National Conference on Recent Research in Engineering and Technology (NCRRET-2015) International Journal of Advance Engineering and Research Development (IJAERD)
e-ISSN: 2348 - 4470 , print-ISSN:2348-6406

Botnet and Detection Techniques: A Review

Binal Panchal
M. Tech Student
Department of Computer Engineering
MEC, Basna.
bebo.ce24@gmail.com

Hardik Upadhyay
Assistant Professor
Department of Computer Engineering,
GPERI, Mehsana.
hardik31385@gmail.com

Abstract- Among the diverse forms of malware, Botnet is the most widespread and serious threat which occurs commonly in today's cyber-attacks. A botnet is a group of compromised computers which are remotely controlled by attacker to launch various network attacks, such as DDoS attack, spam, click fraud, identity theft and information phishing. Botnet has become a popular and productive tool behind many cyber-attacks. The defining characteristic of botnets is the use of command and control channels through which they can be updated and directed. Recently malicious botnets evolve into HTTP botnets out of typical IRC botnets. Data mining algorithms allow us to automate detecting characteristics from large amount of data, which the conventional heuristics and signature based methods could not apply.

Index Terms- Botnet, Botnet detection, HTTP Botnet, Data Mining technique

---

1 Introduction

The improvement and advancement in network bandwidth and computing, parallel and distributed computing are widely accepted. So they have been the obviously targeted by hackers [1]. Botnet is a collection of internet-connected computers whose security defenses have been breached and control ceded to a malicious party, blackhat community. The groups of compromised computers are controlled by one or group of attacker known as "Botmaster" [2]. Botnet operators can use the aggregated power of many bots to exponentially raise the impact of those dangerous activities. A single bot might not be a danger for the Internet, but a network of bots certainly is able to create huge malfunctioning. A study shows that, on a typical day, about 40% of the 800 million computers connected to the

Internet in a botnet in year 2008 [3]. Communication, resource sharing and curiosity have been great motivators for underground research and hacking. The major attacks under Botnet arez, DDos, Scanning, Phishing, Click fraud, spamming [4].

A. History of Botnet

The evaluation of Botnet has enhanced it form single machine to distribute in network. The history of undertaking botnets for destruction roughly dates back to 1990. Prior to this, botnets were the major sources of maintaining control of the IRC channels. In 2003, hackers of Oregon State of U.S. controlled 20,000 botnet hosts and started a DDoS (Distributed Denial of Service) attack on eBay [5]. P2P botnet appeared in 2004. The program itself included the client of P2P and linked to servers adopting Gnutella, and used WASTE to do communication. In 2005, a new type of botnet network virus Zotob started its DDoS to attack many websites of U.S. famous companies [6]. In Feb. 2006, CBI and Microsoft found out that hosts been infected by botnet numbered around 57,783 and up to 88,136 until Sep. 17 after a thorough check. Within only half a year, the number of new infected hosts was up to 30,000 [1]. The [7] Kraken botnet was the world's largest botnet as of April 2008, has infected at least 50 of the fortune 500 companies and grew over 400,000 bots. Damballa has released instruction to remove it, recovered 495,000 bots in Kraken Botnet. A new version of the Zeus botnet, on 5th Oct-2012, was used to steal about $47 million from European banking customers [8]. On 8th Oct, 2012 Dorkbot worm IRC based, People have been allegedly receiving automated messages from what appear to be automated bots. Section 1 gives an introduction of botnet and it's History. Section 2 reviews botnet life cycle, topologies, types, type of targets and attacks and C&C channel. Section 3 introduces detection technique and comparison. In section 4, resent trends of research and detection techniques are considered and compared.

2 Background

The untraceable feature of coordinated attacks is just what hackers/attackers demand to compromise a computer or a network for their illegal activities. Once an attack is initiated by a group of computer nodes having different locations controlled by a malicious individual or controller, it may be very hard to trace back to the origin due to the complexity of the Internet [9]. Because of these reasons it has become very serious problem nowadays.

2.1 Command and Control Centre

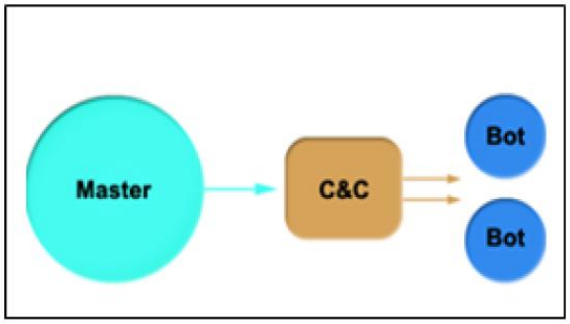

Fig. 1- Botnet Command & Control Architecture

Botmaster will be hiding behind C&C and sending command to perform to infected machine, i.e. a Bot. The backbone of botnet is command and control(C&C) channel that is responsible for setting up the botnet, controlling the activities of the bots, issuing commands, and ultimately reaching the goals. The life of Botnet C&C is until it gets detected. Once a C&C channel is detected, the whole botnet is exposed and better to change the C&C for the BotMaster [10]. The main difference between Botnet and other kind of malwares is the existence of Command-and-Control (C&C) infrastructure. The Botmaster 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 infected machines to their master. The attacked bots, infected machine can also infect other computers enabling them to be botnet members. Basically botnet activities can be classified as three parts:

1) Searching- Searching for vulnerable and unprotected computers.

2) Dissemination (Infection) - the Bot code is distributed to the computers (targets), so the targets become Bots.

3) sign-on - the Bots connect to BotMaster and become ready to receive command and control traffic.

2.2 Botnet Life Cycle

A typical botnet can be created and maintained in five phases including: initial infection, secondary injection, connection, malicious command and control, update and maintenance. This life-cycle is depicted in Fig 2.

1) Initial Infection: The attacker initially scans a target subnet for known vulnerability, and infects victim machines through different exploitation methods. The methods like, binary download from web pages, email attachment, USB autorun and some malicious programs [11].

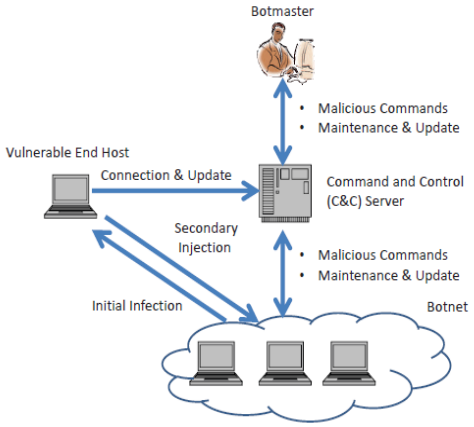

Fig.2 Botnet Life Cycle [11]

2) Secondary Injection: The infected hosts execute s script known as shell-code. It downloads this binary from specific location via FTP, HTTP or P2P. The code gets installed and targeted machine become "Bot" also known as "Zombie".

3) Connection: In connection phase, the bot program establishes a command and control (C&C) channel, and connects the zombie to the command and control (C&C) server. Upon the establishment of C&C channel, the zombie becomes a part of attacker's botnet army.

4) Command and Control: Now actual botnet activity is started. The Botmaster uses the C&C channel to disseminate commands to his bot army. The command communication can be IRC-based, HTTP-based, DNS-based or using P2P protocol to avoid single point of failure. Bot programs receive and execute commands sent by BotMaster. The C&C channel enables the botmaster to remotely control the action of large number of bots to conduct various illicit activities. 5) Update and Maintenance: bots are commanded to be lively and updated. So any new Solution to find and control is detected than control center (Botmaster) can update it with new strategies or may add new functionalities. Sometimes the updated binary move the bots to a different C&C server. This process is called server migration and it is very useful for botmasters to keep their botnet alive.

2.3 Botnet Topologies

According to the Command-and-Control(C&C) channel, we categorized Botnet topologies into two different models, the Centralized model and the Decentralized model.

1) Centralized Model: The oldest and easiest type of methodology is centralized model. The Central point C&C is responsible for exchanging commands and data between Botmaster and Bots. In centralized model BotMaster chooses a host-usually high bandwidth computer- to be the central point (Command-and-Control) server of all the Bots. Generally this C&C server runs certain network services such as IRC or HTTP. The main **advantage** of this model is small message latency which cause BotMaster easily arranges Botnet and launch attacks. As all connection happens through C&C it is a critical point in this model. It is also weak point as somebody manages to discover and eliminate C&C server, the entire botnet will be worthless and ineffective. IRC and HTTP are two common protocols that C&C server used for communication.

a) Botnet based on IRC: The IRC is a form of real-time Internet text messaging or synchronous conferencing. The protocol is based on the Client-Server model, which can be used on many computers in distributed networks. Advantages that encourages botmaster to use IRC as its C&C like, Low latency communication, Anonymous real-time communi cation, Ability of Group (many-to-many) and Private (one-to-one) communication, simple to setup, simple commands, flexibility in communication. Some famous IRC based botnet are, Agobot, SDBot, Spybot, and GT Bot.

b) Botnet based on HTTP: HTTP is another popular protocol used by C&C server. As IRC Botnet detection technique is famous for researcher, attacker started using alternative that is HTTP as C&C. The main advantage of using the HTTP protocol is hiding Botnets traffics in normal web traffics, so it can easily bypasses firewalls with port-based filtering mechanisms and avoid IDS detection. Some known Bots using the HTTP protocol are Bobax, ClickBot, Rustock and the most famous one Blackenergy aswll. Guet al pointed out that the HTTP protocol is in a "pull" style and the IRC is in a "push" style.

2) Decentralized Model: The main disadvantage of centralized C&C is, if researcher mitigate C&C then whole botnet will be down. So attackers started to develop an alternative which would be much harder to discover and destroy. As result, the P2P communication as C&C pattern which is more resilient against detection. In P2P model as shown in fig. 5 there is not any centralized point for control. Each Bot keeps some connections to the other Bots of the Botnet. Bots act as both Clients and servers. A new Bot must know some addresses of the Botnet to connect there. If Bots in the Botnet are taken offline, the Botnet can still continue to operate under the control of BotMaster.

Factors	Centralized (IRC,HTTP)	Hybrid DDNS	Peer-to-Peer P2P
Detection	Easy	Medium	Hard
Resilience	Low	Fairly High	Very High
Latency	Low	Medium	Fairly Hard
Traceback	Fairly Hard	Hard	Very Hard
Complexity	Easy	High	Medium
Experience	Very High	None	Medium

Table 1 Summary of C&C architecture

2.3 Botnet Detection Techniques

1) Honeypots and Honeynet : A honeypot can be defined as an "environment where vulnerabilities have been deliberately introduced to observe attacks and intrusions"(Pouget & Dacier, 2004). It is a computer system that is used as trap to draw the attention to attack this computer system. All Honeypots have a unique concept. They are computer systems that don't have any production value [9]. All Honeypots have a unique concept. They have a strong ability to detect security threats, to collect malware signatures and to understand the motivation and technique behind the threat used by perpetrator. For example, you may put a honeypot web server in the DMZ in your network. The limitations are Limited scale of exploited activities that can track, cannot capture the bots that is using method of propagation other than scanning. So we can conclude that while using Honeynet for Botnet detection, we have to wait until one bot in the network infect our system then we can track or analyze the machine.

2) Intrusion Detection System (IDS): It can be characterized mainly in two ways [9].

a) Signature Based Botnet Detection: Rule based intrusion detection systems like Snort are running by using known malware signatures. They monitor the network traffic and detect sign of intrusions. It is obvious that payload information of network traffic is transformed and embedded into the signature or rule. The IDS detects malicious traffic fitting the communication parameters defined by the rule. Gu et al. (2007) propose a framework, "BotHunter", to correlate IDS based detection alerts.

b) Anomaly Based Botnet Detection: This approach tries to detect Botnet based on number of network traffic anomalies such as high network latency, high volumes of traffic, traffic on unusual ports, and unusual system behavior that could show existence of bots in the network. This approach can detect unknown Botnets. It can be categorized as Host based and Network Based Detection.

In host based detection technique, a detection strategy which monitors and analyzes the internals of a computer system instead of network traffics on its external interfaces. Limitation with this system is high false positive.

A network-based technique is a detection strategy which tries to detect Botnets by monitoring network traffics. We can classify Network-based techniques into two categories: Active monitoring and Passive monitoring. In Active monitoring, it injects test packets in network to measure the reaction of network such that gaining extra traffic on network. Gu et al. have proposed Botsniffer that uses network-based anomaly detection to identify Botnet C&C channels in a local area network.

3) DNS Based detection technique: In order to access the C&C server bots carry out DNS queries to locate the particular C&C server that is typically hosted by a DDNS (Dynamic DNS) provider. So DNS monitoring will be easy approach to detect Botnet DNS traffic and detect DNS traffic anomalies. This is most famous and easy technique to botnet detection but it will be tough to detect recent advanced botnet through this technique.

4) Data Mining Based Detection Technique: Data mining aims to recognize useful patterns to discover regularities and irregularities in large data sets. Packet flow provides full information of flow data but in large file type. Anomaly based techniques are mostly based on network behavior anomalies such as high network latency, activities on unused ports [9]. Data mining technique can be applied for optimization purpose. It enables to extract sufficient data for analysis from network log file. Most useful data mining techniques includes correlation, classification, clustering, statistical analysis, and aggregation for efficiently knowledge discovery about network flows [12].

Flow correlation algorithms are useful to compare flow objects based on some characteristic other than packet content. This technique is very effective when content of packet is not available of encrypted, e.g. might compare arrival time. These kinds of algorithms utilize the characteristic values as inputs into one or more functions to create a metric used to decide if the flows are correlated [12].

Classification algorithms assume that incoming packet will match one of the previous patterns. Therefore, it is not an appropriate approach to detect new attacks [12].

Clustering is a well-known data mining technique where data points are clustered together based on their feature values and a similarity metric. Clustering differs from classification, in that there is no target variable for clustering. Clustering algorithms divide the entire data set into subgroups or clusters containing relatively identical features. Thus, clustering provides some significant advantages over the classification techniques, since it does not require a labeled data set for training [12]. To find particular pattern from large dataset is known as aggregation method, collecting and analyzing several types of records from different channels simultaneously.

Association rule is to find the correlation of different items appeared in the same event. Association rule mining is to derive the implication relationships between data items under the conditions of a set of given project types and a number of records and through analyzing the records, the commonly used algorithm is Apriori algorithm. [13].

2.3 Conclusion

In recent era, there are so many research work has been done for P2P and IRC botnet. The motivations for using the HTTP protocol are multiple. Developing a web-based C&C application is typically easier than implementing customized C&C communication protocols (e.g., peer-to-peer protocols), and there is evidence that web-based "reusable" kits (or platforms) for botnet C&C are available for sale on the Internet. Many of today's botnets and other types of malware leverage HTTP-based network communications for command-and-control (C&C) purposes or to perpetrate malicious activities.

By using various attribute analysis like timeslot, data calculating, mutual authentication and bot clustering analysis they have obtain the various characteristics of the Botnet. Although the characteristics of web botnets in the real world are inadequate, they can be used to understand the regularly behaviors to enhance the detection capability in the future. Data mining and machine learning techniques are easily applicable on network flow information. Flow data have a structured and related nature, which do not require massive preprocessing tasks. Besides, flow information implies patterns inside, which makes data mining algorithms convenient and effective for analysis.

References

[1]Chung-Huang Yang, Kuang-Li Ting. Fast Deployment of Botnet Detection with Traffic Monitoring, Fifth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, pages 856-860, 2009.

[2]Haritha S. Nair, Vinodh Ewards S E A Study on Botnet Detection Techniques, International Journal of Scientific and Research Publications, Volume 2, Issue 4, April 2012

[3]Botnet scams are exploding, 2008 http://usatoday30.usatoday.com/tech/news/computersecurity/2008-03-16-computer-botnets_n.htm.

[4]Nicholas Ianelli, Aaron Hackworth. Botnets as a Vehicle for Online Crime - CERT and CERT Coordination Center are registeredin the U.S. Patent and Trademark Office. 2005

[5]Oregon Man Cops Plea in eBay DDOS Attack, http://www.internetnews.com/security/article.php/3574101

[6]Worm strikes down Windows 2000 systems,http://www.cnn.com/2005/TECH/internet/08/16/computer.worm/index.html

[7]Kraken botnet, Wikipedia, http://en.wikipedia.orglwikilKraken_botnet, 2008.

[8]Zeus botnet steals $47M from European bank customers,2012. http://news.cnet.com/8301-1009_3-57557434-83/zeus-botnet-steals-$47m-from-european-bank-customers/

[9]Erdem Alparslan, Adem Karahoca and Dilek Karahoca. BotNet Detection: Enhancing Analysis by Using Data Mining Techniques, Downloaded from http://dx.doi.org/10.5772/48804 (BOOK)

[10] Sonal P.Patil, Swatantra Kumar. Botnet-A Network Threat, International Conference on Recent Trends in Information Technology and Computer Science (IRCTITCS), Pages 29-35, 2011.

[11] Xiaonan Zang, Athichart Tangpong, George Kesidis and David J. Miller. Botnet Detection Through Fine Flow Classification. Departments of CS&E and EE, The Pennsylvania State University, University Park, PA, 16802. CSE Dept Technical Report No. CSE11-001, Jan. 31, 2011

[12]Alireza Shahrestani, Maryam Feily, Rodina Ahmad, Sureswaran Ramadass. ARCHITECTURE FOR APPLYING DATA MINING AND VISUALIZATION ON NETWORK FLOW FOR BOTNET TRAFFIC DETECTION, International Conference on Computer Technology and Development,IEEE, Pages 33-37 2009

[13]Zhang yanyan,Yao Yuan, Study of Database Intrusion Detection Based on Improved Association Rule Algorithm, IEEE. Pages 673-676, 2010

[14]Sajjad Arshad, Maghsoud Abbaspour, Mehdi Kharrazi, Hooman Sanatkar. An Anomaly-based Botnet Detection Approach for Identifying Stealthy Botnets, Presented in International Conference on Computer Application Industrial Electronics, IEEE, Pages 564-569, 2011

[15]Wang Zilong, Wang Jinsong, Huang Wenyi, Xia Chengyi. The Detection of IRC Botnet Based on Abnormal Behavior. Second International Conference on MultiMedia and Information Technology, IEEE, Pages 146-149, 2010

[16]J. Goebel and T. Holz. Rishi: Identify bot contaminated hosts by irc nickname evaluation In USENIX Workshop on Hot Topics in Understanding Botnets (HotBots 07), 2007

[17]Claudio Mazzariello. IRC traffic analysis for botnet detection, The Fourth International Conference on Information Assurance and Security, IEEE, Pages 318-323, 2008

[18]Hossein Rouhani Zeidanloo, Mohammad Jorjor Zadeh shooshtari, Payam Vahdani Amoli, M. Safari, Mazdak Zamani. A Taxonomy of Botnet Detection Techniques, IEEE, Pages 158-162, 2010

[19]Roberto Perdiscia, Wenke Leea, and Nick Feamstera, Behavioral Clustering of HTTP-Based Malware and Signature Generation Using Malicious Network Traces, College of Computing, Georgia Institute of Technology, Atlanta, GA 30332, USA, USENIX, 2010

National Conference on Recent Research in Engineering and Technology (NCRRET-2015) International Journal of Advance Engineering and Research Development (IJAERD)
e-ISSN: 2348 - 4470 , print-ISSN:2348-6406