with the information sources being monitored. Although this enables the performance of deeper measurements [3], their application may leave traces that influence results, or include activities that can be observed by the botherder. This can cause reactions, such as a DDoS attack against the analyst or the introduction of changes to the botnet structure that will complicate measurements, even including migration of the service to evade monitoring.

A. Sinkholing

In general, the term 'sinkholing' describes a technical counter measure for cutting off a malicious control source from the rest of the botnet. It can be performed against a variety of targets, most likely against botnets command-and-control servers or trojan dropzones.

One of the most common variants of this technique is changing the targeted malicious domain name so that it points to a machine controlled by a trusted party, usually investigators or researchers.

The basic principle of this measurement relies on the fact that the domain is very likely to be contacted by infected machines trying to reach their command-and-control server. As the domain names and IP addresses used for such purposes often only serve malicious purposes, this approach tends to have a low false-positive rate.

B. *Infiltration*

The 'infiltration' of botnets can be divided into software and hardware based techniques. The first covers research on the bot executable and monitored traffic to achieve control and conduct measurements. The latter can be applied if access to the command-and-control server is possible and may be used to wiretap the communication. This includes physical machines as well as virtual machines that might be running in a data centre.

C. *DNS Cache Snooping*

The measurement technique called DNS Cache Snooping [6] is based on the caching property implemented and used by many DNS servers. If a DNS server is queried for a domain for which it has no entry defined, it will issue a query towards the responsible authoritative name server on behalf of the querying client and store the resulting data record afterwards in a local cache. Caching is mainly used to increase the performance of a name server and reduce its traffic load.

D. Anamoly Based Detection

In [4] Wang et al. use an n-gram feature on payload for detection purpose. N-gram is a sequence on n adjacent bytes. For modeling a payload it is first classified into clusters according to some criteria like using port numbers or length.

A payload model is computed for payloads of different lengths from same port for each direction of payload flow. A sliding window of size n is passed over the payload and the occurrence of each n-gram is counted. When n=1, the average byte frequency of each ASCII character 0-255 is obtained. In addition to this mean value, the standard deviation and variance is also calculated. Thus a set of payload models M*xy* is computed where M*xy* stores the average byte frequency and standard deviation of each byte's frequency of a payload of length *x* and port *y*. During detection, each incoming payload is scanned and its byte value distribution is computed.

This newly computed payload distribution is then compared against model M*xy*; if the distribution of the new payload is significantly different from the normal, the detector flags the packet as anomalous and generates an alert.

E. Spam Emails

The techniques based on spam emails usually analyses the email patterns and may also derive certain features of these emails such as sender/recipient address etc. In [5] the authors proposed a method called EsBod which is an email shape based botnet detector which will classify the email into spam or real email.

The shape generator of the Esbod will extract the skeleton of the email that is fed into the system. Using Gaussian kernel density estimator the shape (or template) of the email is derived from the skeleton. The classifier of the EsBod will takes in this derived shape and matches with the botnet signature repository using Hellinger distance.

In [6], Brodsky et al proposes a distributed content independent spam classification method called Trinity which uses source identification along with a peer-to-peer based distributed database. Trinity first determines the source IP of the received email and then updates the database using this IP.

The database is checked for past traces of email sources and number of emails that source recently send within a fixed period and a score is obtained. This score can be used for classification by the MUA (mail user agent). If the score is high and if the sender is not in the sender/recipient address book, then the email must be a spam.

A graph based approach is proposed in [7] by the authors. In this approach large user-user graphs and tightly connected subgraphs are drawn which detects botnet spamming attacks targeting major Web email providers.

V.CONCLUSION

In this comprehensive study on Botnet, readers can gain a deep understanding about the basic definition about bot and botnet and also the methods used so far to detect the botnet. As far as botnet is concerned is also discussed in detail with appropriate diagram to gather much more information about botnet architecture. Hope after reading this survey paper fully, all the readers those who are interested to go for further research to find out new algorithms can utilize this paper as a useful one.

REFERENCES

[1] K. Wang, C.-Y.Huang, S..-J.Lin, "A fuzzy pattern-based filtering algorithm for botnet detection", Computer Networks- 2011.

[2] *Ravi Kishore Sharma, Gajendra Singh Chandel,* "Botnet Detection And Resolution Challenges:A Survey Paper", International Journal of Computer, Information Technology & Bioinformatics (IJCITB) ISSN: Volume-1, Issue-1.

[3] *Daniel Plohmann, Elmar Gerhards - Padilla, Felix Leder, " B*otnets: Detection,measurement, Disinfection & Defence", European Network and Information Security Agency (ENISA) 2011.

[4] *K. Wang, S.ttolfo* "Anomalous payload-based network intrusion detection", in: Proceedings of the 7th

International Symposium on Recent Advances in Intrusion Detection (RAID), (Sophia Antipolis, France), 2004.

[5] *P. Sroufe, S. Phithakkitnukoon, R. Dantu, J. Cangussu,* "Email shape analysis for spam botnet detection", in: Sixth IEEE Consumer Communications and Networking Conference, (Las Vegas, NV), January 2009, pp. 1–2.

[6], *A. Brodsky, D. Brodsky,* "A distributed content independent method for spam detection", in: Proceedings of the First Conference on First Workshop on Hot Topics in Understanding Botnets, (Cambridge, MA, 2007), p. 3.

[7] *Y. Zhao, Y.L. Xie, F. Yu, Q.F. Ke, Y. Yu, Y. Chen, E. Gillum,* "BotGraph Large-scale spamming botnet detection", in: *Proceedings of the 6th USENIX Symposium on Networked Systems Design and Implementation*, 2008, pp. 321–334.

[8] Maryam Feily , Alireza Shahrestani and Sureswaran Ramasdass, "A Survey of Botnet and Botnet Detection" , IEEE Computer Society Third International Conference on Emerging Security Information, Systems and Technologies- 2009.

[9] Hossein Rouhani Zeidanloo, Mohammad Jorjor Zadeh Shooshtrai, Payam Vahdani Amoli, M.Safari, and Mazdak Zamani, "A Taxonomy of Botnet Detection Techniques", IEEE 2010.

[10] Sukhdilpreet Kaur and Amandeep Verma, " Design of Generic Framework for Botnet Detectionin in Network Forensics", International Journal of Computer Science and Information Security (IJCSIS) 2010.

