The Rapid Extraction of Suspicious Traffic from Passive DNS

Wenbo Wang[1], Tianning Zang[2], Yuqing Lan[1]

[2]*National Internet Emergency Center,*

[1]Beihang University,

ICISSP 2018

4th International Conference on Information Systems Security and Privacy

ABSTRACT

The network traffic is filled with numerous malicious requests, most of which is generated by amplified attacks, random subdomain name attacks and botnets. Through using DNS traffic for malicious behavior analysis, we often need to test each domain alone. Besides, the amount of data is very large and simple filtering cannot quickly reduce the need to detect the number of domain names. As a result, it takes a lot of time to calculate on the premise of limited resources. Therefore, this paper introduces a extraction scheme for DNS traffic. We designed a simple and efficient method for extracting three kinds of attack traffic with the largest proportion of traffic. Besides, the method of statistics and classification was used to deal with all the traffic. We implemented a prototype system and evaluated it on real-world DNS traffic. In the meanwhile, as the recall rate reached almost 100%, the number of secondary domain names to be detected was reduced to 8% of the original quantity, and the DNS record to be detected was reduced to 1% of the original number.

OVERVIEW

We conduct a high-level overview of our system here. According to Figure 1, we divide the system into two parts, respectively, the Passive DNS pre-processing module and the traffic extraction module. Besides, we will describe each module function and discuss how to achieve the goal together as well as maximize the recall rate and efficiency.

The process of pre-processing mainly achieves three goals. One is to remove the DNS records containing the wrong domain name, that is, to clean data. For example, there are some illegal characters like *'!'* and *'_'* appeared in domain names. The second is to eliminate unrelated traffic, which will not appear in malicious traffic \, such as traffic of reverse DNS. On the other, we will try to remove domain names that do not appear in malicious traffic during a certain type of traffic extraction process. For example, the domain name in the whitelist can be removed during the extraction of DGA traffic. The third is to calculate the eigenvalues for the traffic extraction module, such as the proportion of ANY queries for each domain name.

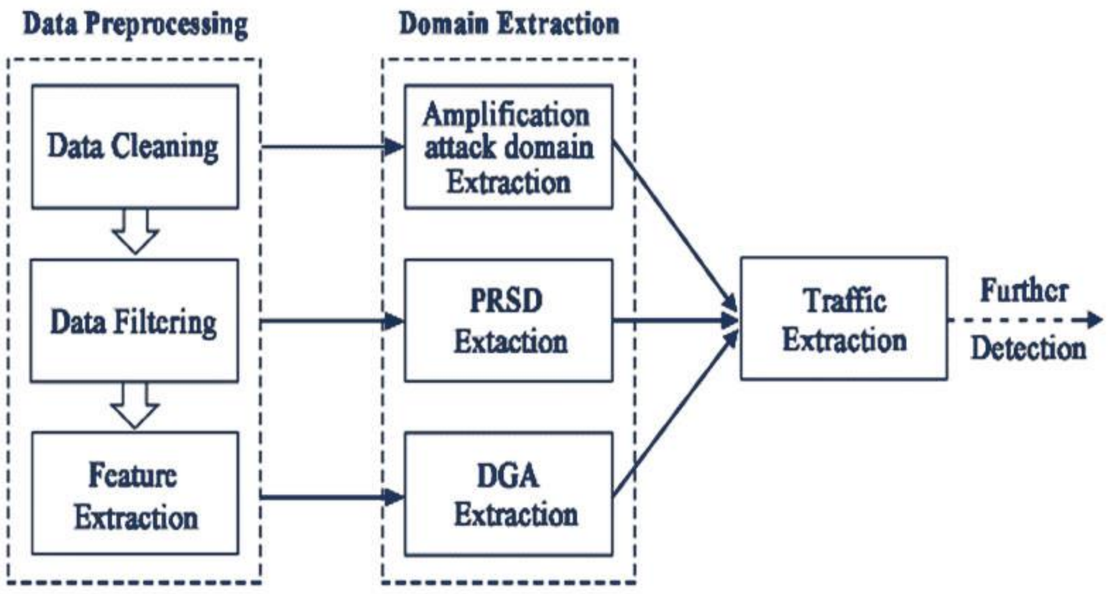

Figure 1: System Overview.

DATA PREPROCESSING

- **Data Cleaning**

A valid domain name contains only 26 alphanumeric characters (including uppercase and lowercase), numbers, dashes, and points used to split each segment.

- **Data Filtering**

We filter the reverse domain name and the domain name generated by the configuration errors.

- **Feature Extraction**

We count the number of queries group by each SLD within a period as *qc*, the number of queries which type is ANY as *qac* and the number of queries which type is TXT as *qtc*. We set any query ratio *qar* = *qac* / *qc*, txt record query ratio *qtr* = *qtc* / *qc*. We calculate the proportion of non-existent domain name *nxdr*. We count the number of sub-domain names for each SLD in the time interval, named *sdc*. We also get entropy, bigram, trigram, fourgram, length of SLD.

DOMAIN EXTRACTION

We take advantage of the SLD and its features, and extract the domain name which involves malicious behaviours. The extraction process aims at three types of attacks, namely, amplification attacks, random subdomain attacks, and DGA domains.

By formula 1, we set a parameter β. When *qar*+*qtr*<=β, the result is 0. When *qar*+*qtr*>β, the sum of *qar* and *qtr* was positively correlated with S_1. At the same time, we set threshold α, in which S_1 >α, and identified as the suspected amplification attack of traffic.

$$S_1 = \max\{0, 1 - e^{-\frac{qar+qtr-\beta}{\beta}}\} \qquad (1)$$

The attackers generate numerous sub domains randomly under the SLD, and these domain names do not exist. Therefore, we multiply *sdc* with *nxdr* to represent the possibility of malicious use, and this value range is large. We use formula 2 to change the result to between 0 and 1. And we have done a lot of experiments to determine the most appropriate parameters in this two formulas.

$$S_2 = \frac{e^{\theta(sdc*nxdr)} - 1}{e^{\theta(sdc*nxdr)} + 1} \qquad (2)$$

ADGs means the domain name generated by the DGA algorithm. The domain name generated by DGA is second-level domain, so this part of the extraction is concerned with the SLD in the traffic. Our model is trained with black and white lists. We extracted the SLD of each domain name in the list, and calculated their length, entropy and n-gram respectively, where n=2,3,4. We chose Random Forests as classifiers. When each tree is trained, a data set of the same size as N can be trained (bootstrap sampling) from a full training sample (sample number N). Random Forests are trained at a faster rate, which can balance errors.

We get the domain name from the previous module. For each SLD, we first remove it if it is in the white list, and then classify the remaining domains into categories in turn to obtain the suspected ADGs.

Further detection often requires complete DNS resource record. Therefore, after obtaining the domain name, the module will restore this set of secondary domain names to the corresponding traffic, that is, the original DNS resource record. The operation is very simple. This batch of secondary domain name after cleaning is listed. Then, the records are retained which have the same SLDs.

RESULTS

We use the data from Shanxi Province to experiment with the relevant parameters, and the results shown in Figure 2. In addition, we have also experimented with the data in Guangdong using the same parameters, which can be found in Figure 3. Besides, Figure 2(a) shows the trend of the number of domain names involved in subdomain attacks and amplification attacks. The peak value is about twice that of the valley value. From Figure 2(b), the change trend is not the same as the change of the number of domain names. The peak value is about 7 times that of the valley, which is more likely to reflect an attack of amplification attacks and random subdomain attacks. Figure 2(c) shows the trend of the number of DGA domain extraction. Since we only judge by the character of the domain name, the extracted domain name contains numerous legitimate domain names. Therefore, the number of DNS records does not really reflect the size of the actual attack traffic. Moreover, the recall rate of our DGA domain is 92% here. In using the results of the experiment on the data in Guangdong, the recall rate of the amplified attack and the random subdomain attack was also 100%, but the recall rate to the DGA was close to 90%.

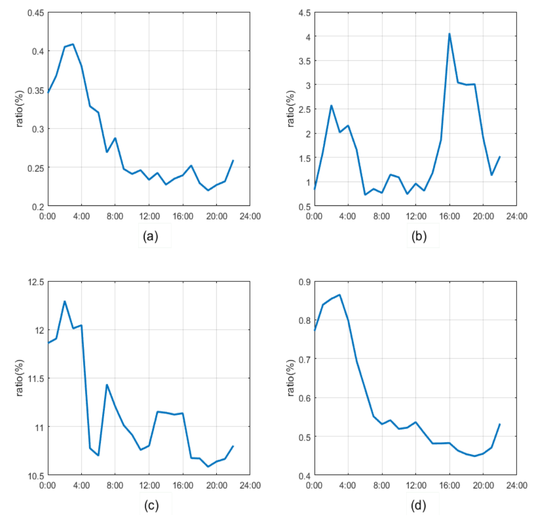

Figure 2: The results of experiments using 23 hours of data from Shanxi Province. (a) (b) are the ratio of number of SLDs and DNS records extracted from PRSD and zoom attacks in an interval. (c) (d) are the ratio of number of SLDs and DNS records extracted from DGA in an interval.

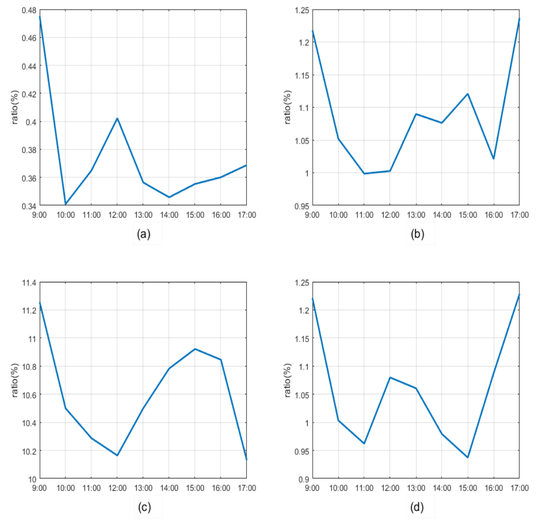

Figure 3: The results of experiments using 23 hours of data from Guangdong Province.

CONCLUSIONS

we propose a malicious traffic extraction model system. The system refers to the relevant detection system, selects features, and preliminarily filters the domain names and traffic related to some attacks. Compared with the simple pre-processing process, the system can select the malicious traffic to a smaller range, while ensuring the recall rate. However, it is also very fast to achieve their goals, which is to narrow the range for further detection. Our evaluation uses real data from passive DNS data of provincial telecommunication at different times. Amplification attacks and random sub-domain name attacks involved in the domain name recall rate reached 100%, DGA domain name recall rate of 90% or more.