ISSN: 2321-7782 (Online)
Volume 2, Issue 1, January 2014
International Journal of Advance Research in Computer Science and Management Studies
Research Paper
Available online at: www.ijarcsms.com

# A Survey on Anomaly Detection of Botnet in Network

**S. Nagendra Prabhu**[1]
Computer Science and Engineering
RVS Educational Trust Group of Institution
Dindigul - India

**D. Shanthi**[2]
Computer Science and Engineering
P.S.N.A College of Engineering & Technology
Dindigul - India

***Abstract:*** ***Botnets are a major threat of the current Internet. Understanding the recent procreation of botnets relying on peer-to- peer networks is critical for diminishing this threat. Today botnets are seen to be one of the main sources of malicious activity. Expeditiously growing botnets find modern methods for spreading malicious codes and dispatch attacks. Nowadays, botnet traffic is merged with a huge volume of favorable traffic due to almost everywhere high speed networks. This article is a survey of recent proposal in botnet research and its mitigation. The survey distinguishes the botnet research into three fields: Anomaly Detection - Botnets, botnet attacks and latest botnet behaviors, and techniques for defending against botnets. While botnets are boundless, the investigation and clarification for botnets are still in their childhood. The paper also summarizes the existing research and recommends future directions for Botnet research.***

***Keywords: Command-and-Control (C&C), denial-of-service (DDOS), IRC (Internet Relay Chat), Peer-to-Peer (P2P).***

## I. INTRODUCTION

Cloud computing refers to the delivery of computing and storage capacity as a service to a heterogeneous community of end-users. It is a network-based environment that target on sharing computations or resources. It refers to both the applications delivered as services over the Internet and the hardware and software in the datacenters that provide those services. Cloud providers use virtualization technologies combined with self service abilities for computing resources via network infrastructure. In cloud, costumers must only pay for what they use and have not to pay for local resources which they need to such as storage or infrastructure. Cloud computing itself is in principle an abstraction of the physical infrastructure which is offered as cloud services to service users. The abstraction levels of these cloud services are Software-as-a-Service (SaaS), Platform-as-a-Service (PaaS), and Infrastructure-as a-Service (IaaS). Popular examples are GoogleDocs for SaaS, Google's AppEngine for PaaS, and Amazon's EC2 for IaaS.

In cloud computing, we have three types of cloud environments deployment model: Public, Private and Hybrid clouds. A public cloud is standard model which providers make several resources, such as applications and storage, available to the public. Public cloud services may be free or not. In public clouds which they are running applications externally by large service providers and offers some benefits over private clouds. Private Cloud refers to internal services of a business that is not available for ordinary people. Essentially Private clouds are a marketing term for an architecture that provides hosted services to particular group of people behind a firewall. Hybrid cloud is an environment that a company provides and controls some resources internally and has some others for public use. Also there is combination of private and public clouds that called Hybrid cloud. In this type, cloud provider has a service that has private cloud part which only accessible by certified staff and protected by firewalls from outside accessing and a public cloud environment which external users can access to it.

In recent times cloud computing has become more and more popular and is applied for various purposes. It offers great potential to improve productivity and reduce costs, but at the same time it possesses many security risks.

 

## II. ANOMALY DETECTION – BOTNET

A key challenge in anomaly detection [3] is defining what is normal and identifying the boundary between normal behavior and the outlier or abnormal behavior. One of the most considerable anomaly threats is Botnet. A botnet is a collection of internet-connected computers whose security guard have been breached and control surrender to a third party. Each such compromised device, known as a "bot", is created when a computer is penetrated by software from a malware code distribution; otherwise known as malicious software. Nowadays, the most serious manifestation of advanced malware is Botnet. To make distinction between Botnet and other kinds of malware, we have to comprehend the concept of Botnet. For a better understanding of Botnet, we have to know two terms first, Bot and BotMaster and then we can properly define Botnet.

*Bot & BotMaster*

Bot – Bot is actually short for robot which is also called as Zombie. It is a new type of malware installed into a compromised computer which can be controlled remotely by BotMaster for executing some orders through the received commands. After the Bot code has been installed into the compromised computers, the computer becomes a Bot or Zombie. Contrary to existing malware such as virus and worm which their main activities focus on attacking the infecting host, bots can receive commands from BotMaster and are used in distributed attack platform. Botnets with a large number of computers have enormous cumulative bandwidth and powerful computing capability. They are exploited by botmasters for initiating various malicious activities, such as email spam, distributed denial-of-service (DDOS) attacks, password cracking and key logging.

*Botnet C&C Channel*

The major difference between Botnet and other malicious threats is the presence of Command-and-Control (C&C) source. The main difference between Botnet and other kind of malwares is the existence of Command-and-Control (C&C) infrastructure. The C&C enable Bots to receive commands and malicious activities, as committed by BotMaster. BotMaster must confirm that their C&C infrastructure is sufficiently powerful to manage thousands of distributed Bots across the network, also it should be against to any attempts to close down the Botnets. The other side detection and mitigation methods against Botnets have been increased [3], [7], [9]. Currently, attackers are also steadily improving their way to preserve their Botnets. The first generation of Botnets make use of IRC (Internet Relay Chat) channels as their Common-and-Control (C&C) centers. The centralized C&C mechanism of such Botnet has made them vulnerable to being detected and disabled. Therefore, new generation of Botnet which can hide their C&C communication have emerged, Peer-to-Peer (P2P) based Botnets. The P2P Botnets do not suffer from a single point of failure, because they do not have centralized C&C servers. Attackers have accordingly developed a range of strategies and techniques to protect their C&C infrastructure. Therefore, considering the C&C function gives better understanding of Botnet and help defenders to design proper detection or mitigation techniques.

*Categorization of Botnet*

According to the C&C channel the botnet [12] are categorize into three different topologies: a) Centralized; b) Decentralized and c) Hybrid.

*a) Centralized Model*

It is one of the oldest model, here one center point is accountable for exchanging commands and malicious data between the BotMaster and Bots. In this model, BotMaster chooses a host (usually high bandwidth computer) to be the central point (Command-and-Control) server of all the Bots as shown in fig: 1.

 

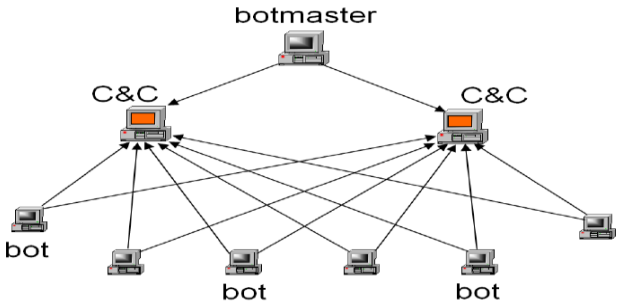


Fig: 1. Centralized Botnet

Since all connections are under working through the C&C server, therefore, the C&C is a critical point in this model. In fact, C&C server is the weak point in this model. If anyone manages to discover and eliminates the C&C server, the entire Botnet will be worthless and ineffective. Thus, it becomes the main drawback of this model.

*b) Decentralized Mode*

To overcome the problem occurred in centralized model, attacker focused to create alternative. Hence a new model is created my attacker in which the communication does not weightily depends on few selected server even though detecting and destroying a large number of Bots. As a result, attackers exploit the idea of Peer-to-Peer (P2P) communication as a Command-and-Control (C&C) pattern which is more flexible to failure in the network. The P2P based C&C model will be used dangerously in Botnets in the near future, and doubtlessly Botnets that use P2P based C&C model establish much bigger challenge for protecting the networks. Since P2P based communication is more robust than Centralized C&C communication, this is the reason why Botnets will move to use P2P protocol for their communication. In the P2P model, as shown in Fig. 2, there is no Centralized point for communication. 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.

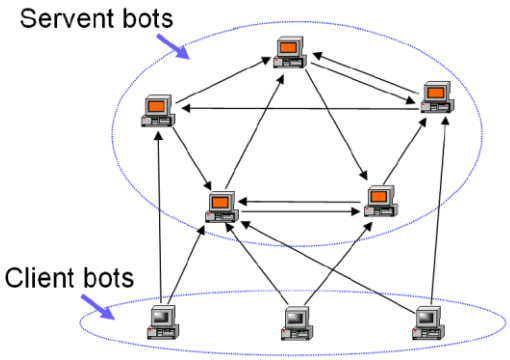


Fig: 2. Decentralized Botnet

*c) Hybrid Model*

Hybrid P2P botnet is the most advanced and critical communication for protecting the network. The hybrid p2p [15] Botnet consists of three parts – botmaster, social websites, bot groups. Bot groups consist of servent bots and client bots. Malicious code from botmaster is embedded into social websites. The servent bots in bot groups will link with the social website to get the malware information from the social websites and send to client bots. Client bots attack target after they receive malware information from servent bot. The brief diagram for the process of embedding social website, the process of servent bots, and the process of client bots are explained separately as follows.

 

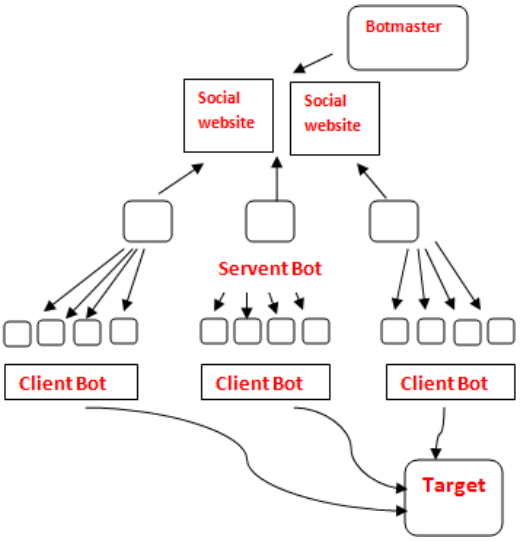


Fig 3: Hybrid Model

## III. LITERATURE REVIEW

Denial of service attack stand as one of the most networks outstanding attacks, by which one victim machine can receive more than its capacity and so other end users requests cannot be served by server, and in the cloud environment, that kind of attack can be most harmful than unclouded environment because of VMs neighboring and resource sharing in cloud computing environment, so one virtual machine can be used as a source of denial of service attack to another virtual machine in same infrastructure and for overcome this security threat , several kind of flooding DoS attacks detecting and prevention approach has been suggested, and every one propose a method to detect or prevent this attack.

*Botnet Exposure Techniques*

Various Botnet detection techniques are listed down. Let see one by one:

*A. Counter-Based Method*

Counter-based detection [6] is a simple method that counts the total traffic volume or the number of web page requests. Since the DDoS attack with the low volume of traffic such as the HTTP GET incomplete attack is prevalent in these days, the frequency of page requests from clients will be a more effective factor. In MapReduce algorithm to detect DDoS with URL counting. To lower the false positive rate, we adopted response rate against page requests as secondary regulation as well as traffic volume, which was proposed by Liu et al [3]. This algorithm needs three input parameters of time interval, threshold and unbalance ratio, which can be loaded through the configuration property or the distributed cache mechanism of MapReduce. Time interval limits monitoring duration of the page request. Threshold indicates the permitted frequency of the page request to the server against the previous normal status, which determines whether the server should be alarmed or not. The unbalance ratio variable denotes the anomaly ratio of response per page request between a specific client and a server. This value is used for picking out attackers from the clients. By using this algorithm it needs a prerequisite to know the threshold value from historical monitoring data in advance.

*B. Access Pattern-Based Method*

The access pattern-based detection method [6] assumes that clients infected by the same bot conduct the same behavior and that attackers could be differentiated from normal clients. This method requires more than two MapReduce jobs: the first job obtains access sequence to the web page between a client and a web server and calculates the spending time and the bytes count for each request of the URL; the second job hunts out infected hosts by comparing the access sequence and the spending time among clients trying to access the same server. The drawback of this approach is the highly computational complexity to spot the DDoS pattern.

 

*C. Entropy Based Anomaly Detection System*

In [4] the author proposed a new architecture for Cloud Computing platform, where the whole Cloud System is divided into multiple administrative domain, which is controlled separately by its own Authentication & Certification Authority i.e. AS. The author also developed ADS for detection & early prevention of DDoS attacks. In this architecture a tree is maintained at every router, by marking every packet with path modification strategy, so that the victim is able to trace the sender of the packet. Any packet which was detected as malicious flow, can be confirmed in a second try i.e. victim node. IDS may be in software form and/or in hardware form, that will monitor the network for disbelieving activity and alerts the network administrator to take a particular action accordingly. Proposed ADS will be implemented in Cloud environment, and also how routers communicate with each other to detect DDoS attack. In this section it also describe that how to mathematically proven the proposed method and examined the simulation results along with charts form with a practical environment.But using this Entropy based Anomaly Detection system it will proven a good detection method but create lot of false positive alarms.

*D. Min-Vertex Cover Method*

In [11] the author has developed a novel P2P botnet detection model that combines session-based analysis and minimum vertex cover theory, this model only analyzes network header information, regardless of extra network data load, and takes use of minimum vertex cover to probe the core nodes of botnet. In addition, it maintains a high success rate even if the attacker changes the features of botnets to counterattack. The system can excavate the core nodes of botnet while there is no guarantee on detection rate and false-positive rate of bots. Furthermore it can use the core nodes to induce the features of botnet, which are feedback to system to increase detection efficiency at next time. The algorithm for solving the core nodes in the simulation results shows a finer performance in face of poor session detection rate and false-positive rate. The analysis results will be stored into the session table in the database. And then core node calculation module, via session table, structures a simple undirected graph which represents the botnet. The author proposed the algorithm for solving the core nodes in the simulation results shows a finer performance in face of poor session detection rate and false-positive rate.

*E. File Allocation Table Method*

Usually when a customer opens an account in the cloud, the provider creates an image of the customer's VM in the image repository system of the cloud. The applications that the customer will run are considered with high efficiency and integrity. In [13] the author proposed to consider the integrity in the hardware level, because it is very difficult for an attacker to intrude in the IaaS level. In this paper the author utilize the File Allocation Table (FAT) system architecture, since its straightforward technique is supported by virtually all existing operating systems. From the FAT table it can know about the code or application that a customer is going to run. We can check with the previous instances that had been already executed from the customer's machine to determine the validity and integrity of the new instance. For this purpose, we need to deploy a Hypervisor in the provider's end. This Hypervisor will be considered the most secured and sophisticated part of the cloud system whose security cannot be breached by any means. The Hypervisor is responsible for scheduling all the instances, but before scheduling it will check the integrity of the instance from the FAT table of the customer's VM. Another approach is to store the OS type of the customer in the first phase when a customer opens an account. As the cloud is totally OS platform independent, before launching an instance in the cloud, cross checking can be done with the OS type from which the instance was requested from with the account holder's OS type. But using FAT it should know about the all the instance from the database.

*F. Page Rank Algorithm*

In [5] the author proposed about a distributed computing framework that leverages a host dependency model and an adapted PageRank algorithm. It is a link analysis algorithm used by the Google web search engine to weight the relative importance of web pages on the Internet. It ranks each web page according to the hyperlink structure among web pages. PageRank is well fitted for MapReduce and this section reviews the basics of executing PageRank in the MapReduce context without taking into account the dangling nodes or the damping factor for clarity sake. The author reported the experimental results from an open-source based Hadoop cluster and highlight the performance benefits when using real network traces from an Internet operator.

 

*G. Honeypots Method*

In general honeypots plays an important role in detecting and analyzing botnets [9],[7]. They can be used to analyze malware code to generate anti-virus signatures. From [9] the author explained, it is clear that in order to obtain maximum benefit from a honeypot, it need to classify network traffic and internal activities clearly. This is not a trivial task, and requires a detailed and close monitoring of all the processes which requires a huge amount of time. The main drawback with honeypot based method is the fact that it can only see the incoming traffic from the IP address assigned to it, which may not be enough for a comprehensive analysis and detection of a botnet. Honeynet work used only unpatched versions of all versions of Windows as Honeypot, and Snort inline used as Honeywall device to track Botnets on a daily basis report**.**

## IV. FORWARD LOOKING RESEARCH

Botnets stand a significant and growing threat against cyber-security as they provide a key platform for many cybercrimes such as DDoS attacks against critical targets, malware dissemination, phishing, and click fraud etc. So now in current days, instead of using a centralized, IRC based C&C channel to perform multiple criminal attacks, the Botnets have been gradually developed into more complicated, and stealthy activity with diverse C&C protocols and structures such as P2P botnet, Hybrid P2P Botnet. Hence existing the long presence of malicious Botnets, only a small amount of studies have examined on Botnet problem and Botnet research is still in its infancy. This Literature survey described about Botnet and various Botnet detection techniques to detect these malicious activity. Finally it is necessary to discuss about further Botnet developments which may arise in future. Hence the following points summarize the future trends to be carried out in Botnet research.

A new idea for detecting Botnet, An approach that can identify anomalies in the information stream does not require any predefined rules or domain specific knowledge base. The concept of my approach simultaneously monitors and analyzes the data stream at multiple temporal scales and learns the evolution of normal behavior over time in each time scale.I am planning to experiment the results from an open-source based Hadoop Cluster using MapReduce Concept [8], [17], MapReduce is a programming model for processing large data sets, and the name of an implementation of the model by Google. MapReduce is typically used to do distributed computing on clusters of computers. The mapreduce concept is mainly used for Simple data-parallel programming model designed for scalability and fault-tolerance.

## V. CONCLUSION

In this paper, the main objective is to design an unknown botnet detection system using cloud computing. Not only to enhance the existing Botnet Exposure techniques, but to raise performance due to cloud computing. Implementing the idea of Information stream approachs in MapReduce concept from an open-source based Hadoop Cluster; there is a lot of chance to achieve high detection rate and low false positive rate.

## References

1. http://en.wikipedia.org/wiki/Botnet.
2. http://en.wikipedia.org/wiki/Cloud_computing,http://en.wikipedia.org/wiki/Zombie_(computer_science)
3. Ding, J. R. Binkley , S. Singh, “An algorithm for anomaly-based botnetdetection,” Proc. of 2nd Workshop on Steps to Reducing Unwanted Traffic on the Internet (SRUTI'06), July 2006, pp 43-48.
4. Muhammad Zakarya & Ayaz Ali Khan "Cloud QoS, High Availability & Service Security Issues with Solutions" IJCSNS International Journal of Computer Science and Network Security, VOL.12 No.7, July 2012
5. J´erˆome Franc¸ois, Shaonan Wang, Walter Bronzi, Radu State, Thomas Engel. “BotCloud: Detecting Botnets Using MapReduce”, University of Luxembourg
6. Yeonhee Lee, Youngseok Lee "Detecting DDoS Attacks with Hadoop"
7. Zhaosheng Zhu, Guohan Lu, Yan Chen "Botnet Research Survey" Annual IEEE International Computer Software and Applications Conference
8. http://cloudcomputing.blognhanh.com/2012/07/mapreduce-based-ddos-detection.html
9. Amit Kumar Tyagi,G.Aghila "A Wide Scale Survey on Botnet" International Journal of Computer Applications (0975 – 8887) Volume 34– No.9, November 2011
10. http://www.infosecuritymagazine.com/view/24987/comment-botnets-the-dark-side-of-cloud-computing/

 

11. Lei Xu,XiaoLong Xu, Yue Zhuo "P2P Botnet Detection Using Min-Vertex cover" journal of networks, vol. 7, no. 8, august 2012
12. Hossein Rouhani Zeidanloo & Azizah Abdul Manaf, "Botnet Command and Control Mechanisms", 2009 Second International Conference on Computer and Electrical Engineering
13. B.Meena, Krishnaveer Abhishek Challa "Cloud Computing Security Issues with Possible Solutions" IJCST Vol. 3, Issssue 1, Jan. - March 2012
14. Bansidhar Joshi, A. Santhana Vijayan, Bineet Kumar Joshi "Securing Cloud Computing Environment Against DDoS Attacks"2012 International Conference on Computer Communication and Informatics(ICCCI-2012), Jan,10-12,2012,coimbatore,india
15. T. T. Lu, H.Y. Liao, M .F. Chen, "An Advanced Hybrid P2p Botnet 2.0", World Academy of Science, Engineering and Technology 57 2011
16. Esraa Alomari,B.B.Gupta,Shankar Karuppayah, "Botnet-based Distributed Denial of Service (DDoS) Attacks on Web Servers: Classification and Art", International Journal of Computer Applications (0975 – 8887) Volume 49– No.7, July 2012
17. Luis A. Trejo, Roberto Alonso, Adri¶an ¶ Avila,"Using Cloud Computing MapReduce operations to Detect DDoS Attacks on DNS servers"
18. S.Renuka Devi & P.Yogesh, "DETECTION OF APPLICATION LAYER DDOS ATTACKS USING INFORMATION THEORY BASED METRICS", CS & IT-CSCP 2012, 10.5121/csit.2012.2223
19. Weili Huang & Jian Yang,"New Network Security Based On Cloud Computing",2010 Second International Workshop on Education Technology and Computer Science.
20. Lanjuan Yang,Tao Zhang, Jinyu Song, JinShuangWang, Ping Chen, "Defense of DDoS Attack for Cloud Computing" PLA University of Science,China.
21. Chi-Chun Lo,Chun-Chieh Huang & Joy Ku "A Cooperative Intrusion Detection System Framework for Cloud Computing Networks" 2010 39th International Conference on Parallel Processing Workshops.
22. Cong Wang & Kui Ren, "Toward Secure and Dependable Storage Services in Cloud Computing", IEEE TRANSACTIONS ON SERVICES COMPUTING, VOL. 5, NO. 2, APRIL-JUNE 2012.
23. Chu Huang, Sencun Zhu and Dinghao Wu "Towards Trusted Services: Result Verification Schemes for MapReduce"2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing.

 