ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

An Efficient Scalable System for Peer-To-Peer Botnet Detection

L. Prabhu, K.Dhivya, V.Geetha

M.Tech, Department of IT, J.J College of Engineering and Technology, Trichy. Tamil Nadu, India

M.Tech, Department of IT, J.J College of Engineering and Technology, Trichy. Tamil Nadu, India

M.Sc, Department of Computer Science, Bishop Heber College, Trichy, Tamil Nadu, India

ABSTRACT: Peer-to-Peer botnets are legally taken by botmasters for the quick recovery against taking down efforts of the system. But it's a harder one for the botmasters, because modern botnets are hidden and performing malicious activities it makes the process inefficient. Additionally because of sudden growth of the network traffic there was an ability to enlarge the malicious activities of the system. In this paper, the hidden P2P botnets are identified using botmasters. Our system first identifies the system which is all engaged in p2p communications. Then it analysis the behavioral characteristics of identifying P2P and it finds the difference between P2P botnet traffic and legal p2p traffic. By doing this our scalability of our system increases. Alternatively it also increases the detection accuracy as well as scalability of our system.

KEYWORDS: Botnet, P2P, Botmaster

I.INTRODUCTION

A botnet also known as a zombie army is a number of Internet computers that, although their owners are unaware of it, have been set up to forward transmissions to other computers on the Internet. Any such computer is referred to as a zombie - in effect, a computer "robot" or "bot" that serves the wishes of some master spam or virus originator. Most computers compromised in this way are home-based [9]. According to a report from Russian-based Kaspersky Labs, botnets not spam, viruses, or worms currently pose the biggest threat to the Internet. A report from Semantic came to a similar conclusion. An increasing number of home users have high speed connections for computers that may be inadequately protected. A zombie or bot is often created through an Internet port that has been left open and through which a small Trojan horse program can be left for future activation. At a certain time, the zombie army "controller" can unleash the effects of the army by sending a single command, possibly from an Internet Relay Channel (IRC) site. The computers that form a botnet can be programmed to redirect transmissions to a specific computer, such as a Web site that can be closed down by having to handle too much traffic - a distributed denial-of-service (DDoS) attack - or, in the case of spam distribution, to many computers. The motivation for a zombie master who creates a DDoS attack may be to cripple a competitor. The motivation for a zombie master sending spam is in the money to be made. Both of them rely on unprotected computers that can be turned into zombies.

The most popular method of Command and control structures in recent years has been the use of Internet Relay Chat (IRC), either in standard form or through use of customized implementations of IRC servers and clients intended to thwart mitigation efforts. Programs that use IRC for Command and control structure are alsoknown as bots, and a distributed network of bots is known as a botnet[1,5]. Botnets will likely be around for some time, causing a huge amount of grief for network operators, victims of DDoS attacks, and other victims, but IRC-based bots are not the be-all and end-all, and the advent of Peerto-Peer (P2P) mechanisms for C2 may spell the eventual death of IRC as a means of C2.

Botnets may structure their Command and control channel in different ways. In the case of centralized architecture bots in botnet contact one or few servers that is controlled by botmasters [7]. The problem in such a system is that there will be single point of failure. To overcome this problem, botmasters are using more resilient C&C architecture using a

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

peer-to-peer structure. In such architecture bots form an overlay network in which any of the nodes can be used by the botmaster to distribute commands to other peers. Detecting botnets is of great importance. Designing an effective P2P botnet detection system is faced with several challenges. A bot-compromised host may exhibit mixed patterns of both legitimate and botnet P2P traffic[11]. Botnets tend to use increasingly stealthy ways to perform malicious activities that are extremely hard to be observed in the network traffic. Another challenge is that volume of network traffic grows rapidly, so deployed detection system must be able to process more information efficiently

II. RELATED WORK

Web-account abuse attack was first reported in 2010[10], in which millions of botnet email accounts were created from major Web email service providers in a short duration for sending spam emails. While each user is required to solve a CAPTCHA test to create an account, attackers have found ways to bypass CAPTCHAs, for example, redirecting them to either spammer-controlled Web sites or dedicated cheap labor 2. The solutions are sent back to the bot hosts for completing the automated account creation. Trojan Hotlan is a typical worm for such automated account signup. Today, this attack is one of the major types of large-scale botnet attacks, and many large Web email service providers, such as Hotmail, Yahoo! Mail, and Gmail, are the popular attack targets. To our best knowledge, BotGraph is one of the first solutions to combat this new attack.

The Web-account abuse attack is certainly not the first type of botnet spamming attacks [8]. Botnet has been frequently used as a media for setting up spam email servers. For example, a backdoor rootkit Spam-Mailbot.c can be used to control the compromised bots to send spam emails.

Storm botnet, one of the most widespread P2P botnets with millions of hosts, at its peak, was deemed responsible for generating 99% of all spam messages seen by a large service provider [7, 8]. Although our work primarily focuses on detecting the Web-account abuse attack, it can potentially be generalized to detect other botnet spamming attacks. In this general problem space, a number of previous studies have all provided us with insights and valuable understanding towards the different characteristics of botnet spamming activities [9]. Among recent work on detecting botnet membership [7,8,9,11], SpamTracker and AutoRE also aim at identifying correlated spamming activities and are more closely related with our work.

Storm botnet, one of the most widespread P2P botnets with millions of hosts, at its peak, was deemed responsible for generating 99% of all spam messages seen by a large service provider [7, 8]. Although our work primarily focuses on detecting the Web-account abuse attack, it can potentially be generalized to detect other botnet spamming attacks. In this general problem space, a number of previous studies have all provided us with insights and valuable understanding towards the different characteristics of botnet spamming activities [9]. Among recent work on detecting botnet membership [7,8,9,11], SpamTracker and AutoRE also aim at identifying correlated spamming activities and are more closely related with our work.

In addition to exploiting common features of botnet attacks as SpamTracker and AutoRE do, BotGraph also leverages the connectivity structures of the user-user relationship graph and explores these structures for botnet account detection. To achieve the already mentioned model, our system involves multiple components. The first one is flow-clustering-based analysis is used to analysis the hosts which are running the P2P applications [2, 3]. These approaches are differ in many ways, they are (1)encryption will make the content signature useless.
(2) It doesn't rely on transport layer because it violated by P2P applications.
(3) It doesn't need any training for the data to corporate the system because it's a very critical to identify the network traffic of botnets P2P before it will be identified.
(4) It identifies the p2p application rather than a very specific one.There are many outlooks are there to find the P2P botnets.

- In this it doesn't create the destructive things to be observable.
- It doesn't contain any botnet specific information for identification purpose.
- It detects both P2P bot as well as legal P2P application simultaneously.
- It also provide high scalability features.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

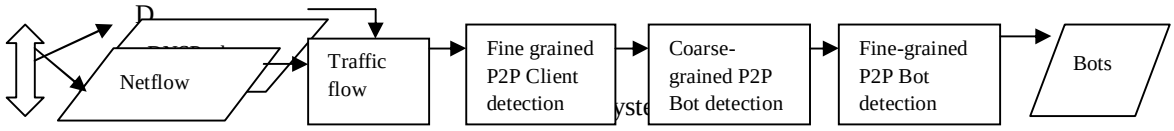
III PROPOSED SYSTEM

In this paper a novel scalable botnet detection system has been proposed. This detection system is capable of detecting stealthy P2P botnets whose malicious activities may not be observable in the network traffic [4].Our system aims to detect stealthy P2P botnets even if P2P botnet traffic is overlapped with the traffic generated by legal P2P applications running on the same compromised host. Our system identifies P2P bots within a monitored network by detecting the C&C communication patterns that characterize P2P botnets. The high scalability of the system can be achieved by using the following techniques.

1. P2P traffic profiling algorithm that is used to built the statistical fingerprints for various P2P applications.
2. A flow-clustering based analysis approach to identify host that engaged in P2P communications [6].
3. A scalable design based on an efficient detection algorithms and parallelized computation.
4. A prototype system based on real world network traffic which demonstrated high detection accuracy.

The new model eradicates the necessity of keeping failed connections. Clustering based client detection algorithm enhances the efficiency of the model. The system is parallelized to boost scalability and efficiency. The proposed system is effective over a large range of parameter values.

IV SYSTEM DESIGN



A P2P botnet depends on P2P protocol to create transmission through C&C channel with botmasters. A P2P bots have a common network traffic patterns that is used to evolve P2P client applications as well as legal applications. It divides in two phases (1) In first phase, Its aims is to detect all the network traffic which involved in peer-to-peer communications. In figure1 we inspect the network flow at the edge and filter it to discard the flow which should be created unexpectedly by peer-to-peer applications. From that we can analysis the network traffic and flow created by peer-to-peer clients. (2) In second phase, our system will examine the network traffic generated by both legal P2P clients and P2P bots. Then we explore the active time of the peer-to-peer client and recognize it as candidate P2P bot and if there is a continuous change in host. We further analyze it by detecting 2 candidates P2P bots.

4.1 Finding out Peer-to-Peer Client

A Filter

Filter component is used to filter the network traffic that is unrelated to P2P communications. This can be achieved by analyzing DNS traffic. P2P clients contact their peers by looking up IPs from a routing table for the overlay network rather than resolving a domain name. Most non P2P applications often connect to a destination address resulting from domain name resolution. This simple filter can eliminate a very large percentage of non P2P traffic and helps in retaining P2P communication.

B Peer-to-Peer Detector

Client detector helps in detecting P2P clients by analyzing the remaining network traffic .For each host within the monitored network we identify two flow sets which contains flows related to sucessful outgoing TCP and UDP connections.TCP connection is considered successful if SYN,SYN/ACK,ACK I handshake is available.

UDP connection is considered successful if there is at least one request and a consequent response packet is found. In order to detect P2P clients we first consider the fact that each P2P client frequently exchanges control messages with other peers. Even though the characteristics of these messages such as size and the frequency of the exchanged packets are same ,they vary depending upon the P2P protocol and network in use. If two network flows are generated by the same P2P applications they carry the same control messages. In addition P2P client exchanges the control messages with large number of peers that is distributed in different networks. The destination IP addresses of network flows that

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

carry these control messages will spread across a large number of networks where each network can be represented by its BGP prefix.

TABLE I
SUMMARIES OF FINGERPRINT CLUSTERS

TRACE	FINGERPRINTS
T-BITTORENT	1 1 145 319, UDP 1 1 109 100, UDP 1 1 146 340, UDP 1 1 346 170, TCP 1 1 145 310, UDP
T-BITTORENT-2	1 1 145.01 317.66, UDP 1 1 109 100, UDP 1 1 146 342, UDP 1 1 346 170, UDP 2 2 466 461, UDP

We implemented the flow analysis component and identified fingerprint cluster for the sample P2P traces including two Bit torrent traces(T-Bit torrent, T-Bittorrent-2) and Skype trace. The summaries of the fingerprint cluster illustrated in table 1, which can successfully cluster flows with similar sizes. Particularly, manual investigation of the flow payload has confirmed flows corresponding to two fingerprint clusters, "(1 1 145 319,UDP)" and "(1 1 109 100,UDP)", are used for node discovery and exchanging ping/pong messages respectively.

4.2. Finding Peer-to-Peer Bots

A Coarse-Grained Detection of P2P Bots

Since bots are malicious programs used to perform profitable malicious activities, they represent valuable assets for the botmasters, who will intuitively try to maximize utilization of bots. This is particularly true for P2P bots because in order to have a functional overlay network, a sufficient number of peers needs to be always online. In other words the active time of the bots should be comparable with the active time of the underlying compromised system. If this was not the case the botnet overlay network would risk degenerating into a number of disconnected subnetworks due to the short lifetime of each single node. In contrast the active time of the legitimate P2P applications determined by users, which is likely to be transient.

B Fine-Grained Detection of P2P Bots

The objective of this component is to identify P2P bots from all persistent P2P clients. We leverage one feature; the overlap of peers connected by two P2P bots belonging to the same P2P botnets is much larger than that contacted by two clients in the same legitimate P2P network. Assume two hosts in the monitored network are running the same legitimate P2P file sharing application (e.g., Emule). Users of these two P2P clients will most likely have uncorrelated usage patterns. It is reasonable to assume that in the general case the two users will search for and download different contents(e.g., different media files or documents)from the P2P network. This translates into a divergence between the set of IP addresses contacted by hosts.

The reason is that the two P2P clients will tend to exchange P2P control messages with different set of peers which "own" the content requested by their users, or peers that are along the path towards the content. On the contrary, if hosts are compromised with P2P bots, one common characteristic of bots is that they need to periodically search for commands published by the botmasters. This typically translates into a convergence between sets of IP connected by hosts.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

An ISO 3297: 2007 Certified Organization Vol.3, Special Issue 1, February 2015

National Conference on Computing and Communication (NC^3 2K15)

Organized by

Dept. of CSE, CARE Group of Institutions, Tiruchirapalli-620009, India on 27th February 2015

V. CONCLUSION

In this paper we proposed a novel scalable P2P botnet detection system that is able to identify stealthy P2P botnets. To perform this task statistical fingerprints of P2P communications have been derived to detect P2P clients and further distinguish between those that are part of legitimate P2P networks and P2P bots. The results shows that the proposed system accomplishes high accuracy on detecting stealthy P2P bots and great scalability.

REFERENCES

[1] A. Ramachandran, N. Feamster, and S. Vempala, " Filtering spam with behavioral blacklisting", In *ACM CCS*, 2007.
[2] M. A. Rajab, J. Zarfoss, F. Monrose, and A. Terzis, " A multifaceted approach to understanding the botnet phenomenon", In *IMC*, 2006.
[3] D. Dittrich and K. E. Himma, "Active Response to Computer Intrusions,"in *The Handbook of Information Security*, edited by H. Bidgoli (Wiley,New York, 2005).
[4] J.Zhang, R.Perdisci and W.Lee, "Building a scalable system for stealthy P2P Botnet Detection", IEEE Transactions on Information Forensics and Security,Vol.9, No.1,January 2014.
[5] Y. Zhao, Y.Xie, F.Yu and Y.Yu, "Botgraph : Large scale spamming botnet detection", in Proc. 6th USENIX NSDI, 2009, pp 1-14.
[6] G.Gu ,R.Perdisci, J.Zhang and W.Lee, "Botminer: Clustering analysis of network traffic for protocol and structure independent botnet detection", in Proc. UNISEX security, 2008, pp.139-154.
[7] M. A. Rajab, J. Zarfoss, F. Monrose, and A. Terzis, " A multifaceted approach to understanding the botnet phenomenon" In *IMC*, 2006.
[8] D. S. Anderson, C. Fleizach, S. Savage, and G. M. Voelker. Spamscatter: Characterizing internet scam hosting infrastructure. In *USENIX Security Symposium*, 2007
[9] Y. Xie, F. Yu, K. Achan, R. Panigrahy, G. Hulten, and Osipkov. Spamming botnets: Signatures and characteristics. In *SIGCOMM*, 2008
[10] Y. Yu, M. Isard, D. Fetterly, M. Budiu, U. Erlingsson,P. K. Gunda, and J. Currey, " DryadLINQ: A system for general-purpose distributed data-parallel computing using a high-level language", In *OSDI*, 2008.
[11] G.Bartlett, J.Heidemenn and J. Pepin, "Estimating P2P traffic volume at USC",USA,Tech. Rep. ISI-TR-2007.