ISSN (ONLINE): 2454-9762
ISSN (PRINT): 2454-9762
Available online at www.ijarmate.com

International Journal of Advanced Research in Management, Architecture, Technology and Engineering (IJARMATE) Vol. 6, Issue 2, February 2020

Application Research on Transformation of Association Rules to Snort Rules

Xuecheng Liu

College of Mathematics and Statistics, Taishan University, 271000, Tai’An, ShanDong, China
Email:controllxc@126.com

***Abstract*—In intrusion detection system used the algorithm of association rules, in the time of detect it needs recalculate some statistical data, but this approach reduce the speed and accuracy of detection, therefore, so this paper proposed a method that put data mining association rules into Snort rules, this not only improves the speed and accuracy of intrusion detection rate, also makes intrusion detection has certain adaptive capacity.**

***Index Terms*— Network security, intrusion detection, association rules, Snort rules.**

I. Introduction

Under the condition of huge network traffic and more and more complex network environment, traditional IDS, which relies on experienced experts to manually update the rule base, is facing the challenge of constantly updating of hacker’s attack methods. Therefore, it is necessary to update the rules in a more automatic and systematic way. Data mining technology can achieve this goal. Data mining technology can analyze the raw data quickly and automatically, and it can make inductive ratiocination and find valuable patterns. Applying data mining into intrusion detection can find useful data association information from the intercepted package, reduce the work intensity of manual analysis, reduce the difficulty of building rules, and improve the accuracy of rule base.

Wenke Lee of Columbia University first introduced data mining into the field of intrusion detection, and establishes an intrusion detection framework based on data mining. Lee has done a lot of work in introducing data mining into intrusion detection, used Ripper algorithm to model intrusion detection, and discovered intrusion behavior through real-time mining. But these methods have some disadvantages, his algorithm need to classify the training data. At the same time, the limitation of classification algorithm makes the classification of training data more complex.

How to use data mining technology to improve the performance of intrusion detection has become a very active research hotspot, many scholars have done research and made some achievements. Snort is a lightweight open source code intrusion detection system. Many detection rule base of commercial IDS are compatible with rule base of Snort, so Snort’s detection base is representative. Then this paper proposed a method of rule updating based association rules mining, this method used association rules data mining technology to update the rule base of Snort, so as to resist effectively the new intrusion mode.

II. Snort Rules

Snort rule can be divided into two logical parts: rule header and rule option. The rule header contains the information of rule action, protocol, source IP and destination IP, subnet mask, source port and destination port, etc. The rule options contain the alert information and the packet area location information to be checked to determine whether to trigger the rule response action. The Snort rule is as follow:

alert tcp any any - > 192.168.1.0/24 111(content:“ 0001 86 a5 ”; msg:“ mountd access”)

The part from the beginning of the rule to the parenthesis is called the rule header; the part enclosed in parenthesis is called the rule option. The word before the colon in the rule option is called option keyword. It should be noted that rule options are not required for every rule, they are just used to better define the type of packets that need to be processed (logging, alerting, and discarding). The corresponding rule action will be triggered only when every element in the rule is true. In this sense, the relationship between rule elements is a kind of logical ‘and’. Meanwhile, among the various rule sets in each rules file, a more extensive logical ‘or’ relationship is formed

III. Improved Association Rules Algorithm

A. *Traditional association rules*

Association rules is one of the main research patterns of data mining, it focuses on finding relationships between different attributes. Running association rules in intrusion detection can discover the relationship between intrusion behaviors.

Definition 1: Association rules

Let $\mathrm{I} = \{I_1, I_2, ..., I_m\}$ be an itemset, an association rules is $R: X \Rightarrow Y$, where $X \subset \mathrm{I}, Y \subset \mathrm{I}, X \neq \varnothing, Y \neq \varnothing$, and $X \cap Y = \varnothing$, it means that if itemset X appears in a transaction, it will inevitably cause itemset Y to appear in the same transaction. X is the precondition, Y is the result of the rule.

Definition 2: The Confidence of association rules

For association rules $R: X \Rightarrow Y$, the Confidence of rule R is defined as:

$$Confidence(R) = P(Y \mid X) = \frac{P(X \cup Y)}{P(X)} \qquad (1)$$

The confidence describes the reliability degree of the rule.

Definition 3: Minimum Support Threshold

The minimum support threshold indicates the minimum support threshold required by discovery association rules for data items, that is *min_sup*, it represents the lowest statistical importance of a data itemset.

Definition 4: Minimum Confidence Threshold

Minimum confidence threshold indicates the minimum confidence that association rules must meet, that is *min_conf*, it represents the lowest reliability of association rules.

Definition 5: Frequent Itemset

The set of item is called the itemset. The itemset with K items is called k-itemset. The appearing frequency of itemset is the number of transactions including itemsets, which is referred to as the frequency or support rate of itemset. If the appearing frequency of itemset is greater than or equal to the product of *min_sup* and the total number of transactions, then the itemset satisfies the minimum support threshold *min_sup*. If the itemset meets the minimum support threshold, it is called frequent itemset.

B. *Improved association rules*

When data mining, if there are some attributes in the database with high probability of appearing, for example, the probability that the value of *duration* attribute is 0 is 0.8, however, the probability of other attributes is relatively low, so it can consider that *duration*=0 is inevitable, so this kind of high probability event can be ignored in mining, after the frequent itemset have been generated, it will be connected to the large probability event, so the final association rules can be got.

Suppose the database has 20 attributes, the number of generated itemsets is $C_{20}^{8} = 125970$ using traditional association rules to generate 8 itemsets. If the improved association rules algorithm is used to generate 8 itemsets, the large probability event will be deleted, and then the number of generated itemsets is $C_{18}^{8} = 43758$. As you can see, when there are many database attributes, the time complexity and the space complexity has obviously reduced by using the improved association rules algorithm.

IV. Constructing Snort rules with improved association rules algorithm

A. *Date preprocessing and data mining*

Because the data of network data set is binary, and the packets of various protocols are interwoven with each other, so the raw data set is not suitable for data mining, it need to be preprocessed. The specific process of data preprocessing is to process data sets into connection records of transport layer, each record includes *timestamp*, *duration*, *protocol*, *bytes_sent_by_responder*, *local_host_IP*, *remote_host_IP*, *state*, *flags*.

The number of association rules obtained by data mining depends on *min_sup* and *min_conf* of the association rules. If the setting value of *min_sup* and *min_conf* are higher, then the association rules gained will be less, this may result in the loss of connections records that are large in number in a short period of time, but have a low proportion of the entire dataset. Conversely, if the setting value of *min_sup* and *min_conf* are lower, a lot of meaningless association rules will be gained, which cause trouble to the analysis process. The solution is to set the axis attributes, for example *state*, if the candidate set does not contain axis attributes, it will be regarded as meaningless rule and abandoned. Choosing the appropriate axis attribute can not only mine the effective association rules, but also reduce the time and space complexity of the data mining algorithm.

In addition, in order to improve the detection rules, the connection records will be needed to make statistics information, the statistics information will be as reference attributes, including the number of total protocols in the dataset, the total number of connection records of each protocol, the number of bytes sent by both host using this protocols, the percentage of completed connection records, which are the end with SF, are the full connections records with normal flag FIN, the percentage of local host initiated connection records, the percentage of connections records initiated by hosts in adjacent networks.

B. *Convert association rules to Snort rules*

The Apriori algorithm is used to mine the connection records, and the understandable rules are obtained after code conversion, one of them is as follows:

195.73.151.50= > 172.16.112.149 http SF[3%,60%]

This association rule indicates that host 195.73.151.50 sends a large number of packets to host 172.16.112.149 by HTTP protocol in the LAN, the support threshold is 3%, and the confidence threshold is 60%.

The duration of the association rule will be calculated, and at the same time, the reference properties such as connection rate, average number of connections, the sending rate of sender and responder should also be calculated. If the behavior is found to send a large number of packets to the target host in a short time, then it is a DDOS attack, and transforms the behavior to Snort rule:

Alert TCP 195.73.151.50 any- > 172.16.112.149

As long as the host 195.73.151.50 sends data packets to port 8080 of the host 172.16.112.149 in the LAN, Snort will consider this as an attack on normal web server, thus will send an alarm.

It is worth noting that the source IP of the above Snort rule is applicable to specific IP address, which may cause many false alarms. However, when the network packets are been mined in real time, if a large number of packets are sent to host in LAN by 8080 port in short time, then this behavior can be considered that there are hackers sending DDOS attacks. In this case, the host's specific IP address is used instead of *any* in the Snort rule; it can shield the attack of the host more effectively.

C. *Transform association rules to Snort rules*

Because Snort is an intrusion detection system based on packet matching, its rule base are composed of rule headers and rule options, the previous transformation process is just to transform association rules into rule headers, and the rule options is still blank. In order to reduce false positives, data packets will be analyzed and the rule options well be generated. The detailed method is to use network packet analysis tools such as Ethereal to analyze the intercepted packets, and finds the packet corresponding to the source IP and destination IP in the association rule, analyzes some information that often appears in these packets, and refers to this attribute to generate corresponding rule options.

V. Simulation Experiment

The network data used in this experiment is downloaded from Lincoln Laboratory of Massachusetts Institute of Technology. The dataset is 64M in size, with 347987 packets in total. The dataset is collected in 1hour and 45minutes using TCPDUMP, which collected from the government websites when a variety of attacks are used against imitative hackers.

When the *min_sup* is set to 0.002, three association rules are obtained:

32.97.253.56 = = > 172.16.116.44 http SF 10.04%100.00% 159 395.636

207.25.71.141 = = > 172.16.116.194 http SF 18.9%73.23% 186 302.946

209.67.29.11 = = > 172.16.116.194 http SF 22.74%29.75% 144 1414.609

The two numbers in the rear are the number of occurrences and duration time. The association rules are transformed into Snort rule base, and then the replay attack is tested with the original dataset, and 3 alerts are obtained, the original data packet have been analyzed using Ethereal, and it is found that one of them is a false alarm. The correct alert is for DDoS, which is based on multiple connections. The reason for false positives is that association rule mining algorithm only relies on frequent program mining, regardless of whether it is normal web page access. Therefore, it is necessary to analyze the authenticity of association rules before the rule is transformed.

VI. Conclusion

Taking snort as an example, this paper proposed a method of constructing IDS rules by using improved association rules algorithm, and used this method to carry out experiments. This method is feasible for DDoS and PROBING type attacks based on frequent connections, because of the limitation of association rule algorithm, it is not suitable for the attack U2R and R2L, which the attack will be completed in one connection. How to reduce human participation, improve intelligence, reduce the generation of redundant rules, improve the performance of rules, and consider using a variety of mining algorithms for processing are further issues to be studied.

References

[1] Peili Qiao, Shibin Zhang, "Implement of the CVE-based intrusioin detection expert system rule base," *Network Security Technology and Application*, vol. 5, pp. 50-52, May 2005.

[2] Feng Yu, Huichan Liu, "Build behavior pattern rule base of intrusion detection system using association rules technology," *Journal of engineering college of armed police force*, No 6,vol. 19, pp.22-24, December. 2003.

[3] FuQian Shi, Zheng Lin, Jiang Xu, "Rotation, scale, and translation resilient public watermarking for images," *IEEE Trans. Image Process.*, vol. 10, no. 5, pp. 767-782, May 2001.

[4] Shuyan Wang, Xiaomei Wang, "Application of association rules for medical data classification," *Compute and Digital Engineering*, vol. 37,No.8, pp. 47-49, August 2009.

[5] Hong Li, Zhihua Cai, "Application of association rules in Medical data analysis," *Microcomputer Development.*, vol. 13, no. 6, pp. 94-97, June 2003.

[6] Lixin Cui, Senmiao Yuan, Chunxi Zhao, "Algorithms for mining constrained association rules," *Chinese Journal of Computers.*, vol. 23, no.2, pp. 216-220, February 2000.

[7] Shijun Li, Hanfei Wang, Dongru Zhou, "Mining and application of statistical relationships in relation databases," *Computer Engineering and Applicationgs.*, pp. 117-118, June 2000.

Xuecheng Liu is a lecturer. He obtained his first degree of Bachelor of Management at the University of Jinan and Master of Science Degree at GuiZhou University. His major fields of study are Network Information Security.