Poster: An Automatic Multi-Step Attack Pattern Mining Approach for Massive WAF Alert Data

Yang Zhang[1,2], Tingwen Liu[1,2], Jinqiao Shi[1,2], Panpan Zhang[1,2], Haoliang Zhang[1,2], Jing Ya[1,2]
[1]Institute of Information Engineering, Chinese Academy of Sciences, Beijing, China
[2]National Engineering Laboratory for Information Security Technologies, Beijing, China
Email: {zhangyang, liutingwen, shijinqiao, zhangpanpan, zhanghaoliang, yajing}@iie.ac.cn

***Abstract*—This paper introduce a three-stage approach that can automatically mining multi-step attack patterns from massive alert data of web application firewalls. The first stage extracts attack sequences, and the second stage clusters similar attack sequences. At the last stage we recognize an attack pattern for each cluster. We conducted our experiments on real-world WAF alert data obtained from a famous Chinese ISP. Experimental results show that different attackers using the same attack pattern may have the same "attack background".**

I. Introduction

WAF (Web Application Firewall), as an appliance to identify common attacks, such as cross-site scripting and structured query language injection, is widely used in web application protection. WAF work depending on well-designed rules customized to the protected web applications. However, it is not easy to configuring and maintaining these customized rules. Once hackers launch a multi-step attack and other advanced attacks, it's difficult to extract attack sequences and further recognize attack patterns from massive WAF alert data. In reality, we have lots of different sequences, but we do not know the rules hidden in these disordered attacks. Therefore, mining multi-step attack patterns is an important and significant task in WAF alert data analysis.

In this study, we propose a novel three-stage approach that can automatically mining multi-step attack patterns from massive WAF alerts. The main idea is to cluster attack sequences that are generated automatically, and then recognize an attack pattern from each attack sequence cluster. Different from prior approaches [1], [2], [3], our proposed approach does not need any additional information except WAF alerts. Moreover, it is can be implemented easily.

We evaluate our proposed approach on WAF alerts that are collected from one real-world system. Experimental results show that some seemingly unrelated alerts may have the same "attack background" and other similar information in common. This conclusion is manually verified by offline tracking the IP addresses used by different attackers in our experiments. Thus, our approach could make sense in analysis of massive WAF alert data.

II. Our Approach

To mine multi-step attack patterns from massive WAF alert data, there are three open problems needed to be addressed. The first open problem is how to extract attack sequences, which is defined as all the alerts that are trigged by the same attack. The second open problem is how to put the similar attack sequences together and constitute a cluster. At last, we need to address the problem of recognizing the inherent attack pattern from multiple similar attack sequences. Each stage of our approach addresses one of the above problems respectively, as described in detail in the following part.

Extract Attack Sequence: a WAF alert mainly consists of *alert time, attacked domain name, attacker IP address* and *alert type* (indicating the specific step during the process of an attack). Each attack step may trigger many WAF alerts of the same alert type. For WAF applications, we make a reasonable assumption that each attacker IP address is involved in every step of a multi-step attack, namely all the steps are not launched sequentially with different IP addresses. Then we can get many attack sequences with *attacker IP address* and *attacked domain name* together as the primary keys[1]. As a attack does not last for ever, we think two continues WAF alerts of the same *attacker IP address* and *attacked domain name* belong to different attacks if the time interval between their *alert times* is over 12 hours. Then a long attack sequence is split into several short but rational attack sequences.

Cluster Attack Sequence: to complete the clustering operation, we should first give a tool to measure the distance of two different attack sequences, which is the sum of the distance of their *attacker IP addresses*, the distance of their *attacked domain names* and the distance of their *alert type* sequences in this paper. Regarding each IP address (domain name or alert type sequence) as a string, we design an improved levenshtein distance, referred to θLD hereafter, to measure its distance with another IP address (domain name or alert type sequence). Different from the classical levenshtein distance that gives the same weight for differences at any position, θLD gives different weights. In this paper we define θLD as follows:

$$\theta LD(x_{[1,i]}, y_{[1,j]}) = \theta \times \min(r, s, t)$$

$$\text{where} \quad \begin{cases} r = \theta LD(x_{[1,i-1]}, y_{[1,j]}) + 1 \\ s = \theta LD(x_{[1,i]}, y_{[1,j-1]}) + 1 \\ t = \theta LD(x_{[1,i-1]}, y_{[1,j-1]}) + \begin{cases} 0 & \text{if } x_i \neq y_j \\ 1 & \text{if } x_i = y_j \end{cases} \end{cases}$$

[1] Continuous WAF alerts of the same *alert type* are merged into one alert.

TABLE I
PRIMARY INFORMATION OF WAF ALERTS USED IN THIS PAPER

# of WAF Alerts	75670
# of Attack Sequences	2574
# of Effective Attack Sequences	368
# of Alert Typess	13
Duration	3 days (2014/12/19-2014/11/21)

Obviously we can get θLD based on the idea of dynamic programming. Parameter θ measures the weight ratio between two adjacent positions in a string. Smaller θ means that the rearward position takes much more weights. For IP address, domain name and alert type sequence, we do not use the same θ. The θ for IP address (denoted as θ_{ip}) should be more than 1, as two IP addresses with the same prefix is more similar than two IP addresses with the same suffix of the prefix's length. The θ for alert type sequence (denoted as θ_{ats}) should be 1, as alert types in different positions play the same importance. The θ for domain name (denoted as θ_{dn}) should be less than 1, the primary reason is opposite to that of IP address. We set the $\theta_{dn} = \frac{1}{\theta_{ip}}$ in this paper. Note that the same θLD implies different meanings for string pairs of different lengths. Thus, we normalize θLD in this paper, namely we use $\frac{\theta LD(x_{[1,i]}, y_{[1,j]})}{\max(i,j)}$ as the distance of two strings $x_{[1,i]}$ and $y_{[1,j]}$. Note that $\max(i,j)$ is the maximum value of $\theta LD(x_{[1,i]}, y_{[1,j]})$. A threshold η is introduced to limit the distance of two attack sequences in the same cluster.

We argue that, in addition to WAF alerts, our approach does not need any additional information to complete the clustering operation.

Recognize Attack Pattern: multi-step attack patterns in a attack sequence cluster are automatically discovered by the longest common alert type subsequence extraction algorithm based on the idea of dynamic programming.

III. Evaluation

We evaluate our multi-step attack pattern mining approach, in terms of the effectiveness of attack patterns, using real-world WAF alert data obtained from a famous Chinese ISP, as outlined in Table I. We get 2574 different attack sequences in total from obtained WAF alerts. As some multi-step attacks are incomplete, we remove these attack sequences of only one alert type, and get 368 effective active sequences.

Table II shows the distribution of alert types in our WAF alerts. We can find that top 4 alert types ($\frac{4}{13} = 30\%$) take more than 90% WAF alerts.

Figure 1 shows the change of the number of clusters with the increase of η. We know the distance of any two attack sequences is no more than 3. From Figure 1 we can find that the number of clusters decreases with the increase of η, and we can only get one cluster when $\eta \geq 1.8$.

We use $\theta_{ip} = 1.5$ and $\eta = 0.8$ in our experiments. Under these values we get 24 clusters, and we show several attack patterns of the most frequently used in Table III.

TABLE II
DISTRIBUTION OF ALERT TYPES

Rank	Alert Type	Number	Ratio
1	SQL Injection	43123	56.99%
2	Vulnerability Protection	14636	19.34%
3	Protocol Violations	6650	8.79%
4	Scanning	4514	5.97%
5	Cross-Site Scripting	3320	4.39%
6	Restricted Files	1656	2.19%
7	Others	1771	2.34%

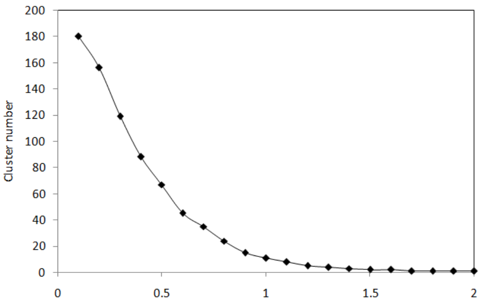

Fig. 1. Change of the number of clusters with η

TABLE III
DISTRIBUTION OF ALERT TYPES

No.	Attack Pattern
1	⟨Vulnerability Protection, SQL Injection⟩, and repeats
2	⟨Vulnerability Protection, Scanning⟩
3	⟨Restricted Files, Protocol Violations⟩
4	⟨SQL Injection, Protocol Violations⟩, and repeats
5	⟨Cross-Site Scripting, Protocol Violations⟩, and repeats

We gather these IP addresses that attack the same domain name with the same attack pattern in a very small time window. We find that some IP addresses are parters to launch an attack although they seem not to be, which is confirmed by our offline investigations. This case verifies the effectiveness of our multi-step attack pattern mining approach.

IV. Conclusion

In this paper, we proposed a novel three-stage approach to mine multi-step attack patterns. We also introduced a new tool to measure the distance of any two attack sequences. Experiments on real-world WAF alerts and offline investigations show that different attackers with the same pattern may have the same "attack background", which verifies the effectiveness of our approach.

References

[1] L. Wang, A. Ghorbani, and Y. Li, "Automatic Multi-Step Attack Pattern Discovering," *International Journal of Network Security*, vol. 10, no. 2, pp. 142–152, 2010.

[2] B.-C. Cheng, G.-T. Liao, C.-C. Huang, and M.-T. Yu, "A Novel Probabilistic Matching Algorithm for Multi-Stage Attack Forecasts," *IEEE Journal on Selected Areas in Communications (JSAC)*, vol. 29, no. 7, pp. 1438–1448, 2011.

[3] Z. Liu, C. Wang, and S. Chen, "Correlating Multi-Step Attack and Constructing Attack Scenarios based on Attack Pattern Modeling," in *Proc. of Information Security and Assurance (ISA)*, 2008, pp. 214–219.