Indian Journal of Science and Technology, Vol 9(41), DOI: 10.17485/ijst/2016/v9i41/86987, November 2016
ISSN (Print) : 0974-6846
ISSN (Online) : 0974-5645

Automatic Firewall Rule Generator for Network Intrusion Detection System based on Multiple Minimum Support

S. Nithya[1]* and C. Jayakumar[2]

[1]SITE, VIT University, Vellore - 632014, Tamil Nadu, India; ms_nithya@yahoo.co.in
[2]RMK Engineering College, Gummidipoondi Taluk, Kavaraipettai, Tiruvallur - 601206, Tamil Nadu, India; cjayakumar2007@gmail.com

Abstract

Background: Association rule mining plays a vital role in predicting the attacks and generating the firewall rules automatically. Data mining techniques discover the knowledge by counting the frequently occurring items, whereas most of the real-world datasets are non-uniform containing both frequently and relatively rarely occurring items. This paper discusses about how to generate the automatic firewall rules to detect anomalies using multiple minimum support. **Methods**: Mining association rules based on single minimum support approach suffers from the dilemma known as 'rare item problem' it requires multiple scans of database which increase the load and time consuming. To avoid this problem Multiple Minimum Support with Probability based approach (MMSP) is used to generate rules. **Findings:** To create a model of current user behavior from the dataset the probability will be compute with threshold value and the alarm will be raised accordingly. By using MMSP, the number of false alarm are reduced during intrusion detection and automatic firewall rules will be generated.

Keywords: Apriori, Firewall, Intrusion Detection, Minimum Support, Probability Approach, Rare Association Mining

1. Introduction

1.1 Association Rule Mining

The patterns hidden in large databases are discovered using the data mining techniques like Association rule mining, clustering and classification techniques. The correlations among the entities in a dataset are discovered by the process of association rule mining. Correlation refers to statistical relationship between two set of items. The association rule mining is governed using two statistical measures and they are *support* and *confidence*.

Definition 1: Given a set of items I= $\{I_1, I_2,\ldots,I_k\}$ and a database of transactions D=$\{t_1, t_2,\ldots, t_n\}$ where ti= $\{I_{i1}, I_{i2},\ldots, I_{nk}\}$ and $Iji \in I$, an *association rule* is an implication of the form X => Y where X,Y$\in I$ are set of items called itemsets and X$\cap$ Y = φ[4].

Definition 2: The *support* (s) for an association rule X=>Y is the percentage of transactions in the database that contain X$\cup$ Y[6].

Definition 3: The *confidence or strength* (α) for an association rule X=>Y is the ratio of the number of transactions that contain X$\cup$ Y to the number of transactions that contain X[6].

For T, the association rules are generated by mining the data. To discover all rules that satisfy min_supand-min_conf constraints. An itemset that satisfies min_sup constraint is called frequent itemset or frequent pattern[8,10].

Based on the literature survey, there exists useful knowledge pertaining to rare entities[7,9]. The rare entities are complex to detect and generalize from the dataset because it has fewer data. The rare knowledge patterns concerning to rare events which can be mined by using

*Author for correspondence

improved approaches like rare association rules. An association rule formed between either frequent or rare items or among rare items is referred as rare association rule. There exists a useful knowledge discovery in rare associations. Some network attacks may occur rarely, but they harm rigorously.

1.2 Intrusion Detection System

An intrusion detection system will detect malicious activities that can compromise the security of the system. There are two ways to detect intrusion: static analysis and dynamic analysis[1,2]. Network intrusion detection system analyses packets on a network which harms the system or cause a Denial of Service (DoS) attack. An NIDS typically runs on a hub or router, it monitors and analyzes the traffic flow. Anomaly intrusion detection and misuse intrusion detection are the classifications of intrusion detection. In anomaly intrusion detection, any action that significantly deviates from the normal behavior is considered asintruder[1,2]. The advantage of anomaly detection is detecting unknown attacks based on audit data but the major issue is high false-alarm and limited by training data.

The rest of the paper is organized as follows. Section 2 discusses the related works. Section 3 explains the existing approach and its pitfalls. Section 4 explains the proposed approach for generating rules for Network Intrusion Detection System. Finally, Section 5 concludes with mention of the future likely enhancements of the system.

2. Related Works

In datamining, the problem of deriving associations from the data has become more popular solution. Several algorithms were proposed for association rule mining. For discovering association rules between items in a large database the Apriori, AprioriTid and a hybrid algorithm Apriori Hybrid has been proposed[11]. These algorithms suffer from the dilemma called 'rare item problem'.

An approach known as Multiple Support Apriori (MSApriori) has been proposed[8,10] to improve the performance of extracting frequent itemsets, involving rare items. In this approach, if an itemset satisfies the lowest MIS value among the respective items, then the frequent itemsets and the support percentage are generated. Based on the generated percentage, each item is assigned with a minimum support value known as "Minimum Item Support (MIS)". This method still suffers from the rare item problem.

Relative Support Apriori Algorithm (RSAA) has been proposed[9] for discovering frequent itemsets involving both frequent and rare items. First supports 1, second supports 2 (s2>s1), and relative support Rsup are the three user specified measures used in this algorithm. The main issue in this algorithm is providing values to these parameters for the given dataset.

To extract rare association rules an improved approach by using the notion of "Support Difference (SD)" in calculating min_sup for each item has been presented[7]. It still suffers from 'rule missing' problem.

3. Existing Approach

As anomaly detection is a data analysis process data mining approach is used to generate rules for intrusion. Apriori algorithm is used because the audit dataset of user behavior is too large.

```
Apriori (T, ε)
k<- 2
        while Lk-l ? 0
                Ck, <- Generate(Lk- 1)
                for transactions t Є T
                        Ct <- Subset (Ck, t)
                        for candidates c Є Ct
                          count |c|<- count|c| + 1
                Lk, <-{c Є Ck | count |c| =ε}
                k<-k+l
return U Lk:
        k,
```

Figure 1. Apriori pseudo code.

In existing approach Figure 1, Apriori algorithm suffers from the problem known as 'rare item dilemma' since it uses single minimum support for mining rules. At high minimum support value, rare itemsets are missed, and at low minimum support value, the number of frequent itemset explodes. To generate candidate sets, the database must be scanned multiple times which will increase the I/O load and also this consumes more time. To solve these problems a novel algorithm based on multiple minimum support and probability concepts is proposed.

4. Proposed Approach

In the proposed approach, Apriori training dataset is generated using the most popular intrusion detection software called snort which is also used to record and logs of user's activities. This data set is used to create a model to compute the probability of raising the alarm by the current activities. If this probability exceeds defined threshold, it will generate a rule to be added to firewall in order to block the intruder.

In Figure 2 instead of using single minimum support, MMSP are used to evade 'rare item problem'. To circumvent multiple scanning of large database, probability approach is used.

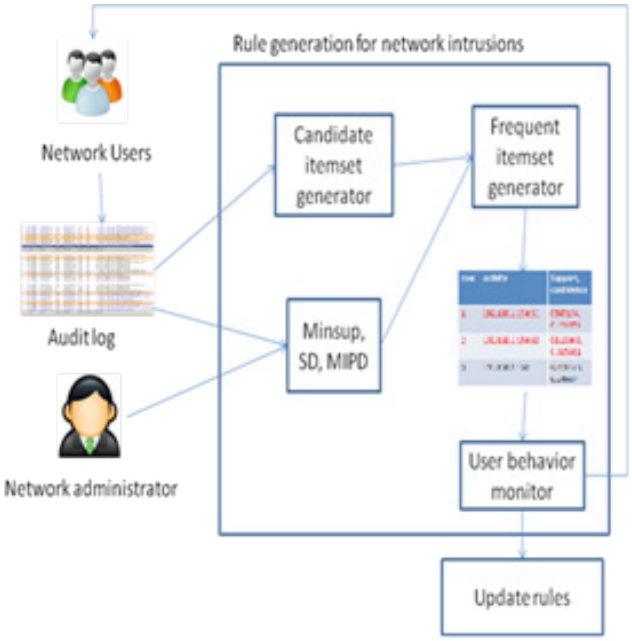

Figure 2. Proposed system architecture.

Pseudocode of the proposed algorithm:

Start;

Calculate Support Difference (SD), Transaction (T), Maxitem Pattern Difference (MPD), Min support and confidence threshold

Generate candidate itemset C1;

Calculate support Support(c1)=count(c1) / T;

Calculate d and e;

d= No of items / No of Transactions;

e=1-d;

Calculate MIS for each item in C1;

MIS=calculate-MIS(c1, Support);

Max_value = max(Support) – MPD If Max_value>min_mis then min_sup1=Max_value; Else min_sup=min_mis; Generate frequent 1-itemset

L1= {<i> | i ε c1, Support (i) ≥ min_mis(i)};

L1= sort(L1,MIS); Calculate MIS and IPD;

for k=2; Lk-1≠φ;k++ do // repeats till no frequent item occurs Ck=candidate-gen(Lk-1);

Ct=subset(Ck,t); for each candidate c ε Ct do

temp_prob(j)= Support (c)

ifmin_mis>MIS(c) then min_min=MIS(c)

j++;

end for Support (ck) = a * min(temp_prob) + (b * min(temp_prob) * max(temp_prob)); Max_value=max(temp_prob);

ifmax_value>= min_mis then

min_sup=max_value

elsemin_sup= min_mis

IPD=max_value – prob_sup(ck);

Lk ={ c ε C | IPD<=MPD &Support(ck) ≥ min_mis& Support(ck)>=min_sup| for all i ε c) } e n d for

return U Lk:

k,

Generate Association Rules. End.

Figure 3. Pseudo code for MMSP algorithm.

By using the Figure 3 MMSP algorithm the number of false alarms is reduced by generating real time rules for firewall and the minimum support values reduces the 'rule missing' and 'rule explosion' problems. Frequent itemsets are generated using probability, without scanning the database so many times.

5. Conclusion

In this paper, an approach to improve the network anomaly detection by generating automatic firewall rules is proposed. To reduce the number of false alarms in anomaly detection and to build better rules for firewall, the data mining approach like association rule mining is used. To generate firewall rules MMSP algorithm is used which involves in extracting the interesting correlation relationship among large set of data items. It avoids 'rare item problem' and reduces the I/O load and execution time. The logs of user activities are recorded using snort and generate the rules for firewall based on the MMSP algorithm. The proposed approach can be extended to

generate rules for heterogeneous network based on multiple-level rule mining technique.

6. References

1. Amudha P, Karthik S, Sivakumari S. An experimental analysis of hybrid classification approach for intrusion detection. Indian Journal of Science and Technology. 2016 Mar; 9(13). DOI: 10.17485/ijst/2016/v9i13/81977.
2. Prasad SNSE, Srinath MV, Basha MS. Intrusion detection systems, tools and techniques – an overview. Indian Journal of Science and Technology. 2015 Dec; 8(35). DOI: 10.17485/ijst/2015/v8i35/80108.
3. Wankhade AD, Chatur PN. Comparison of firewall and intrusion detection system. International Journal of Computer Science and Information Technologies. 2014; 5(1):674–8.
4. Rawat SS, Rajamani L. Probability Apriori based approach to mine rare association rules. 3rd Conference on Data Mining and Optimization (DMO), Selangor, Malaysia; 2011 Jun. p. 253–8.
5. Chang R,Liu Z, An improved Apriori algorithm. International Conference on Electronics and Optoelectronics (ICEOE2011); 2011. p. 476–8.
6. Saboori E, Parsazad S, Sanatkhani Y. Automatic firewall rules generator for anomaly detection systems with Apriori algorithm. 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE); 2010. p. 57–60.
7. Kiran RU, Reddy PK. Mining rare association rules in the datasets in which items' frequencies vary widely at Center for Data Engineering. International Institute of Information Technology, Hyderabad. 2010; 5981:49.
8. Kiran RU, Reddy PK. An improved multiple minimum support based approach to mine rare association rules. IEEE Symposium on Computational Intelligence and Data Mining (IEEECIDM), TN; 2009. p. 340–7.
9. Yun H, Ha D, Hwang B, Ryu KH. Mining Association rules on significant rare data using relative support. The Journal of Systems and Software. 2003; 67(3):181–91.
10. Liu B, Hsu W, Ma Y. Mining association rules with multiple minimum supports. SIGKDD Explortions; 1999. p. 337–41.
11. Nithya S, Jerlin MA, Charanya R, Jayakumar S, Rathi R. Self-restorative cluster head selection in heterogeneous network. Global Journal of Pure and Applied Mathematics. 2015 Jun-Jul; 11(3):1655–2522.
12. Agrawal R, Srikant R. Fast algorithms for mining association rules. Proceedings of 20th International Conference on Very Large Data Bases, Santiago, Chile; 1994 Sep. p. 487–99.
13. Jerlin MA, Jayakumar C. A dynamic malware analysis for windows platform-a survey. Indian Journal of Science and Technology. 2015; 8(27).