Clustering Analysis of Network Traffic for Peer-to-Peer Botnet Detection

Futai Zou

School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University

800 Dongchuan RD. Minhang District, Shanghai, China

China Telephone:+86 21 54740000

Qiaolun Zhang

School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University

800 Dongchuan RD. Minhang District, Shanghai, China China Telephone:+86 21 54740000

*Abstract*—Network traffic is important for tracing back peer-to-peer botnet. Since the network traffic is difficult to catch, we generated the network traffic in our virtual machines using improved mirai botnet. We cluster the network traffic of peer-to-peer botnet to find the abnormal feature and classify the network traffic, which can provide useful information for tracing back peer-to-peer botnet. We found that the traffic of bots and nodes differs and the traffic of peer-to-peer botnet is different to normal network traffic. But it is hard to find the network traffic of C&C server because there is few network traffic of C&C server. The results shows that it is practical to cluster the network traffic to find the feature of p2p botnet.

Keywords- peer-to-peer botnet; network traffic; clustering analysis

# Introduction

Botnet has been widely used to compromise computers , launch distributed denial-of-service attack and other illegal things, which is a great thereat to the web applications. Now hackers has moved to peer-to-peer architecture to implement botnet, which is really difficult to take down. Each bot can act as a server to send command to other bots. On the one hand, we can’t take down the botnet even if we find some bots. On the other hand, the peer-to-peer architecture makes it more difficult to find the botmaster.

However, the command and control protocol is embedded into the peer-to-peer botnet because the botmaster has to send the original command to the bots. And the traffic of C&C server is different from common bots. The traffic now can be found using a 2-tuple “consversation-based” approach with an accuracy of more than 95%( Pratik Narang, Subhajit Ray, Chittaranjan Hota 2014). Moreover, the command and control channels can also be found in network traffic(Guofei Gu, Junjie Zhang, and Wenke Lee 2008).

In this paper, we’ll take a closer look at the network traffic of peer-to-peer botnet. First we’ll implement a botnet environment using mirai. Then we’ll using clustering to analysis the network traffic of peer-to-peer botnet.

# Related Work

The command and control botnet have been widely analyzed by XXX . They observed that . XXX have .

# Hypothesis

Our hypothesis is that the peer-to-peer botnet can be detected by analyzing the communicating and malicious network traffic. There are two hypothesis to detect the peer-to-peer network. The two hypothesis is to find out the abnormal network traffic and explore more information from it to detect the peer-to-peer botnet.

*The discovery hypothesis.* The peer-to-peer network traffic being different from normal network traffic, we can cluster it to identify it from normal network traffic.

*The exploration hypothesis.* The main job of C&C server is to control the bots and the bots is to launch attacks. The functions being different, the network traffic of C&C servers and bots are different. Thus, we can cluster the network traffic of them and find out the C&C server and even trace back to the botmaster.

# Methods

## System Overview

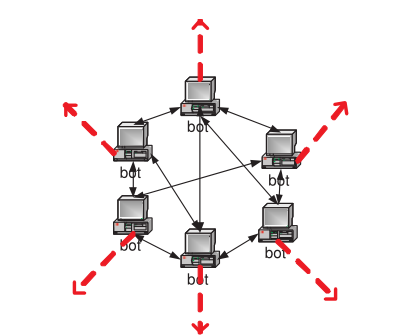


Figure1: Architecture of peer-to-peer botnet

The architecture of peer-to-peer botnet is shown in figure1. The bot connect to each other. Each bot can act as client and server. And the traffic of them is different from normal network traffic. Once a bot receives a command as a client, it will soon act as a server and give the command to lots of other bots. And the network traffic between the bots and servers can be detect by clustering analysis.

Moreover, the network traffic of bots and servers is also different. A server will connect to lots bots but a bot just connect several hosts. Thus clustering analysis is also effective for the detection of C&C server in detecting peer-to-peer botnet.

## Implementation

Materials. The network traffic was generated from the botnet we set up using 100 virtual machines. We simulated a peer-to-peer botnet using the latest released malware mirai. Using mirai, we simulated UDP flood, SYN flood, ACK flood, TCP stomp flood and GRE IP flood. The results of ACK flood was removed because our experiment environment can’t launch effective ACK flood.

Procedure. The figure2 shows the procedure of detecting abnormal traffic. First, a peer-to-peer botnet of mirai was setup. We used 1 VPS for database server, 1 VPS for scan receiver and distributor, 1 server for CNC, and 3 servers for loading. Next, we launched different attack command in C&C server. In the meantime, we saved the network traffic for analyzing. Next we extracted the features of peer-to-peer botnet and used the features to clustering the network traffic. After this, we found out the abnormal network traffic.

Figure2: Detection of abnormal traffic

Next, we used the abnormal network traffic found above to detect C&C traffic. First we gathered the abnormal network traffic. Then, we extract the features that differ between C&C and bots. Using the features, we clustered the network traffic to find out the C&C servers.

Figure3: Detection of C&C traffic

# Results

Table1 shows the results of our experiment. We launched 100 UDP flood attacks and detected 77 times. For SYN flood attack, we detected 66 times among 100 attacks. We launched 100 times TCP stomp flood attacks but just detected 58 times. For GRE IP flood attacks, we detected 83 times among 100 attacks. The C&C server was found 396 times among 400 times of attacks.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Type | UDP flood | SYN flood | TCP stomp flood | GRE IP flood | C&C server |
| Total Count | 100 | 100 | 100 | 100 | 400 |
| Found | 77 | 66 | 58 | 83 | 260 |

Table1: Results of clustering analysis of network traffic。

# Discussions

From the results of the experiment, the clustering analysis if effective for UDP flood, SYN flood and GRE IP flood but it does not really effective for TCP stomp flood. As the clustering analysis is to find the difference of the network traffic, it is difficult to find out the network traffic of botnet if it is similar to the normal network traffic. As the network traffic of GRE IP flood is similar to the network traffic of browsers, it is difficult to use clustering analysis to detect it. The C&C needs to connect to lots of bots and it is really different to the behavior of bots. So the accuracy of detecting is closely to the rate of detecting the abnormal traffic.

# Conclusions

The network traffic of botnets can be effectively found using clustering analysis. We can extract the network traffic of botnet if it is different with normal network traffic. After extracting the network traffic, it is likely to find out the C&C server using clustering analysis as the network traffic is different from bots’. But we also need to notice that the network of C&C servers is rare and maybe we can’t find out the C&C server.

# Acknowledgements

This research was supported by Participation in Research Program funds granted by Shanghai Jiaotong University. We would like to thank Yin Jin for her thoughtful comments. We also thank anonymous for providing the mirai botnet source code. Transcripts of the materials can be obtained by writing to either of the authors.

说明论文获得的基金的资助，论文获得的相关帮助等信息。

##### References

1. Ming Wang, Qian Zhang, The constrction of E-business: a case study. Journal of Peiking University, Vol 243, pp.102-103, April 2003 (In Chinese).
2. G. Eason, B. Noble, and I. N. Sneddon, “On certain integrals of Lipschitz-Hankel type involving products of Bessel functions,” Phil. Trans. Roy. Soc. London, vol. A247, pp. 529–551, April 1955. *(references)*
3. J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
4. I. S. Jacobs and C. P. Bean, “Fine particles, thin films and exchange anisotropy,” in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
5. K. Elissa, “Title of paper if known,” unpublished.
6. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
7. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
8. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.