Real-time FFSN Detection Method Based on Hurst Exponent

Jiajia Wang1*, Yu Chen1

1 School of Information Technology, Taizhou Polytechnic College, Taizhou, Jiangsu, 225300, China

*Corresponding author’s e-mail: wangjiajia_99@163.com

Abstract. In this paper, a real-time detection method of FFSN is designed. The response delay time of domain name is extracted by passive method. According to the self-similarity of response delay time of legitimate domain name and the difference of response delay time between legitimate domain name and FFSN, VTP analysis method based on Hurst exponent is used to detect FFSN. In order to avoid false alarm caused by the network itself, the suspected network also needs to be used the active method to judge the fast flux attribute in order to confirm the existence of FFSN. Our method can not only detect the existence of FFSN in real time, but also give the specific FFSN domain name. The experimental results show the effectiveness of our method.

1. Introduction

Botnet is made up of a large number of hosts that are installed with bots. Bots are remote control software developed by malicious users. These programs are parasitic programs that can be turned on automatically with the startup of computers. As a result, malicious users can manipulate a wide range of hosts in the Internet by controlling bots. Malicious users always try to control as many hosts as possible. Using botnet, malicious controllers could launch a variety of malicious acts, such as DDOS attacks, spam, phishing sites, click fraud and so on. Botnet, the biggest threat to the Internet, costs no less than a million dollars a year. In order to avoid exposure, the traditional botnet uses fast-flux service technology to hide the location of malicious server through the fast transformation of mapping relationship of domain name and IP address. Fast Flux Service Network (FFSN) increases the robustness and concealment of botnet.

In this paper, we propose a method that can detect FFSN in real time. By using passive and active two-step detection, our method can accurately judge the existence of FFSN and have little influence on network performance. Section 2 describes the existing work. Section 3 analyzes the essential features of FFSN. According to these features, the detection methods are proposed in Section 4. Section 5 illustrates that this paper adopts the VTP analysis method based on Hurst exponent and gives the algorithm flow and sequence construction methods. Section 6 is the experimental results and analysis.

2. Related Work

Please In view of the huge harm of FFSN, there are many detection methods at present.

Hsu[1] proposed a real-time FFSN detection method. According to the request agent mechanism of FFSN, the three-stage communication delay time between client and server is calculated. However, the calculation of the delay time of the three stages is difficult to be completed in the terminal, it needs to be deployed on routers, firewalls and other network devices, which undoubtedly increases the burden
of equipments. Moreover, the network delay is strictly related to the network performance of the local area network. The network with large delay is not necessarily FFSN, may be network congestion.

The existing FFSN detection methods are mainly based on the attributes of DNS packets, such as the mapping relationship between domain name and IP address\textsuperscript{[2,3]}, the size of DNS packets\textsuperscript{[4]} and so on. However, the number of DNS packets in the network is very large, which requires a large storage space. Passive methods need a long detection period, and each time the captured DNS packets must to be compared, it is also a great challenge to the performance of the device.

Wang\textsuperscript{[5]} proposed a FFSN detection method based on domain name system attributes. According to the essential attributes of FFSN domain name system, the detection features are extracted. However, the number of FFSN in the network is smaller than the legitimate network. If only the passive method is used to collect the characteristics of FFSN, it will take more time, such as a few months, to obtain the communication packets between users and FFSN. This problem exists in many FFSN detection methods based on domain name attributes\textsuperscript{[6, 7]}. We prefer a real-time detection method.

Tyagi\textsuperscript{[8]} proposed a spatial feature-based detection method that requires the computation of the number of Autonomous Systems (AS) corresponding to each domain name. A legal domain name may also be present in several AS, which may result in a higher error rate.

3. FFSN Feature Analysis
In this section, the characteristics of FFSN are analyzed, and then the detection methods are proposed according to these characteristics.

3.1 Fast Flux Attribute
The IP address of legal network domain name mapping is often limited. These IP addresses belong to high performance servers and can provide stable services. Therefore, the mapping relationship between domain name and IP is almost the same. Compared with traditional botnet, the biggest feature of FFSN is the use of fast flux technology. A FFSN domain name can map multiple IP addresses, these IP addresses are often neglected personal hosts, as long as the owners of these personal hosts turn off the computer or use the software to check and kill the virus, it is possible to disconnect from the FFSN controller at any time. However, as long as the number of bots is large enough, the domain name of FFSN can always be resolved to online IP address, even if a IP address is detected, the mapping relationship between this IP and FFSN would terminate immediately to ensure that FFSN is not exposed.

3.2 Request Agent Mechanism
Different from the working process of HTTP server in legitimate network, FFSN introduces the request agent mechanism. Once the client infects the bot and becomes a zombie host, the working process is shown in Figure 1.
The specific work process is as follows:
1) Zombie hosts send HTTP GET requests to FFSN proxy;
2) The proxy redirects the HTTP GET request to the FFSN server;
3) The server processes the request and sends the reply to the proxy;
4) The proxy returns the reply to the zombie hosts.

Compared with the legitimate network, the working process of FFSN has the two extra steps: the request agent of the proxy and the returns about the reply content. Therefore, from the user’s point of view, the response delay time of FFSN is longer than that of legitimate network, and even can not be accessed. As long as FFSN adopts request agent mechanism, this kind of delay is inevitable.

However, it is not possible to perform an accurate determination of the FFSN depending on the value of the delay time, such as the reduction of the network bandwidth, the insufficient processing capacity of the hardware, and the like, the fluctuation of the delay time may be caused. The congestion caused by the network burst data is more likely to cause false positives of the FFSN.

4. FFSN Detection Flow
We use a method combined of passivity and initiative to detect FFSN. Firstly, the passive detection method is used to analyze the delay time between the sending of HTTP GET packets and the receipt of the response nodes. This process does not involve query, only collects information at the terminal, so it would’t have any impact on the network. Influenced by the network itself, the value of delay time can not explain the attributes of the network (such as legitimate network or network manipulated by botmasters). In this paper, the value will not be calculated directly. Because the normal network packets have self-similarity (explained in Section 6.2), we calculated the Hurst exponent about the delay time. Once there is an abnormal, it is possible to detect a malicious network.

In order to avoid misinformation, we made a query on the fast flux attribute of the domain name by adopting the active method after detecting the suspected malicious network. A domain name may have multiple IP addresses, whether legal or illegal, but the difference is that the IP address corresponding to the domain name of the legal network belongs to the high-performance server, the corresponding relation between the domain name and the IP address is almost the same. Most of the IP addresses of the FFSN belong to the personal computers, and the corresponding relation between the FFSN domain name and the IP address is always changed due to the reasons such as shutdown and off-line. Since the number of FFSN in the network is very small, the active detection packets do not have a large impact on the network.

If the active test results are positive, it indicates that the current domain name has the property of fast flux, it is the FFSN, and should alarm to the administrators by the detection system. If the detection results are negative, it means that the situation of the current network may suddenly change (such as network failure, etc.). In addition to directly forwarding of the current data, the network administrator should check the local area network to ensure that the line is unobstructed.
5. Formatting the text VTP Analysis Method Based on Hurst Exponent

5.1 VTP Analysis Method
Currently, the effective Hurst exponent calculation method mainly includes three methods: the RS analysis method, the wavelet analysis method and the VTP analysis method. In which, the VTP analysis method is mainly used for analyzing the communication data and the actual network flow, and the difference between VTP analysis method and the other two methods is mainly that there is no need to divide a plurality of sub-sequences or sub-windows, and is more suitable for practical network application. In this paper, the Hurst exponent calculation is carried out by using the VTP analysis method, and the calculation steps are as follows:

1) Given a discrete time series X of length N, each w is divided into blocks, as follows: \(X^{(w)}_{j} = \frac{1}{w} \sum_{i=(j-1)w+1}^{jw} X_i\), in which, \(j = 1, 2, ..., \frac{N}{w}\), \(w = 1, 2, ..., \frac{N}{2}\);
2) Calculate sequence variance: \(\text{Var}[X^{(w)}] = \frac{1}{N/w} \sum_{j=1}^{N} \left( \frac{1}{w} \sum_{i=(j-1)w+1}^{jw} X_i - \bar{X} \right)^2\), in which, \(\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i\);
3) The curves are plotted on the abscissa and the ordinate of \(\log(w)\) and \(\log[\text{Var}(X^{(w)})]\), respectively, to obtain the slope \(\beta\): \(\log[\text{Var}(X^{(w)})] = \log[\text{Var}(X)] - \beta \log(w)\);
4) Finally, calculate Hurst exponent: \(H = 1 - \beta/2\).

5.2 Sequence Model
As can be seen from Section 5.1, using the VTP analysis method based on Hurst exponent, it is first necessary to construct a discrete time sequence X with length of N, in the detection flow of Section 4, we have collected the delay time of response data, the delay time is close to 0 if the network has no congestion and the host has access to the legitimate network. Once the bots have access to the FFSN, there is a large fluctuation in the delay time, and the similarity is destroyed, so that the value of Hurst exponent is changed.

Therefore, a delay time series based on the continuous time t is constructed: \(X_t = \{X_t, t = 0, 1, 2, \ldots\}\), \(X_t\) represents the sum of the delay times for all HTTP GET request packets collected by the system at time t.
When \( t = t_0 \), the sum of delay times is counted as \( X_{t_0} \), when \( t = t_1 \), the sum of delay times is counted as \( X_{t_1} \), and so forth. When \( t = t_c \), the detection sequence is \( X = \{X_{t_c-N}, X_{t_c-N+1}, \ldots, X_{t_c-1}, X_{t_c}\} \), \( c > N > 0 \). The next detection time is \( t = t_c + w \), the new detection sequence is \( X' = \{X_{t_c+w-N}, X_{t_c+w-N+1}, X_{t_c+w-1}, X_{t_c+w}\} \), \( 0 < w < N \).

With the continuous change of time \( t \), \( X_t \) is also constantly updated, and the sequence \( X \) could change with time. Section 6.2 shows that sequence \( X \) meets the analytical conditions of VTP method.

According to the conclusion of Ceballos\(^9\), with the increase number of calculations, the accuracy of the results will also increase. However, the longer the length \( N \) of the single detection sequence is, the longer the detection time would be, and the smaller the degree of polymerization \( w \) would be, the greater the computational complexity would be. In section 6, the appropriate parameters would be selected according to the training dataset.

6. Experimental Results and Analysis

6.1 Analysis of Response Delay Time in Training Data Set

We select training data from www.Alexa.com and www.malwaredomains.com, 100 legitimate domain names from www.Alexa.com and 100 FFSN domain names from www.malwaredomains.com. In order to ensure universality, each domain name is accessed multiple times at different time, and finally averaging access delay time is obtained.

![Figure 3. delay time corresponding to legitimate domain name](image)

![Figure 4. CDF function of legal Domain name response delay time](image)

![Figure 5. delay time corresponding to FFSN](image)

![Figure 6. CDF function of FFSN response delay time](image)

Figure 3 is the legal domain name response delay time, and Figure 4 is the corresponding CDF function. It can be seen that 96% of the legitimate domain name response delay time does not exceed 0.2s. Figure 5 shows the response delay time of the FFSN domain name, and Figure 6 is the corresponding CDF function. It can be seen that 90% of the FFSN domain name response delay time exceeds 0.3s. It can be seen that the response delay time of FFSN is different from that of legitimate domain name, and it is feasible to distinguish FFSN from legitimate domain name by delay time.
6.2 Self-similarity Analysis of Detection Sequences

The VTP analysis method is used to determine whether the legitimate domain name is mixed with FFSN, and the delay time series $X$ of the legitimate domain name needs to have the self-similarity of slow decline in the process of accumulation\textsuperscript{[10]}. This section analyzes the properties of sequence $X$ under normal network behavior.

We grasp data packets in local area network, in order to reflect the universality, we select five different time periods of a day to collect the data packets, and calculates the five sequences $A$, $B$, $C$, $D$, $E$, which meet the conditions of this paper, the time length of each sequence is $N=1000$ms, and the value of the degree of polymerization $w$ is 1, 2, ..., 10. The $\log(w)$—$\log(\text{Var}(X_w))$ diagrams of the five sequences are calculated, and the results are shown in Figure 7.

According to Figure 7, the slope and Hurst parameters of the sequence $A$, $B$, $C$, $D$, $E$ can be calculated separately, as shown in Table 1. According to Table 1, the slopes of sequences $A$, $B$, $C$, $D$ and $E$ are all $\gamma \in (-1, 0)$, and $\frac{1}{2} < H < 1 \left( H = 1 - \frac{\beta}{2}, \beta = -\gamma \right)$. Therefore, the delay time series of legal domain name in normal network environment has self-similarity, which accords with the condition of using VTP detection method.

### Table 1. The slope and Hurst exponent of legal domain name response delay time

| Sequence | Slope   | Hurst exponent |
|----------|---------|----------------|
| A        | -0.30346| 0.84827        |
| B        | -0.25875| 0.870625       |
| C        | -0.49496| 0.75252        |
| D        | -0.45434| 0.77283        |
| E        | -0.32846| 0.83577        |

6.3 Detection of FFSN by VTP Analysis

In this section, a proper amount of FFSN domain name access data is mixed in normal network data as the detection sequence. The time length $N$ of the detection sequence is 1000ms, the sampling interval is 100ms, and the degree of polymerization $w$ is 1, 2, ..., 10. The collected detection sequences include $A$, $B$, $C$, $D$, $E$, $F$, $G$, $H$, a total of 8. Among them, sequence $A$ ~ $E$ is composed of normal network data, FFSN domain name response data in sequence $F$ account for 10%, FFSN domain name response data in sequence $G$ account for 80%, and sequence $H$ is composed of normal network data.

The specific test results are shown in Figure 8. It can be seen that the slope of the curve of sequence $A$ to $E$ is normal. The sequence $F$’s slope is suddenly increased due to the influence of the time change of response delay time of FFSN, this is different from sequence $A$ ~ $E$. The slope of the sequence $G$ tends to be flat with the increase of the number of access times of the FFSN domain name, and finally, due to the disappearance of the FFSN domain name in the sequence $H$, the slope slightly changes.

### Table 2. The slope and Hurst exponent of mixed domain name response delay time

| Sequence | Slope   | Hurst exponent |
|----------|---------|----------------|
| A        | -0.23221| 0.883893       |
| B        | -0.13002| 0.93499        |
| C        | -0.17625| 0.911874       |
| D        | -0.55782| 0.72109        |
| E        | -0.32736| 0.836321       |
| F        | -1.06431| 0.467843       |
| G        | -0.45752| 0.771241       |
| H        | -0.49067| 0.754663       |
The slope and the Hurst exponent of the sequence A to H can be calculated according to the detection result of Figure 8, as shown in Table 2. As can be seen from Table 2, the method described herein is able to detect the FFSN on-line.

6.4 Determination of FFSN

In order to improve the accuracy and avoid false alarm caused by network congestion, and in order to further locate the FFSN domain name accurately, we also need to analyze the fast flux attribute of the suspected domain name. In the initial stage of Hurst exponent change, we select the domain name with the largest delay time in the sequence, and uses dig tool to mine the IP address of this domain name. Because the mapping relationship between the FFSN domain name and the IP address is usually 3~10 min, the time to detect the suspected domain name fast flux attribute is at T and T+600s, respectively.

(a) Mapping relationship of the domain name and IP address at time Ts         (b) Mapping relationship of the domain name and IP address at time T+600s
Figure 9. Mapping relationship of the domain name and IP address at time Ts and T+600s

If the coincidence rate of the domain name and the IP address is more than 50% in this process of both T and T+600s, the anomaly of the Hurst exponent is considered to be caused by the network itself, and the alarm is not carried out; if the coincidence rate of the domain name and the IP address is lower in this process of both T and T+600s, as shown in Figure 9, the anomaly of the Hurst exponent is considered to be due to the detection of a FFSN, and an alarm is required.

7. Summary and Prospect

In this paper, a method which can detect FFSN in real time is proposed. Based on the difference of response delay time between legitimate domain name and FFSN, the existence of suspected malicious network can be judged by detecting the response delay time of domain name without any query and no effect on network bandwidth. In order to reduce false positives and improve accuracy, the fast flux attribute detection of suspected malicious networks is carried out. The number of FFSN in the network is much less than the legitimate network, so the active detection packets will not affect the network traffic. In this paper, the existence of FFSN can not only be detected, but also the domain name information of FFSN can be determined accurately. In the future, the FFSN should be further accurately located and eliminated.

Acknowledgments

This work was financially supported by Taizhou Polytechnic College Boshuo Fund(TZYBS-17-6).

References

[1] Hsu C H , Huang C Y , Chen K T . Fast-flux bot detection in real time[C]// International Conference on Recent Advances in Intrusion Detection. Springer-Verlag, 2010,p:1-20.
[2] Yunfeng Xu, Yansheng Lu, Zhengbiao Guo. The availability of fast-flux service networks[C]//iCOST, 2011 International Conference on Selected Topics in Mobile and Wireless Networking, IEEE, Shanghai, China, 2011.10, 89-93.
[3] Wang H T , Mao C H , Wu K P , et al. Real-Time Fast-Flux Identification via Localized Spatial Geolocation Detection[C]//COMPSAC, 2012 IEEE 36th Annual Computer Software and Applications Conference. IEEE, Izmir, Turkey, 2012, 244-252.
[4] Knysz M, Hu X, Shin K G. Good guys vs. Bot Guise: Mimicry attacks against fast-flux detection systems[C]// INFCOM, 2011 proceedings IEEE INFCOM, IEEE, 2011, 1844-1852.
[5] Wang Jiajia et al. Fast-Flux Detection Method Based on Domain Name System Attribute[J], Journal of Nantong Vocational University, 33(3), 2019.09, P: 77-82.
[6] C. Zhou, K. Chen, X. Gong. Detection of Fast-Flux Domains based on Passive DNS analysis[J], Acta Scientiarum Naturalium Universitatis Pekinensis, 2016,52(3), p:396-402.

[7] Martinezbea S, Castilloperez S, Garciaalfaro J. Real-time malicious fast-flux detection using DNS and bot related features[C]/ PST, Eleventh International Conference on Privacy, Tarragona, Spain, IEEE, 2013:369-372.

[8] Tyagi A K, Aghila G. Detection of fast flux network based social bot using analysis based techniques[C]/ ICDSE, International Conference on Data Science & Engineering, Cochin, Kerala, India, IEEE, 2012, 23-26.

[9] Ceballos R F, Largo F F. On The Estimation of the Hurst Exponent Using Adjusted Rescaled Range Analysis, Detrended Fluctuation Analysis and Variance Time Plot: A Case of Exponential Distribution[J]. 2017, 3(8):424-434.

[10] Fen Y , et al. Real-time SYN Flooding Attacks Detection Method Based on Hurst Parameter[J]. Computer Science, 2008, 35(12):109-114.