Grouping Domain Names using DNS Query Graph

Jehyun Lee[1], Jonghoon Kwon[1], Hyo-Jeong Shin[2], and Heejo Lee[1]
Division of Computer and Communication Engineering, Korea University[1], Korea Telecom[2]
Seoul, Korea
[{arondit, signalnine, heejo}@korea.ac.kr[1], hshin@kt.com[2]]

Abstract

Many malwares have the network activities for command and control, update, propagation, and so on. Because of the use of domain names while the network activities, those malicious activities are observed and have been blocked on the DNS. However, to evade static blacklists, recent malwares are using numbers of newly generated domain names. In this paper, we introduce a domain name grouping using DNS query graph containing query strategy of DNS clients. By grouping the domain names which have the statistically and sequentially similar query strategy, we extract the malicious domain names groups of the multi-domain malwares from the numerous numbers of domain names. From the experiments with the DNS trace of an ISP network, we find tens of multi-domain malwares, and commonly observed unique DNS query strategies. As the contribution of method, the grouping result enhances the efficiency of blacklists by detecting newly appeared malicious domain names.

Keywords: Malicious domain names, Domain name grouping, Query graph, Malware

1. Introduction

Malicious software, called malware, has been considered as a main source of Internet threats, such as spying, DDoS attack, spamming, and etc. As the more computers are connected to a network, the more malwares works using the network for their malicious activities.

Against the malicious network activities, one of the monitoring points is the Domain Name System(DNS), and firewalls and IDSes have been takes blacklists to qurantine the malicious network activities. However, to evade the blacklist based approach, recent malwares including botnets, such as Conficker, Torpig, Rustock, Tykib and so on, are using numbers of malicious domain names.

In this paper, we propose a domain grouping method which enables to find newly generated malicious domain names. By grouping the domain names which show the statistically and sequentially similar query strategies using a graph based approach [1], the mechanism makes the domain name groups of malwares, legitimate servicies, and related domains. Compared with previous studies [2], [3], this method considered the similarity of query strategy caused by the shared clients.

2. Domain Grouping Mechanism

Our domain name grouping mechanism is based on the DNS query graph generated from DNS traffic. For extracting groups of domain names,

This research was supported by the MKE(Ministry of Knowledge Economy), Korea, under the ITRC(Information Technology Research Center) support program supervised by the NIPA(National IT Industry Promotion Agency)" (NIPA-2010-C1090-1031-0005), and the R&BD Support Center of Seoul Development Institute and the South Korean government (Project title: WR080951, Establishment of Bell Labs in Seoul / Research of Broadband Convergent Networks and their Enabling Technologies).

the mechanism finds sub graphs which have domain names with similar strategy as their nodes. The sub graph extracting method is the graph filtering with increasing threshold. The graph properties the mechanism use for grouping similart query strategy are the number of clients, the number of queries, and the average query amount per a client. According to the edge properties in the query graph, these three properties represent commonly appeared query strategies, in terms of query statistics and sequence. The domain names sharing similar clients are connected relatively stronger, higher edge weight, in the query graph, than the domain names in the other groups.

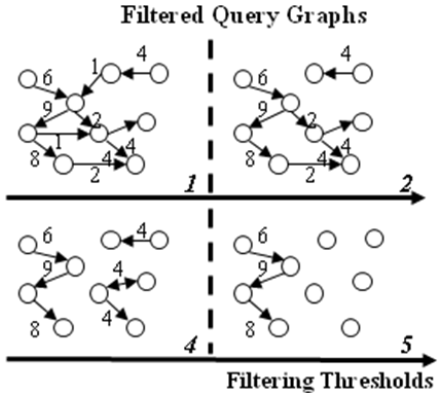

Fig. 1 An example of chromatograhic filtering

The increasing threshold graph filtering is removing the edges which have the less edge weight than the moving threshold. The filtering is applied with increasing thresholds and finds the separated components which have higher intra edge weight and less inter edge weight than a threhold. This characteristic of the query graph is illustrated in **Fig. 1**.

As the result, we get the component query graphs which have classifiable query strategy and infection status in each filtering step, and, compositly, a domain names in a component share the similar query strategy and the infection clients. We named this filtering method as the chromatographic filtering.

3. Experimental Result

In out experimentations, we attempt to evaluate how well the domain names which are generated by same malware are grouped without noise. Our experimentations are performed with DNS traffic captured at an ISP level DNS server.

From the two hours of DNS trace which have near ten million of queries and one million of domain names, the filtering method extract near a thousand of groups. Tens of them have known malicious domains as a member node. **Fig. 2** is some example of detected group of malicious domain names.

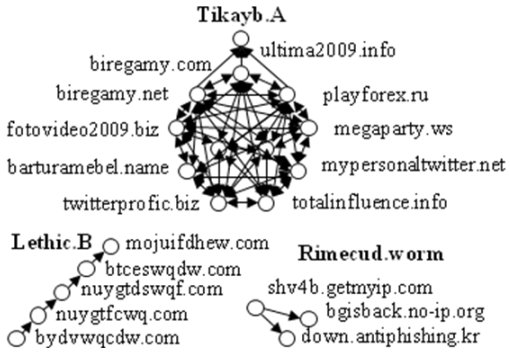

Fig. 2 An example of malicious domain groups

4. Conclusion

Recent malwares are using multiple domain names for evading blacklist based detection. In this paper, we propose a domain grouping method using DNS query graph. By using the chromatographic filtering to a query graph, we get the domain groups having similar query strategies. This grouping result contributes to enhance the efficiency of static black lists and white lists, and lastly to quarantine malwares.

References

[1] J. Lee, J. Kwon, H. Shin, and H. Lee, "Tracking Multiple C&C Botnets by Analyzing DNS Traffic," in *Proc. of 6th Workshop on Secure Network Protocols*, pp.67-72, 2010.

[2] M. Antonakakis, R. Perdisci, D. Dagon, W. Lee, and N. Feamster, "Building a Dynamic Reputation System for DNS," *in Proc. of the 19th USENIX Security Symposium*, pp.273-289, 2010.

[3] N. Shishir, M. Prateek, H. Chi-Yao, C. Matthew, and B. Nikita, "Bot-Grep: Finding P2P bots with structured graph analysis," *in Proc. of the 19th USENIX Security Symposium*, pp. 95-110, 2010.