International Journal of Security and Its Applications
Vol.8, No.3 (2014), pp. 87-96
http://dx.doi.org/10.14257/ijsia.2014.8.3.10

A P2P Botnet Detection Method Used On-line Monitoring and Off-line Detection

Yuhui Fan[1] and Ning Xu[2]

Department of Computer and Information Engineering, Huainan Normal University, Huainan, China
[1]122336956@qq.com, [2]amyxuning@sina.com

Abstract

P2P botnet has become a significant threats in security network. In this paper, we propose a new method to detection the P2P botnet through the analysis of the P2P botnet host's life cycle, use the method of off-line detection to find the suspected botnet hosts, and determine the P2P botnet host through online monitoring method. In this way, the efficiency and the accurate rate of P2P botnet detection have raised then only use one method, and reduce the harm of the P2P botnet.

Keywords: *P2P botnet; life cycle; detection; monitoring; netflow*

1. Introduction

The number of Chinese Internet users has reached 450 million, but CNCERT claimed the number of IP for botnet control server in 2010 was 13782 in The 2010 China Internet Network Security Report. Domestic IP for Botnet Control Server was 7251, while IP for botnet control ever outside our country was 6531.The total number of the controlled hosts by Botnet was 5622023, among which was 470120 within the country and 5151903 overseas [1]. Government departments, commercial institutions and common users suffered a lot from net theft, DDoS attack and a mass of spam caused by Botnet, which has been affecting seriously Internet security.

2. Correlational Rationale

Botnet commands and communication mechanism have evolved into three forms through years: (1) Botnet based on traditional IRC protocol; (2)Botnet based on HTTP, DNS protocol; (3)Botnet based on P2P protocol.

Commands and control mechanism based on traditional IRC protocol and HTTP protocol are to have centralized control positions, which make the botnets based on client-server framework easily tracked, detected and countered, once the defenders obtain the bot, they can easily identify the position of botnet controller so as to monitor and know well the overall information of botnet. Then defenders can also turn network controllers off to wipe off the threat. While the newly emerging botnet based on P2P protocol (as shown in Figure 1) has different control nodes distribution throughout the net, which do no virtual harm to the whole net when one or several nodes destroyed. So botnet based on P2P protocol is more elusive and has more damage resistance than ever.

ISSN: 1738-9976 IJSIA

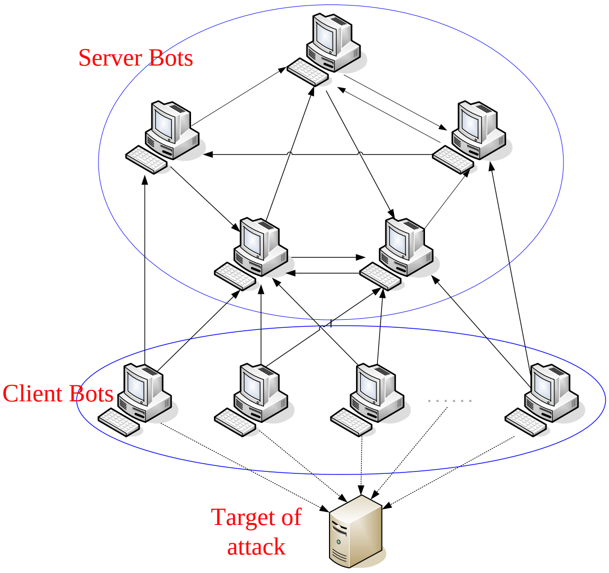

Figure 1. The Botnets Model based on P2P Protocol

There are two ways to detect P2P botnet at present: one is off-line detection, that is, collect the network flow data over a period of time and analyze it with arithmetic; the other is on-line detection, namely, monitor the potential botnet, collect and analyze the network flow, and report the suspected hosts. The former can distinguish the existing botnet during the detection but cannot take control measures immediately, while the later can detect the botnet instantly but is powerless to deal with the mass of net flow. In all, it is urgent to detect, distinguish and control botnets in the network study.

3. Behavioural Analysis of P2P botnet's Survival Features

The present detections of P2P botnets focus on the analysis of the botnet flow [2~7], thus they are distinguished from other P2P applications. After that, the infected hosts by botnets are recognized through off-line detection and on-line detection. While this paper observes the different competency behaviors of botnet in different stages, analyses with synthetic judgment, thereby obtains the information of P2P botnet host. The design here suggests the steady monitoring of the inactive and no-harm botnet hosts but takes measures to stop the contact between the control nodes and the attack-stage botnet hosts.

3.1. The Life Cycle of P2P Botnet

Wang(Wang Ping,2009) and his members group the botnets into three stages according to the execution sequence: selecting members, forming botnet, awaiting commands, selecting members means to infect lots of hosts, and forming botnet means to make up the botnet by uniting the infected hosts through P2P, and awaiting commands means the infected hosts waiting for the commands from the botnet controller [8].

P2P botnets are grouped into initial stage, trance stage and attack stage according to network flow suggested by Chai (Chai Sheng, 2010). During the initial stage, noticeable features, which can be detected, out of P2P botnets are the drastic increase in linking numbers and the low rate of linking success. During the trance stage, the trance host is not involved in other, but linked to the equal node in the botnet to keep the simple contact. The great amount of session flow, the same communication traffic and the small amount of communication are typical of the trance hosts. While in the period of attack stage, lots of commands are emitted from the P2P botnets controllers to start the cyber attack like attacking the intended DDOS, stealing the information of the present host, sending out the jumble of junk mail .The last one among the attack usually causes the obvious flow, which can be easily detected [9].

3.2. Important Features of P2P Botnets

In the above analysis, the typical features of P2P botnets in different stages, according to the grouping criteria from Chai (Chai Sheng, 2010), are helpful to analyze and detect its existence, as in [9]. The important features are:

(1)P2P botnet hosts produce many ICMP reports with low rate of successful linking during the initial stage,

(2)P2P botnet hosts are linked to many nodes with the same communication traffic during the trance stage,

(3)P2P botnet hosts produce too much SMTP contact with too much similar data package of destination address during the attack stage.

Above all, the features in (1), (2) can be extracted off-line to detect the address of P2P botnets hosts and the hosts can be monitored on-line. If behaviors like in (3) are spotted, the hosts can be stopped immediately.

4. Sign for Detection of Botnet Hosts' Behavior

4.1. Design Ideas

In most cases, P2P botnet hosts are in the initial and trance stages and do no harm to others in their life cycle. They begin to attack other users on the net, like sending spam or attacking DDOS only when receiving commands from the controllers of the botnets, as in [10]. So, the present detection of them is mainly based on their features of networks, which may be easily mixed with the normal P2P operation and give rise to the misreport. However, this paper designs a detection of combining the off-line detection and on-line monitoring to minimize the impact from the P2P botnet hosts to other users. It is shown in Figure 2.

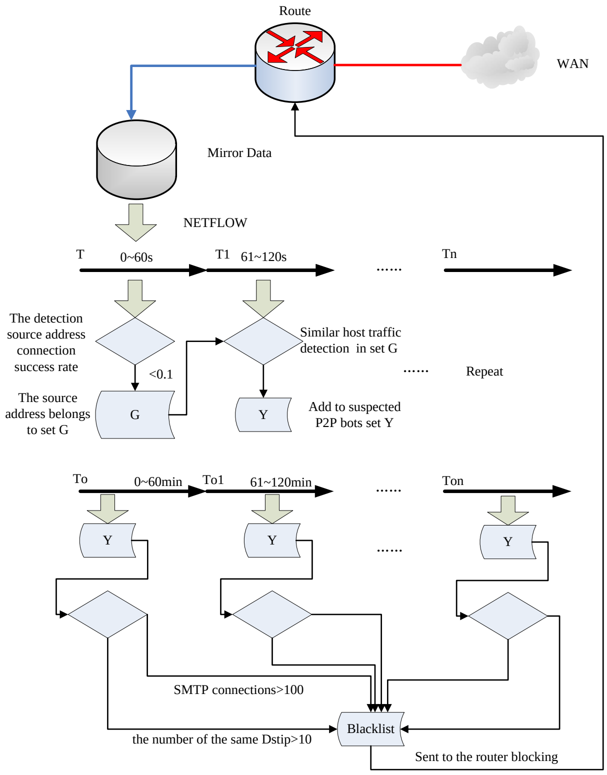
Route
WAN
Mirror Data
NETFLOW
T
0~60s
T1
61~120s
......
Tn
The detection source address connection success rate
<0.1
The source address belongs to set G
G
Similar host traffic detection in set G
......
Repeat
Y
Add to suspected P2P bots set Y
To
0~60min
To1
61~120min
......
Ton
Y
Y
......
Y
SMTP connections>100
the number of the same Dstip>10
Blacklist
Sent to the router blocking

Figure 2. The Off-line Detection and on-line Monitoring to Minimize the Impact from the P2P Botnet Hosts

4.2. Design of the Off-line Detection

After a certain period of collection of net flow from the network exports, if the low rate of connection success and many nodes with the same communication traffic distinguished as the behavioral features in those botnet stages, the source host address can be separated as the suspected botnet host address.

(1) Detection of connection success rate

When P2P botnet hosts are linked the botnet nodes in the network, the connection success rate will be very low due to the firewall block, network address conversion or not-online hosts and so on. There will be a lot of "destination unreachable" ICMP package and TCP package different from the normal net communication. After the number of successful linking package received by the source address is divided by that of the requesting linking package, the connection success rate is obtained to tell the abnormal network of the hosts. If the rate is between 0 and 0.1, it shows the low connection between the source address and outside destination address, which can be caused by the suspected botnet hosts, as in [10,11].

Define 1 T: the detection cycle.

Define 2 Flow{Srcip, Dstip, Protocol, Times, Bytes, Sport, Dport}: in a detection cycle expressed as T, the collected net flow data is divided into 7-element data stream collection according to Srcip, Dstip, Protocol, Times, Bytes, Sport, and Dport.

Define 3 SpackIP{Srcip, NumS}: the set include the source IP address as Srcip, and the numbers of SYN packets sent from the source IP address to the different destination addresses as NumS.

Define 4 RpackIP{Dstip, NumD}: the set include the destination IP address as Dstip, and the numbers of SYN/ACK packets sent from the different source IP addresses to the destination addresses as NumD.

Define 5 G{ Srcip1, Srcip2, …, Srcipn}: the set include the local suspected hosts are defined as hosts set expressed as G { Srcip 1, Srcip 2, …, Srcip n}.

Define 6 V{IPAdress, Rate}: the set include local IP address as IPAdress, and in a detection cycle, this local IP address's connection success rate as Rate.

Based on the above definition, the connection success rate algorithm is as follows:

```
The Flow{} in ascending order according to the Srcip;
For each flow{}
{
  i=0;
  if (Flow[Srcipi]=Flow[Srcipi+1])
       {SpackIP[Srcipi]=Flow[Srcipi];
       i++;
}
  Else
       SpackIP[NumSi]=i;
}
The Flow{} in ascending order according to the Dstip;
For each flow{}
{
i=0;
  if (Flow[Dstipi]=Flow[Dstipi+1])
       {RpackIP[Dstipi]=Flow[Dstipi];
       i++;
}
  Else
       DpackIP[NumDi]=i;
}
i=0;
for each DpackIP{}
{
  i=0;
       for each SpackIP{}
            {
            j=0;
            if(DpackIP[Dstipi]=SpackIP[Srcipj])
                {
V[IPAddressi]= SpackIP[Srcipj]);
                V[Ratei]= DpackIP[NumDj]/ SpackIP[NumSi];
                }
            j++;
}
  i++;
}
```

So we got the set V, the set V is that the number of successful linking package received by the source address is divided by that of the requesting linking package. If the rate is less than 0.1, the source address belongs to set G, it is shown in Figure 3.

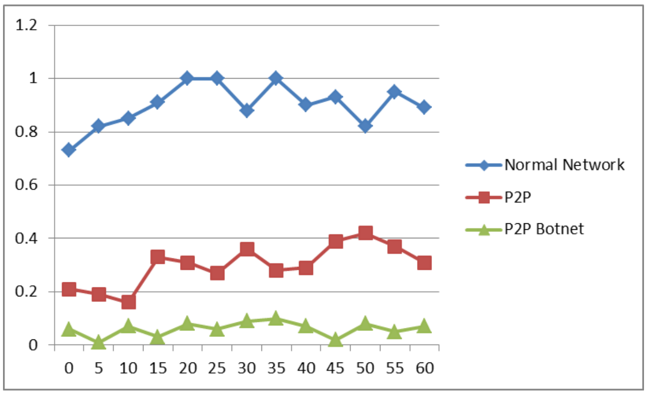

Figure 3. The Rate of Connection Success

The connection success rate is helpful to distinguish the abnormal behavior of intranet hosts. But there will be many misreports if the detection is just based on that. So the further analysis is necessary to get the accurate information of P2P botnet hosts.

(2) Detection of similarity of communication traffic

Since P2P botnet hosts will keep the continuous contact with the botnet nodes to wait for the controller's commands after the linking, and P2P botnet hosts mostly are in this trance stage in its life cycle, there will be P2P botnet hosts behaviors like a great deal of nodes linking, similar communication traffic between nodes and little communication traffic.

Sang-K yun Noh (2009) proposed the design of detecting P2P botnet based on Multi-Phased Flow Model [12]:

Define the sampling period, T1=T+1, that is one detection of successful linking ration moves down to the next sampling period.

Define the sampling Time, Time = 60s, 7-element data stream collection is obtained respectively.

Based on the suspected local hosts collection, according to the source IP address, 7-element data stream set in 60 seconds can be obtained.

Markov chain then is employed to calculate the change in those status values, the result of which is compared to the normal network flow to tell whether there is the botnet host.

After that, the source IP address is recounted to form the set Y. The collection Y includes the suspected botnet hosts' source IP address of the further screening.

(3) Feasibility of off-line detection

Since detection of similarity of communication traffic follows detection of connection successful rate, the detection range is greatly reduced, the speed increased and misreport ratio decreased. But those two detections are based on 2 sampling time, and are likely to fail to report. For this, sampling interval and repeated sampling are suggested to deal with the detection. The less the sampling interval is, the more the infected hosts are to be distinguished, the less the failure to report is. However, the oversampling badly decreases computational speed, so the paper here suggests the off-line detection of those two detections in the beginning and monitors the IP address of local hosts in set Y on-line.

4.3. Design of On-line Monitoring

The suspected botnet hosts through the off-line detection in the address set Y are all in their initial stage and trance stage and do no harm to other hosts or network. There will be misreports if they are judged as botnet hosts. So the paper here suggests the on-line and continuous monitoring of the suspected hosts to precisely locate P2P botnet hosts. Once there is the harmful behavior, the suspected host will be stopped to minimize the harm.

Exports router cannot detect the behavior since the P2P botnet hosts infect the nodes, steal the information and monitor the net flow in the interior of local network. DDOS attack takes up little time in their life cycle. Sending junk mail is the main way to make profits by the botnet controllers, which is the typical behavior to be easily detected [13].

(1) Design of monitoring

There will be too much SMTP connections to the mail server after P2P botnet hosts receive the controller's command of sending spam. Besides that, there will be SMTP reports of the same address from the same P2P botnet hosts. Thus, the design here is based on the two: too much SMTP connections and SMTP reports of the same address.

Deng Guoqiang (Deng Guoqiang,2011) employs the sampling time expressed as To in the design[14], and To is 60min. SMTP flow data will be captured in every T2 period. According to7-element data stream collection, the data flow number in the collection Y is counted. If the flow number is over 100, the host can be thought to be sending spam. The source address should be put on blacklists. At the same time, the number of the same destination address with different source address in SMTP flow data should be counted, if the number is over 10, the source address of the hosts is the address of the same P2P botnet, and the source address should be put on blacklists too.

(2) Confirmation of control of P2P botnet hosts

The host address in the blacklists can be confirmed as the member of P2P botnet, and the host can be blocked off at the network exports router. So the host cannot be linked to outer nets including botnets to decrease its danger.

(3) Feasibility of on-line monitoring design

Since the above on-line monitoring is based on the suspected P2P botnet hosts collection, which is detected off-line, the monitoring is more targeted and its conclusion is more accurate. Yet, there may be the failure to report the P2P botnet hosts whose key feature is to send spam in the detection, but the chance of misreport is slim.

5. Simulation Experiment

5.1. The Experimental Set

We use two networked PCs with 4G memory as an experimental platform. We install Vmware on PC A, and simulate 4 P2P zombie host, are A1, A2, A3, A4. The control program is installed on PC B, this PC is used to send control commands to the P2P bots. These hosts are connected to form network, it is shown in Figure 4.

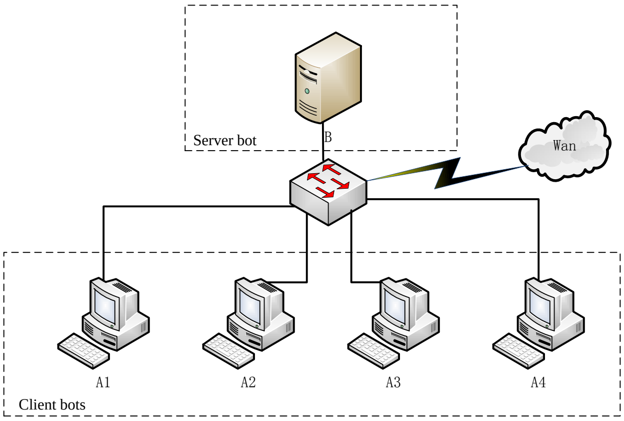

Figure 4. Experiment Environment

The Peacomm P2P zombie virus is deployed in A1 to A4, and we capture each virtual machine's traffic for 2 minutes with an interval of 1 minute, these are network characterized in traffic in P2P bot's initial stage and trance stage. The first time, we mixed the sample data up with the normal network export traffic, and detected of mix data through off-line detection, then we got the suspected bot's IP address. The second time, we copied sample data three times and mixed up the normal network export traffic, then we got the suspected bots' IP address through off-line detection twice.

We capture each virtual machine's net traffic after the control terminal in host B sent the control command of sending spam, then we mixed the sample data up with the normal network export traffic. On-line monitoring is adopted to monitor each collection of the suspected botnet host' IP address which were obtained by two off-line detection experiments. Then we get two blacklists of the P2P botnet hosts.

5.2. Experimental Result

Table 1 shows the quantities of the suspected botnet hosts IP address which were obtained by two off-line detection experiments.

Table 1. The Result of off-line Detection

	The number of bots	The number of suspected bots	The number of false positives	The number of misreport
First time	4	5	2	1
Second time	4	6	2	0

It can be seen from the Table 1 that the false alarm rate and miss rate are higher when detecting in the first test in the case of fewer sample. And more samples are added in the second test, which can greatly decrease the misreport rate.

Table 2 shows the blacklists of P2P botnet hosts which were obtained by two on-line monitoring experiments.

Table 2. The Result of on-line Monitoring

	The number of bots	**The number of suspected bots**	**Blacklist**	**The number of misreport**
First time	4	5	3	1
Second time	4	6	4	0

It can be seen from the Table 2, Since the on-line monitoring is based on the suspected P2P botnet hosts list, which is detected off-line, the botnet hosts' address which failed to be reported in the off-line detection cannot be monitored in the on-line monitoring. And on the other hand, the hosts' address to be misreported in the off-line detection cannot be detected in the on-line monitoring without sending spam. Then they won't be misreported.

Through analysis of the experimental data, we draw an conclusion that enhancing sampling frequency will effectively decrease the misreport rate, and has an influence on the implementing effect of the whole scheme.

6. Conclusion

In all, the combination of the on-line monitoring and the off-line detection is helpful to detect P2P botnet hosts, and greatly relieves the load of detection platform. Off-line detection can be carried out by time period to obtain the blacklists of the suspected botnet hosts. On-line monitoring is supposed to capture and analyze the SMTP data traffic, which is a very small part of the protocol flow at the network outlet. So the monitoring can be implemented easily at the large outlet flow. Through the above analysis, the botnet hosts should be confirmed in the attack stage. There may be the failure to report the botnet hosts if they are not in the attack stage, but if the botnet hosts are stopped and separated in the attack stage, the danger of the botnet is lessened because botnet hosts mainly do harm to others in the attack stage.

Acknowledgment

This research was supported by the Anhui provincial college and university Natural Science Foundation, China (No. KJ2012Z363, No.KJ2013Z302).

References

[1] CNCERT/CC. CNCERT/CC Annual Report 2010. http://www.cert.org.cn/UserFiles/File/CNCERTAnnualReport2010v2.pdf.

[2] J. Zhang, R. Perdisci, W. Lee, U. Sarfraz and L. Xiapu, "Detecting stealthy P2P botnets using statistical traffic fingerprints", Proc of the 41st Annual IEEE/IFIP International Conference on Dependable Systems and Networks Piscatawa:IEEE Press, **(2011)**, pp. 121-132.

[3] S. Saad, I. Traore, A. Ghorbani B. Sayed, D. Zhao, W. Lu, J. Felix and P. Hakimian, "Detecting P2P botnets through network behavior analysis and maching learning", Proc of the 9th Annual International Conference on Privacy, Security and Trust Piscatawa, IEEE Press, **(2011)**, pp. 174-180.

[4] J. Kang, J.-Y. Zhang, Q. Li and Z. Li, "Detecting new P2P botnet with multi-chart CUSUM", Proc of International Conference on Networks Security, Wireless Communications and Trusted Computing. Woshington D C:Computer Society, **(2009)**, pp. 688-691.

[5] G. Guofei, R. Perdisci and Z. Jun-jie, "BotMiner: Clustering analysis of network traffic for protocol-and-sturcture-independent Botnet Detection", Proc of the 17th USENIX Security Symposium. Berkeley:USENIX Association, **(2008)**, pp. 139-154.

[6] M. M. Masud, J. Gao, L. Khan, J. Han and B. Thuraisingham, "A Multi-partition Multi-chunk ensemble Teachnique to Classify Concept-drifting data streams", Proc of the 13th Pacific-Asia Conference on Knowledge Discovery and Data Mining. Berlin:Springer Verlag, **(2009)**, pp. 363-375.

[7] H. R. Zeidanloo, A. Bt Manaf, P. Vahdani, F. Tabatabaei and M. Zamani, "Botnet detection based on traffic monitoring", Proc of the 1st International Conference on Networking and Information Technology. Piscataway, IEEE Press, **(2010)**, pp. 97-101.

[8] P. Wang, L. Wu, B. Aslam and C. C. Zou, "A systematic study on Peer-to-Peer Botnets. Computer Communications and Networks", 2009. ICCCN 2009. Proceedings of 18th Internatonal Conference on. IEEE, **(2009)**, pp. 1-8.

[9] C. Sheng, H. Liang and L. Bo, "The P2P Botnet Online Detect Approach Research", Acta Electronica Sinica, vol. 4, **(2010)**, pp. 906-912.

[10] R. Schoof and R. Koning, "Detecting peer-to-peer botnets", System and Network Engineering. University of Amsterdam, **(2007)**.

[11] L. Jianbo, "Detection of P2P Botnet Based on Analysis of Flow", Computer & Digital Engineering, vol. 3, **(2011)**, pp. 90-91.

[12] S.-K. Noh, J.-H. Oh, J.-S. Lee, B.-N. Noh and H.-C. Jeong, "Detecting P2P botnets using a multi-phased flow model", 3rd International Conference on Digital Society. Cancun: Computer Society, **(2009)**, pp. 247-253.

[13] L. Zhuang, J. Dunagan, D. R. Simon, H. J. Wang and J. D. Tygar, "Characterizing Botnets from Email spam records", Proc of First USENIX Workshop on Large-Scale Exploits and Emergent Threats, **(2008)**.

[14] D. Guo-qiang, L. Zhi-tang, L. Dong and L. Zhan-chun, "Design and implementation of a behavior based algorithm to detect spam zombie client", Journal of Guangxi University(Natural Science Edition), vol. 36, **(2011)**, pp. 100-110.

Authors

Yuhui Fan, received the B.S. degree in Department of Educational Technology from Anhui Normal University, China in 1999, and the M.S. degree in Department of Educational Information Technology from East China Normal University, China in 2008. He is currently an Instructor in the Department of Computer and Information Engineering at the Huainan Normal University. His research interests include network security, computer networking.

Ning Xu, received the B.E. degree in Department of Computer Science from Anhui Normal University, China in 2002, and the M.E. degree in Department of Computer Science from Anhui University, China in 2009. She is currently an Instructor in the Department of Computer and Information Engineering at the Huainan Normal University. Her research interests include network security, computer networking.