Application of Revised Ant Colony Optimization for Anomaly Detection Systems

Chia-Mei Chen[1*], Wen-Ling Lo[2], Ya-Hui Ou[3], Gu-Hsin Lai[3], Tse-Yao Wang[3]

[1] Department of Information Management National Sun Yat-Sen University, Kaohsiung, Taiwan.
[2] Department of Technology Crime Investigation Taiwan Police College, Taipei, Taiwan.
[3] Department of Aviation & Communication Electronic Air Force Institute of Technology, Kaohsiung, Taiwan.

* Corresponding author. Tel.: +886-7-5252000#4726; email: chiamei.chen@gmail.com
Manuscript submitted March 28, 2018; accepted June 20, 2018.
doi: 10.17706/ijcee.2018.10.3.241-247

Abstract: Botnets have caused significant security threat and huge loss to organizations. Bot masters control the botnet through command and control servers (C2 servers); they often adopt the most commonly used communication channel, HTTP, in order to blend in malicious communication messages into massive normal traffic for detection evasion purpose. By analyzing malicious and normal traffic, this study discovered the anomalies of the botnet communication patterns. Botnet connections exhibit some similarity behaviors which are not possessed by normal traffic. This study develops an anomaly score function to represent the anomalies and proposes a botnet detection method based on a revised ant colony optimization algorithm. The experimental results show that the proposed anomaly botnet detection method identifies botnets efficiently.

Key words: Botnet, anomaly detection, ant colony optimization.

1. Introduction

The advances in Internet technology and the rapid growth of the Internet and mobile network provide ubiquitous computing. The Internet plays an essential role in our daily activities and has also brought a relative surge in cyber-attacks. Businesses have been exposed varieties of security threats. According to the incident reports and security reports [1], botnets have become one of the most powerful attack tools for hackers.

A botnet is constructed by a number of compromised machines through infection; the compromised machines are called robots, bots, or zombies. Malware infections could be done easily through social engineering attacks; victim users browse a malicious email attachment or website and get infected. Once victims are infected, they report to the C2 servers and are ready for attacking. Bots are remotely controlled by a botmaster; they receive instructions through a communication channel to the command and control (C&C or C2) servers.

Bots could be a combination of various forms of malware, such as Trojan horse, virus, worm, and spyware. Hackers often modify bot program code to quickly develop a new botnet attack, creating a serious threat that is difficult to prevent. Botnet size may vary ranging from thousands to hundred millions of bots and the power of a botnet is proportional to the size [2], [3].

For the purpose of survivability and intrusion evasion, bot masters construct a botnet involving multiple C&C servers [4], [5] and zombies would report and receive commands from one of the servers. Therefore,

command and control servers are the heart of botnets; taking down the servers means disrupting the attacking power of botnets.

Despite the wide deployment of many defense mechanisms, such as firewall, anti-virus software, and intrusion detection system, cyber attacks are still increasing [6]. Traditional firewalls, anti-virus software, and IDS (intrusion detection systems) seem ineffective against botnet attacks. Intrusion detection systems (IDSs) are an essential security defense component. Therefore, the purpose of this study is to propose an efficient anomaly-based botnet detection system.

The structure of the paper is as follows: Section 2 reviews anomaly-based botnet detection, Section 3 presents the proposed botnet detection method, Section 4 discusses the experimental results and Section 5 concludes the paper.

2. Related Work

A botnet detection system BotGAD (Botnet Group Activity Detector) [7] was developed based on the group activities, such as group uniformity, activity periodicity, and activity intensity. Akiyama *et al*. [8] proposed three metrics for determining the botnet behaviors: relationship, response, and synchronization. The relationship presents the connection between bot master and bots over one protocol, such as IRC, HTTP, or P2P. The response means that bots respond immediately and accurately after they receive commands from the botmaster. The synchronization means bots simultaneously carry out programmed activities, such as DDoS attack, reporting their status, or sharing information, based on the botmaster's commands.

Lakhina *et al*. [9] adopted sample entropy to find the traffic flow distribution characteristics. The work could detect DDoS and port scanning during the progress of the attacks, but it is not suitable for identifying botnets. Yen and Reiter [10] proposed a detection system called TAMD to identify infected hosts in the enterprise network by finding out aggregated communication involving multiple internal hosts. The features include flows communicating with the same external network, sharing similar payload, and involving internal hosts with similar software platforms. The experimental results show that the proposed approach has a low false positive rate.

Chen and Lai [11] applied ACO to identify botnet connections with control and command servers. Some normal connections such as regular updates are unable to distinguish by the above ACO-based approaches, as they exhibit similar connection behaviors as bots.

Some studies employed data mining technique as countermeasure for botnets. Livadas *et al*. [12] and Strayer *et al*. [13] apply machine learning algorithms, like C4.5 Decision Tree, Naïve Bayes, and Bayesian Networks, to analyze IRC-based botnet. The network traffic is classified into two groups: IRC and non-IRC. The Euclidean distance is calculated to correlate similar IRC traffic together. Kondo and Sato [14] adopted support vector machine (SVM) algorithm to identify C&C sessions from the traffic data. Their work finds out the packet histogram vector of the C&C session, including packet payload size and packet interval time, can better identify C&C sessions than the other vector definitions, such as session information vector and packet sequence vector. Lu *et al*. [15] employed n-gram, decision tree and clustering algorithms to classify network traffic into different application communities. Lu's work analyzes the temporal-frequent characteristics of the 256 ASCII bytes on the payload over a predefined time interval to distinguish malicious bot traffic from normal one. Huang [16] proposed a bot detection mechanism which analyzed failure packets by means of machine learning approach.

3. Proposed Detection Approach

Zombies maintain a connection with their C2 server. Based on our observation, botnets behave differently from the normal users and various botnets exhibit different anomaly connection behaviors. Some advanced

botnets might apply a certain degree of randomness to invade detection. As they need to ensure a connection is maintained with the C2 server, the randomness might be able to identify. More advanced botnets might adopt some randomness and multiple connection frequencies.

DBSCAN (Density-Based Spatial Clustering and Application with Noise) is a density-based clustering algorithm, introduced in Ester *et al*. 1996 [17], which can be used to identify clusters in a data set containing noise and outliers. Clusters are dense regions in the data space, separated by regions of lower density of points. Due to network latency, the connection frequency might generate noise or error, but DBSCAN can overcome such noisy data. Therefore, the proposed solution applies DBSCAN to cluster the connection behaviors to identify the multi-frequency regularity of botnet connections.

According to the literature review, ant colony optimization (ACO) and data mining algorithms provide promising results on identifying the anomalies, but each has some limitations. ACO could identify anomalies with great similarity with high false positive rate on regular update messages. Data mining requires training and the detection performance heavily relies on the training set. By utilizing the clustering algorithm, the connections can be characterized in groups. Two anomaly scores are defined to express the anomalies of group behaviors found in botnet connections. To enhance detection performance, this study develops a revised ACO-based botnet detection method.

To analyze the abnormal network flows, all the HTTP network traffic is collected. The first part of the proposed method adopts DBSCAN to cluster the connections of a given pair of source and destination in terms of the inter-connection times and packet sizes. The anomaly scores of each pair of connection are calculated as the basis of the pheromone function used in the revised ACO algorithm. The second part applies ACO to identify the connection patterns of a given pair of source and destination. Different from the previous work, this study extends the observation time in order to discover low frequency stealthy connection and proposes an adaptive threshold function to identify malicious connections.

An initial work of ACO algorithm [18] solved the TSP problem which has three variation algorithms: ant-density, ant-quantity, and ant-cycle. An isolated ant moves randomly. It decides to follow a trail with high pheromone trail and reinforces the trail by laying its own pheromone. The collective behavior emerging from ants forms an autocatalytic reaction where the more the ants follow a trail, the more attractive the trail becomes. In the literature, the change of the pheromone attempts to search for short edge and then induce a positive feedback. In this revised ACO algorithm, the idea of finding shortest paths is transformed into identifying anomaly connection behaviors; the change of the pheromone is based on the anomaly scores of the connection behaviors, not the edge distance in TSP.

Advanced botnets may connect the servers in various frequencies or random. The revised ACO algorithm utilizes such anomalies to define the pheromone function and to discover the malicious botnet connections. At an iteration of the revised ACO algorithm, ants explore the connections in the network during a period of time. For each path (the connections of a given pair of source to destination), the amount of pheromone generated is based on the two attributes: inter-connection period and packet size. Hence, the anomaly score function is defined by the two anomalous behaviors: the regularity of the inter-connection times and the regularity of the flow sizes, which will be explained below.

The inter-connection times of the connections of a given pair of a connection explored during a time frame are clustered into groups. The connections with regular inter-connection times have a larger score than those with random connection times. The anomaly score of the inter-connection times indicates the degree of the regularity of the inter-connection periods. The packet sizes of the botnet messages might be similar; the proposed method clusters the sizes into groups. The anomaly score of the packet sizes represents the degree of randomness.

The above two anomaly behaviors are applied to the revised ant colony optimization algorithm as the

basis of the pheromone; the anomaly connections will accumulate an amount of pheromone.

The proposed detection algorithm combines ACO and DBSCAN to identify various types of anomalous botnet connection behaviors, including botnets with one or multiple connection frequencies and those with a mixture of connection regularity and randomness. This study utilizes the clustering algorithm DBSCAN to discover if the connections of a pair of source to destination exhibit the group behaviors. The cluster results are used to define anomaly scores of the connections; the anomaly scores indicate how anomaly the connections are.

In Traveling Salesperson Problem, the heuristic information is inversely proportional to the distance; in this study, the heuristic information indicates the degree of traffic anomalies observed on a path (representing the connections of a pair of source to destination in a given time frame). Therefore, the path exhibiting anomaly connection behaviors has high pheromone and the chance that ants explore it would increase. If the same path continues performing the anomalous connections, the ACO algorithm would form a positive feedback and finally most ants would explore the same path.

The suspicious botnet servers are the ones with high pheromone, while normal ones have lower. Therefore, given the list of servers (destination IPs) found by ranking in the order of pheromone, the cut-off point of the list of the suspicious is the maximum gradient descent. An illustration will be shown in the next section.

4. Performance Evaluation

To evaluation the detection performance of the proposed detection system under various network environments, this study used traffic datasets from real networks as well as research labs [19-21]. The average detection performance of the datasets is summarized in Figure 1. The results demonstrate that the proposed detection method captures malicious connections efficiently and has a very low false positive rate.

The results were further investigated manually for identifying the causes of the false positives. One false positive occurred as the connections of the IP are in the ending point of the dataset; the other was Plurk connections. The former false positive could be avoided in real environments as the traffic comes continuously. Plurk is a social networking service, which connects its server every 2 seconds, very similar to the botnets' behavior. The previous work [22] excluded such traffic without analysis, while this study considers it as a false positive.

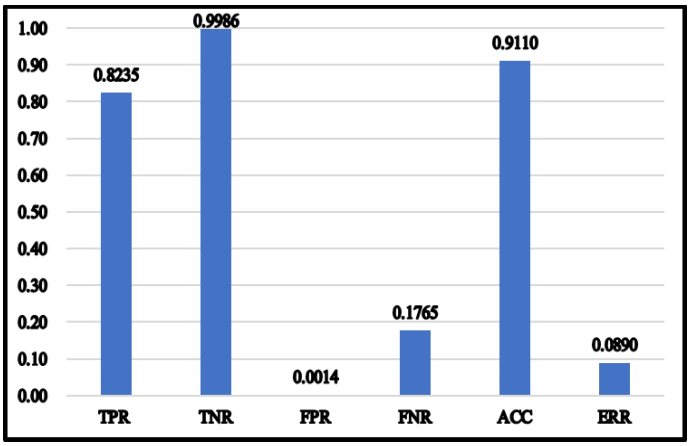

Fig. 1. The average detection performance.

5. Conclusion

Many cyber attacks utilize botnets to launch attacks. Detecting botnets could reduce the damage. Botnets

become stealthy to evade rule-based intrusion detection system. A small amount of malicious traffic is generated and mixed into a massive amount of normal traffic. The previous works identify botnets based on their regular connection behaviors, while some botnets connects to the C2 servers with multiple connection frequencies or randomness to evade detection.

By analyzing malicious and normal traffic, this study discovered the anomalies of the botnet communication patterns. Botnet connections exhibit some similarity behaviors which are not possessed by normal traffic. This study develops an anomaly score function to represent the anomalies and proposes a botnet detection method based on a revised ant colony optimization algorithm. The experimental results show that the proposed anomaly botnet detection method identifies botnets efficiently.

The proposed solution was evaluated using datasets. More evaluations can be done using real botnet traffic collected from a large real network. Further investigation can be done by extending to peer-to-peer botnets.

Acknowledgment

We would like to thank the research grant support from MOST (Ministry of Science and Technology) and TWISC (Taiwan Information Science Center), Taiwan.

References

[1] Li, C., Jiang, W., & Zou, X. (2009). Botnet: Survey and case study. *Proceedings of the 4th International Conference on Innovative Computing, Information and Control (ICICIC)* (pp. 1184-1187).

[2] Cai, T., & Zou, F. (2012) Detecting HTTP botnet with clustering network traffic. *Proceedings of the 8th International Conference on Wireless Communications, Networking and Mobile Computing (WiCOM)* (pp. 1-7).

[3] Crotti, M., Dusi, M., Gringoli, F., & Salgarelli, L. (2007). Detecting http tunnels with statistical mechanisms. *Proceedings of the IEEE International Conference on ICC'07* (pp. 6162-6168).

[4] Al-Bataineh, A., & White, G. (2012). Analysis and detection of malicious data exfiltration in web traffic. *Proceedings of the 7th International Conference on Malicious and Unwanted Software (MALWARE)* (pp. 26-31.)

[5] Zeidanloo, H. R., & Azizah, A. M. (2009). Botnet command and control mechanisms. *Proceedings of the Second International Conference on Computer and Electrical Engineering* (pp. 564-568).

[6] Cybersecurity, Cybint news. (2017, September). The scary truth about cyber security. Retrieved from the website: https://www.cybintsolutions.com/cyber-security-facts-stats/

[7] Choi, H., Lee, H., & Kim, H. (2009). BotGAD: Detecting botnets by capturing group activities in network traffic. *Proceedings of the Fourth International ICST Conference on Communication System Software and Middleware*.

[8] Akiyama, M., Kawamoto, T., Shimamura, M., Yokoyama, T., Kadobayashi, Y., & Yamaguchi, S. (2007). A proposal of metrics for botnet detection based on its cooperative behavior. *Proceedings of International Symposium on Applications and the Internet Workshops (SAINT)*.

[9] Lakhina, A., Crovella, M., & Diot, C. (2005). Mining anomalies using traffic feature distributions. *Proceedings of the ACM SIGCOMM* (pp. 217-228).

[10] Yen, T. F., & Reiter, M. K. (2008). Traffic aggregation for malware detection. *Lecture Notes in Computer Science 2008*, 207-227.

[11] Chen, C. M., & Lai, G. H. (2017). Ant-based botnet C&C server traceback. *Proceedings of the International Conference on NCS*.

[12] Livadas, C., Walsh, R., Lapsley, D. E., & Strayer, W. T. (2006). Using machine learning techniques to

identify botnet traffic. *Proceedings of the 31st IEEE Conference on Local Computer Networks* (pp. 967-974).

[13] Strayer, W. T., Walsh, R., Livadas, C., & Lapsley, D. E. (2006). Detecting botnets with tight command and control. *Proceedings of the 31st IEEE Conference on Local Computer Networks* (pp. 95-202).

[14] Kondo, S., & Sato, N. (2007). Botnet traffic detection techniques by C&C session classification using SVM. *Proceedings of the Security 2nd International Conference on Advances in Information and Computer Security (IWSEC)* (pp. 91-104).

[15] Lu, W., Rammidi, G., & Ghorbani, A. A. (2011). Clustering botnet communication traffic based on n-gram feature selection. *Proceedings of the 2011Computer Communications* (pp. 502-514).

[16] Huang, C. Y. (2013). Effective bot host detection based on network failure models. *Computer Networks, 57(2)*, 514-525.

[17] Ester, M., Kriegel, H. P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. *Proceedings of the 2nd Int. Conf. on Knowledge Discovery and Data Mining* (pp. 226-231).

[18] Brezina, Jr. I., & Čičková, Z. (2011). Solving the travelling salesman problem using the ant colony optimization. *Management Information Systems, 6*, 10-14.

[19] TACERT. (2017). Taiwan academic network computer emergency response team case studies. Retrieved from the website: http://tacert.tanet.edu.tw/prog/index.php

[20] ISOT Research Lab. Datasets. (2017, August). Retrieved from the website: https://www.uvic.ca/engineering/ece/isot/datasets/

[21] Mila, collection of PCAP files from malware analysis. (2017). *CONTAGIO*. Retrieved from the website: http://contagiodump.blogspot.tw/2013/04/collection-of-pcap-files-from-malware.htm

[22] Lee, J. S., Jeong, H., Park, J. H., Kim, M., & Noh, B. N. (2008). The activity analysis of malicious http-based botnets using degree of periodic repeatability. *Proceedings of the SECTECH'08, International Conference on Security Technology* (pp. 83-86).

Chia-Mei Chen has joined in the Department of Information Management, National Sun Yat-Sen University since 1996. She was section chief of Network Division and Deputy Director, Office of Library and Information Services in 2009-2011. She had served as a coordinator of TWCERT/CC (Taiwan Computer Emergency Response Team/Coordination Center) during 1998 to 2013 and established TACERT (Taiwan Academic Network Computer Emergency Response Team) in 2009. She is a deputy chair of TWISC@NCKU, a branch of Taiwan Information Security Center. She continues working for the network security society. Her current research interests include anomaly detection, malware analysis, network security, and cloud computing.

Wen-Ling Lo has earned a master's degree after graduating from the department of information management at National Sun Yat-Sen University of Taiwan in 2015. She is the analyst and team leader of Research and Development for Taiwan Computer Emergency Response Team / Coordination Center (TWCERT/CC). Her job is to lead the team to develop systems, including incident report system, ticketing system, malware analysis and report system and so on. She is also responsible for the international collaboration with international CERTs and related organization.

Ya-Hui Ou received the M.S. degree in Department of Information Management from the I-SHOU University, Taiwan, in 2009, the Ph.D. degree in Department of Information Management National Sun Yat-Sen University, Taiwan, in 2017. She is currently working as research assistant at National Sun Yat-Sen University. Her research interests include network security and statistical analysis.

Gu-Hsin Lai received the M.S. degree in information management from the National Chi Nan University, Taiwan, in 2002, the Ph.D. degrees in information management from National Sun Yat-Sen University, Taiwan, in 2009. He was a principal engineer of information technology security office in Taiwan Semiconductor Manufacturing Company from 2009 to 2011. He joined the Faculty of the Department of Information Management, Chinese Culture University, in 2012. Now, he is an assistant professor in Department of Technology Crime Investigation, Taiwan Police College. His research focus on spam mail filter, wireless sensor network and system security.

Tse-Yao Wang is an assistant professor at the Faculty of Department of Aviation & Communication Electronic at Air Force Institute of Technology, Kaohsiung, Taiwan. He has received the Ph.D. degree in Department of Information Management National Sun Yat-Sen University, Taiwan, in 2016. His research interests include network security and machine learning.