Review of the Research on Botnet

Gao Jian
Information technology and
network security College,
People's Public Security
University of China,
BeiJing,China
People's Public Security of
China,Daxing District,
Beijing,China,102600

ABSTRACT
The botnet is controlled by an attacker, which is formed by a lot of vulnerable hosts. The botnet is one of the biggest threats on the Internet. The attacker usually uses it to attack, such as: spam, distributed denial of service attacks, fraud and so on. In this paper, we mainly study the control channel of the botnet, including the IRC protocol, the P2P protocol and the HTTP protocol. At the same time, this paper also studies the detection method of the botnet, which includes the host based detection method and the network based detection method.

General Terms
Botnet, malware.

Keywords
Botnet,Command and Control, P2P,Detection,DDoS.

1. INTRODUCTION
CNCERT/CC definition of the botnet is: the attacker to use the Internet to establish a secret, you can focus on the control of the computer group. Institute of science and technology of information security engineering research center of Peking University computer botnet is defined as: the attacker for malicious purposes, spread bots control a large number of hosts, and is composed of many of the command and control channel network. Botnet (botnet) is a new type of controllable attack platform which is formed by the fusion of the traditional malicious code technology, such as network worm, Troy Trojan, virus, back door tool.

The first malicious botnet PrettyPar appeared on the Internet in 1999. 2002 SDbot and Agobot botnet source code published on the Internet and widely circulated, many hackers on this basis to modify the new variants, the botnet has quickly become a serious security threat to the internet. In 2003, Phatbot began to use P2P construction technology to form its control channel, Phatbot based on the Agobot, and integrated into the P2P technology. From now on the P2P botnet continues to surge on the Internet. P2P botnet is distributed, so it solves the problem of single point failure of traditional botnet. In 2006 BlackEnergy Botnet, it mainly uses the HTTP protocol, its advantages can be generated by the entire network information on the Internet in the mixed traffic is huge, so the check and counter difficult. 2007 P2P protocol based on the botnet Storm began spreading in Europe, as of June 30, 2007, has been infected with 170 million hosts. In 2008, Microsoft's windows operating system for the target of the botnet Conficker appears on the internet. Symantec reported from July 2008 to June 2009, hackers steal credit cards, e-mail lists, bank accounts and other personal information to sell the price, it is estimated that the total value of more than $276 million. 2010, still use the P2P technology Kenzero botnet appears, its important target is to download illegal files and browse pornographic websites users. This year (2011), a new, HTTP based botnet Zeus is being spread, its main use of social engineering to spread, in order to obtain a bank card password, personal information for the purpose of.

2. THE COMMAND AND CONTROL OF BOTNET
Puri[1] et al, Research on the more active IRC Botnet, a comprehensive and systematic overview of the composition of the various components of the network are summarized and explained. In 2006, Barford[2] et al. System analysis of several IRC botnet is well-known: GTBot, SDBot, AgoBot and SpyBot, put forward the classification method of bots function structure, and analyzed from seven aspects of functional characteristics of each botnet. Nazario[3] Jose et al. On the HTTP protocol based on the BlackEnergy protocol of the communication mechanism of the Botnet, zombie program configuration, DDoS attack behavior and the results of anti virus checks were analyzed. Stover[4] et al specific protocols of P2P and Storm and Nugache botnet command encryption method is analyzed, and the comparison of two different Trojans, studied how to use the P2P mechanism. Wang[5] et al. Proposed the existing P2P technologies are applied directly to the botnet is not appropriate, and proposed a new hybrid P2P Botnet, including its command and control mechanism, and the robustness and defense methods are analyzed. In 2009, Bailey[6] Michael, who had a review of the existing Botnet, its evolution process and its future shape. Fitzgibbon Conficker et al. Analysis and Research on the communication mechanism of Niall worm [7] and its P2P. Vogt[8] et al. Proposed a hierarchical super-botnet botnet construction methods, this structure can improve the robustness of the botnet to a great extent, and the feasibility and characteristics of the communication mechanism of super-botnet is verified by simulation. Based on the P2P protocol control mode of the Storm Botnet, Grizzard[9] et al introduced the characteristics of the P2P botnet and the Trojan.Peacomm based Kademlia samples were systematically analyzed. Porras[10] et al from multiple perspectives on the Storm zombie program is analyzed, mainly for the different samples of the Storm zombie program, the static analysis and network communication analysis.

A. Command and control based on IRC protocol

In the early IRC[11] chat network, administrators in order to prevent the channel from being abused, better management authority,

Recording channel events and a series of functions, the preparation of a smart program to complete this series of services. In 1993, for the purpose of service, chat network administrator to write the first good zombie program Eggdrop, used to manage the IRC chat network. Inspired by this design ideas, hackers then write a malicious zombie program, implanted many victims of the host, the use of IRC chat channel issued a command to control them, launched a variety of malicious attacks. In June 1999 the first to use the IRC server for remote control of the zombie PrettyPark, founder of the zombie zombie called network technology, it contains some still widely used even now functions, such as access to the host information, search the user name and password and other sensitive information, self renew, upload and download files, redirect the communication, launch denial of service attack etc.. Then a large number of BOT procedures based on the IRC protocol appear, such as GTbot, SDBot, SpyBot and Agobot, etc.

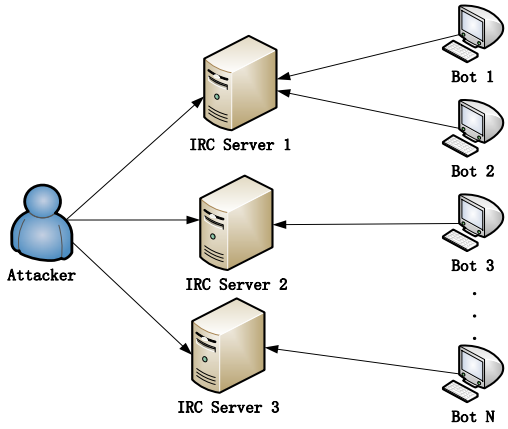

Fig. 1. IRC Botnet Topology Graph

B. Command and control based on HTTP protocol

In addition to the IRC protocol, HTTP protocol in recent years is another popular botnet command and control protocol, HTTP protocol botnet using HTML language based communication, first controlled machines will access a URL request information sent to the control by the Web server, if the connection is successful, the Web server will feedback this request, and contains the control commands issued to the current attackers botnet in return content. The zombie program then parse the command from the returned content and execute it.

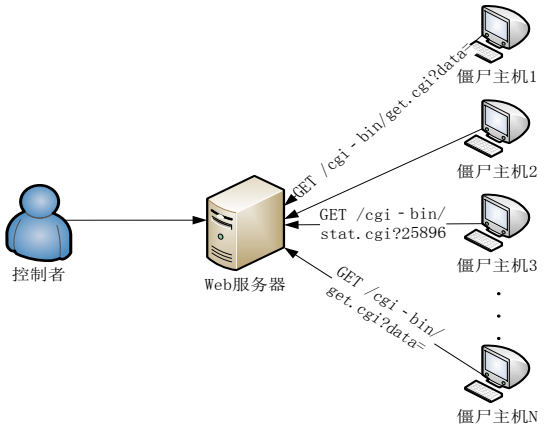

Fig. 2. HTTP Botnet Topology Graph

C. Command and control based on P2P protocol

The command and control mechanism based on IRC protocol and HTTP protocol has centralized control point, which makes it easy to track, detect and counter the botnet based on client server architecture. Once the defenders get bots, they can easily find the botnet controller position, global information and the use of monitoring and tracking means to control the Botnet, by closing these centralized botnet controller can easily attack botnet. In order to make the botnet more resilient and hidden, some of the new zombie process began to use the P2P protocol to build its command and control mechanism.

Grizzard[9] and others on the development process of P2P botnet were reviewed, and the different P2P control mechanisms of the botnet Slapper, Sinit[12], Phatbot[13], SpamThru, Nugache and Peacomm were introduced. Wang[5] et al. Proposed a hybrid P2P Botnet, and the command and control mechanisms of the framework for the detailed design, the whole structure, the zombie all nodes into super nodes and ordinary nodes, similar to the Gnutella agreement. Vogt[8] et al proposed a cascade of "Super-Botnets" botnet group construction, the continuous decomposition in the propagation process of botnets, to ensure that the botnet size limits, and the relationship between the neighbor nodes through small botnets between and based on public key encryption communication mechanism to construct the botnet group. P2P botnet now can be divided into two categories: the first category is [14] P2P in parasitic botnets, parasitic P2P Botnet, all nodes are vulnerable to some zombie node on the existing P2P network, the nodes use their P2P network protocol to communicate; second is the P2P Botnet, deposit P2P Botnet, the use of communications itself to create the agreement does not depend on any P2P network, so it has better invisibility, but also because of its own protocol is a new protocol, without the Internet test, so its stability needs to be improved.

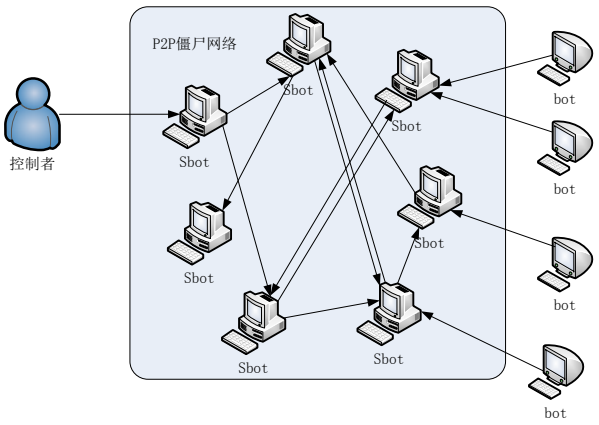

Fig. 3. P2P Botnet Topology Graph

3. BOTNET DETECTION TECHNOLOGY

The academic circles in 2003 began to pay attention to the development of the Botnet, some international researchers engaged in honeypot, Honeynet of Botnet activities in-depth tracking and analysis. There are Pacific Azusa University's McCarty[1] Bill, France's Clarke[4] Dave, University of Washington's Dittrich[6] Richard and Germany's honey network group. German honey network group analysis of the situation of the Botnet, which includes the use of the Botnet, categories and behavior, and other aspects. Wang[15] et al proposed a method to detect honeypot Honeynet environment in botnet. Trolle Borup [16] et al. Lasse from the analysis

method, sample acquisition, defense means all aspects of a new type of Botnet Waledac for a comprehensive analysis. Stock[17] Waledac et al. Ben has also been studied. Zhu[18] Zhaosheng and others from three different parties in the face of the existing botnet overview: understanding the Botnet, detection of Botnet and defense. Guenther Starnberger[19] et al introduced P2P Overbot botnet based on Kademlia protocol in Overbot, even if the attacker caught some zombie nodes can gain other node's IP address, or interrupt communication node and control terminal, and analyze some possible means of defense. Dittrich[20] David and others deeply analyzed the Nugache Botnet, through the analysis of its network traffic, the study of its communication mechanism, DDoS attacks. Florio[13] et al introduced Storm botnet is how to use the current Overnet network communication control, from many aspects of BOT host single behavior, network behavior and malicious behavior analysis of Storm Botnet, and how to track and destroy botnet gives some methods.

Botnet is a new type of attack which is evolved from the traditional malicious code, and the detection of Botnet still belongs to the category of malicious code identification. The research of Botnet detection methods can be divided into two aspects: the host recognition method based on the direct use of existing malicious code detection method to identify botnet; network recognition method based on the use of network communication, the characteristics of Botnet, inductive learning, in order to identify the zombie host or server.

Host based signature recognition method has a lot of [21][22][23], such as anti-virus tools, anti-virus software, these tools for malicious code checks are very effective. But the creators of the botnet are also very familiar with this, they can be faster than antivirus software vendors to update their signature code to avoid this type of inspection. The inspection method based on the host computer is largely similar to that of the Trojan horse. Trojan and the main difference between the botnet is the ability to do not have the ability to spread Trojans, and the structure of the botnet is much more complex than trojan.

A. Host based Detection

1) port based detection [24]

In the past some old-fashioned Trojans use fixed ports, according to this characteristic can easily determine whether infected with the Trojan horse, only need to check the specific port will know what infected with Trojan, so many new Trojans are now in between 1024 to 65535 choose a port number as the Trojan port (generally not less than 1024 of the selected port), so as to give the Trojans difficult judgment. At present, there are many new technologies of Trojan horse, using the technology of non linking and port hiding, some of which can be used by DLL injection technology. The emergence of these new technologies has brought no small challenge to the network intrusion detection technology.

2) based on the detection of feature code [25]

Feature code is the anti-virus software company in the analysis of Trojan or virus, the virus samples extracted binary string, and can identify the virus samples through these features. Detection technology based on feature code is first used in virus and Trojan horse, the same technology can also be used to check the zombie process, it is also one of the most popular techniques to detect zombies. Usually in a new zombie program appears, carries on the scanning, according to the characteristics of binary feature file string, establish binary feature library or script feature library, and finally the feature matching in the scan file. Feature extraction is based on a basic point: whether it is virus, Trojan, worm or zombie program, whether it is encrypted or deformed, in the binary file, there must be a constant component. Extract the invariant parts, and make it characteristic, can be used to distinguish between different characteristics of the characteristics of the zombie library. Currently more popular three kinds of feature code formats: MD5 format, string format, and both ends of the parity check and format.

3) based on the behavior of the detection

All malicious code, whether it is virus, Trojan or worm, in the initial stage of implantation, there are some common, such as: in the registry settings from the start, to hide their own processes, documents. Based on the behavior of the inspection is to determine the extent of these suspicious. In the previous inspection, which is based on the feature code, it is widely spread in the zombie process after the analysis of the sample. And once the system is known to be implanted into a zombie process, the malicious code will be in a short period of time to get the machine's sensitive information. Therefore, in the reality of the detection technology, the urgent need to check the unknown zombie program. At present, a lot of anti-virus software, active defense, including the behavior analysis, feature scanning and other functions.

4) based on virtual machine detection [26]

Virtual machine technology is not a new thing, anti-virus software provided by the heuristic search technology, its essence is virtual machine technology. Research on the virtual machine technology anti-virus software vendors have never stopped, "virtual machine anti-virus technology" is a virtual run time environment in memory, the execution of the program will be detected in the virtual environment, according to their behavior or release of known virus signatures, to determine whether the virus program. Virtual machine technology is the main function is to run a certain rules of the description language. Objectively speaking, in all kinds of virus checking methods, the characteristic value method is the most widely used, the fastest, the most simple and effective method. But due to defects in itself, it is only suitable for known viruses, for unknown viruses, if can make the virus first run for a period of time under control, let it have its own reduction of real environment damage. At present, antivirus software vendors are optimistic about the virtual machine technology, but based on the mainstream technology of virus signature inspection of anti-virus anti-virus software technology is adopted, auxiliary means of virtual machine technology to detect virus is still in a fairly long period of time.

B. Network based Detection

Network based detection systems and tools have a lot of. The lightweight intrusion detection system Snort[27] and Bro[28] system are the two most typical examples. But they and the traditional anti-virus tools have a very similar shortcomings, for the new attack is powerless to check the way. Abnormal intrusion detection system can be described by what is the normal traffic to alleviate this limitation and based on the normal traffic is obviously not the same is considered abnormal flow, but this kind of system may lead to many false positives. Botnet development to the present stage, the relevant technology has been relatively mature. Different botnets can choose different control and command structure, to communicate using a variety of different protocols can be open protocol or homemade proprietary protocol; in addition because the target is different, the botnet attack module

selection in the design of the present development, the diversity of visible, botnet type complex. Accordingly, the detection method for botnet is also very diverse. According to the characteristics of the detection method, we can classify the existing botnet detection methods from many aspects.

1) Protocol related Detection Method

According to the control protocol of the Botnet, the main protocol used by the botnet is IRC, HTTP and P2P protocol. Due to the P2P and HTTP zombies are relatively small, the emergence of these well-known zombies at the beginning, the general use of the port matching, feature matching and other methods to identify such a class of. The detection of protocol is mainly focused on the research of [29] protocol based on IRC protocol. R. Binkley[30] J. et al. Proposed an anomaly detection algorithm based on the presence of a botnet. The Botsniffer method proposed by Gu[31] et al, method of network anomaly detection based on IRC and HTTP types of bots, depend on the methods of supervision belong to the same botnet within the network of zombie behavior will be showing the relevance and similarity of space-time. The Rishi method proposed by Goebel[32] et al is a detection method based on feature matching IRC user nickname, first use regular expressions to describe the zombie nickname mode, to achieve score function method using n-gram analysis, and then to capture network packets from the nickname field, the score is marked for meet the threshold of zombie hosts.

2) Protocol Independent Detection Method

The protocol independent detection method does not rely on the specific communication protocol adopted by the control channel of the Botnet, but is recognized by matching the characteristics of the network characteristics of the botnet or network behavior characteristics to identify the host. The network behavior includes the control of the botnet communication behavior, the use of the communication behavior of the hidden technology and the malicious network behavior of the botnet. Chang and Daniels[27] proposed a method based on the behavior of nodes to capture the cluster nodes of the Botnet, so as to check the communication of the botnet. Currently most of the botnet is used to DDoS attacks and send spam platform. So many researchers through the analysis of the characteristics of spam to detect the presence of the Botnet, but also a lot of researchers for the flow of DNS to check the presence of zombies [28].

4. SUMMARY

With the continuous progress of the information society in the process of information society, the attacks against computer networks are emerging, the botnet has gradually become one of the biggest threat to the modern internet tool. Botnet fusion of the advantages of numerous malicious code, provides an ideal platform to launch a variety of attacks. Traditional centralized botnet has been tested for many years, and finally shows its inherent structural defects, and the P2P botnet is becoming a hot research topic in the strong robustness of its communication architecture.

5. REFERENCES

[1] R.Puri, Bots&botnet: An overview, SANS White Paper, 2003,http://www.sans.org/reading_room/whitepapers/malicious/1299.php

[2] G.Eason, B.Noble, I.N.Sneddon, On Certain integrals of EggDrop:Open source IRC bot,1993,Http://www.eggheads.org

[3] J.Nazario, BlackEnergy DDoS Bot Analysis, Arbor Networks, 2007:26-30

[4] S.Stover, D.Dittrich, J.Hernandez, et al. Analysis of the Storm and Nugache Trojans:P2P is here, In proceedings of USENIX,2007:18-27

[5] Moheeb Abu Rajab, Jay Zarfoss, Fabian Monrose, and Andreas Terzis. A multifaceted approach to understanding the botnet phenomenon. In Proc. of the 6th ACM SIG-COMM Conference on Internet Measuremen, Rio de Janeiro, Brazil, October 2006.

[6] Michael Bailey, Evan Cooke, Farnam Jahanian, Yunjing Xu, and Manish Karir. A Survey of Botnet Technology and Defenses. In Proc. of the 2009 Cybersecurity Applications & Technology Conference for Homeland Security, March 2009.

[7] Liang Xie and Sencun Zhu. A Feasibility Study on Defending Against Ultra-Fast Topological Worms. In Proc. of The 7th IEEE International Conference on Peer-to- Peer Computing (P2P'07), Galway, Ireland, September 2007.

[8] Ryan Vogt, John Aycock, and Michael Jacobson. Army of Botnets. In Proc. of the 2007 Network and Distributed System Security Symposium (NDSS), Febuary 2007.

[9] Julian B. Grizzard, Vikram Sharma, Chris Nunnery, Brent ByungHoon Kang, and David Dagon. Peer-to-Peer Botnets: Overview and Case Study. In Proc. of the 1st USENIX Workshop on Hot Topics in Understanding Botnets (HotBots '07), Cam- bridge, MA, April 2007.

[10] Phillip Porras, Hassen Saidi, and Vinod Yegneswaran. A Multi-perspective Analysis of the Storm (Peacomm) Worm. Technical report, SRI, November 2007.

[11] C.Kalt, Internet Relay Chat:Client protocol, Reauest for Comment(RFC)2812(Informational),2000

[12] Sinit P2P Trojan analysis. Http://www.lurhq.com/sinit.html

[13] E.Florio, M. Ciubotariu, Peerbot: Catch me if you can, White Paper, Symantec Security Response,2007

[14] Jun Li, Toby Ehrenkranz, Geoff Kuenning, Simulation and Analysis on the Resiliency and Efficiency of malnets. Workshop on Principles of Advanced and Distributed Simulation (PADS'05),2005

[15] Ping Wang, Lei Wu, Ryan Cunningham, and Cliff C. Zou. Honeypot Detection in Advanced Botnet Attacks. In International Journal of Information and Computer Security (IJICS), 4(1), 30-51, 2010.

[16] Lasse Trolle Borup. Peer-to-Peer botnet: a case study on Waledac. Mathematical Modelling. 2009

[17] Ben Stock, Jan Gobel, Markus Engelberth, Felix C.Freiling, and Thorsten Holz. Walowdac-Analysis of a Peer-to-Peer Botnet. 2009 European Conference on Computer Network Defense.2009

[18] Zhaosheng Zhu, Guohan Lu, Yan Chen, Zhi Judy Fu, Phil Roberts, and Keesook Han. Botnet Research Survey. In Proc. of the 32nd Annual IEEE International Computer Software and Applications (COMPSAC '08), July 2008.

[19] Guenther Starnberger, Christopher Kruegel, and Engin Kirda. Overbot - A botnet protocol based on Kademlia. In Proc. of the 4th International Conference on Security and

Privacy in Communication Networks (SecureComm), September 2008.

[20] Clarke R.Building an Early Warning System in a Service Provider Network. Black Hat Briefings Europe, 2004

[21] P.Szor, The Art of Computer Virus Research and Defenses, Addison-Wesley Professional,2005

[22] M. Roesch, Snort-lightweight intrusion detection for networks, In Proceedings of the 13th systems Administration Conference(LISA'99), Seattle,Washington,USA,1999

[23] V. Paxson, Bro: A System for Detecting Network Intruders in Real Time, In Proceedings on the 7th USENIX Security Symposium(Security'98), San Antonio, Texas,USA,1998

[24] D.Wagner and P.Soto. Mimicry attacks on host based IDS. ACM CCS, 2002

[25] Su Chang and Thomas E. Daniels. P2P botnet detection using behavior clustering & statistical tests. In Proc. of the 2nd ACM workshop on Security and artificial intelligence (AISec '09), Chicago, November 2009.

[26] Ulrich Bayer. TTanalyze:A tool for Analyzing Malware, Master Thesis of Vienna University of Technology, 2006

[27] Evan Cooke, Farnam Jahanian, and Danny McPherson. The Zombie Roundup: Understanding, Detecting, and Disrupting Botnets. In Proc. of the Workshop on Steps to Reducing Unwanted Traffic on the Internet (SRUTI'05), July 2005.

[28] Ricardo Villamarin-Salomon and Jose Carlos Brustoloni. Bayesian bot detection based on DNS traffic similarity. In Proc. of the 24th Annual ACM Symposium on Applied Computing (SAC '09), Honolulu, Hawaii, March 2009.

[29] Y. Chen. IRC-based botnet detection on high-speed routers, 2006. ARO/DARPA/DHS Special Workshop on Botnet.

[30] J. R. Binkley and S. Singh. An algorithm for anomaly-based botnet detection. In USENIX 2nd Workshop on Steps to Reducing Unwanted Traffic on the Internet (SRUTI 06), June 2006.

[31] Guofei Gu, Junjie Zhang, and Wenke Lee. BotSniffer: Detecting Botnet Command and Control Channels in Network Traffic. In Proc. of the 15th Annual Network and Distributed System Security Symposium (NDSS'08), February 2008.

[32] J. Goebel, T.Holz, Rishi.identify bot contaminated hosts by irc nickname evaluation, In Proceeding of the first conference on First Workshop on Hot Topics in Understanding Botnets, Berkeley,CA,USA,2007,USENIX Association.