Framework for False Alarm Pattern Analysis of Intrusion Detection System using Incremental Association Rule Mining

Won Yong Chon, Eun Hee Kim, Moon Sun Shin, Keun Ho Ryu

Database Laboratory, Chungbuk National University
CheongJu Chungbuk, 361-763, Korea
{ chonwy2000, ehkim, msshin, khryu }@dblab.cbu.ac.kr

Abstract: The false alarm data in intrusion detection systems are divided into false positive and false negative. The false positive makes bad effects on the performance of intrusion detection system. And the false negative makes bad effects on the efficiency of intrusion detection system. Recently, the most of works have been studied the data mining technique for analysis of alert data. However, the false alarm data not only increase data volume but also change patterns of alert data along the time line. Therefore, we need a tool that can analyze patterns that change characteristics when we look for new patterns. In this paper, we focus on the false positives and present a framework for analysis of false alarm pattern from the alert data. In this work, we also apply incremental data mining techniques to analyze patterns of false alarms among alert data that are incremental over the time. Finally, we achieved flexibility by using dynamic support threshold, because the volume of alert data as well as included false alarms increases irregular.

Keywords: Intrusion detection system, Incremental Mining, Association rules.

1. Introduction

There are several unsolved problems of the network intrusion detection systems (NIDS). In order to manage the network efficiently, we must solve the problem about how to reduce the false alarm in Intrusion detection system. False alarm is defined as alert that NIDS make to respond to normal data. There are two kinds of false alarm. One is a false negative that declines accuracy of the NIDS. The other is a false positive that declines efficiency of the NIDS. These kinds of a large quantity false alarms caused decline of security service. Because of reducing the correctness of the network, network administrator can't protect the network. To protect the network system, we need the framework that analyzes the false alarms frequently. And, this tool can improve the efficiency of the NIDS. According as alert data is usually growing, alert data character continuously changes. In order to analyze the growing alert data, we propose the incremental mining technique.

In order to state this paper efficiently, we guide it as follows. Section 2 describes the false alarm reduction methods and the incremental mining technique as a related works. Section 3 describes the association methods and incremental association technique. To test the incremental association technique, we compared the proposed mining technique with the previous one in the fourth chapter. In the last chapter, we stated the profit of this framework and conclude this paper.

2. Related Work

There are many studies of the methods to reduce the false alarm to improve the network intrusion detection system (NIDS) [3]. There are several tropical examples, log data correlation analysis methods [4], an alert correlation analysis [5, 6], a study of alert minimizing to utilize the data mining technique [1][2], and so on. But, these kinds of studies have a defect in dependence on selected data feature. Data mining is an act of analyzing a database and searching for new facts based on the original database. The process of data mining provides knowledge in the form of rules and patterns based on statistical analysis of data. The process is challenging because the source databases from which the knowledge is extracted are large and growing. The knowledge itself is time- varying, as some rules and patterns may hold at present but not in the future [8]. We can use this mining technique for reducing the volumes, using the method of extracting the rule and filtering the alert data.

In this paper, we suggest the framework for false alarm analysis using incremental mining technique. Incremental mining technique maintains the previous mining rules. When some data are inserted into original data, generally speaking, an old large itemset has the potential to become small in the updated database. Similarly, an old small itmeset could become a large one in the new database [9]. In conclusion, Incremental mining is a maintaining technique of the mined rules using the minimum calculating [6].

3. Framework for the analysis of false alarm

In this paper, we made a framework using the incremental techniques. The volumes and features of False alarm data will increase and change by the time. For analysis the false alarm data, previous data mining techniques mined the whole data. But incremental mining technique is just using the inserted data and mined data. Following figure describes the whole framework of the false alarm pattern analysis.

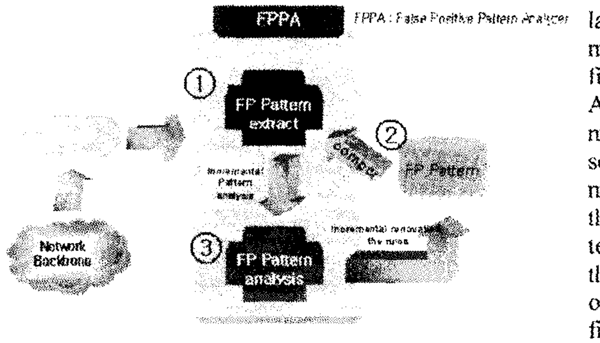
FPPA
FPPA : False Positive Pattern Analyzer
①
FP Pattern extract
②
compa
FP Pattern
Incremental Pattern analysis
Network Backbone
③
FP Pattern analysis
Incremental renovation the rules

Fig. 1. False alarm analysis framework

The whole framework consists of classification, NIDS and analysis parts. The classification part (1) selects the false positive data using the classification rules. And then hand the selected data to the NIDS. The NIDS make the alert corresponding to the rules (2). And, continually analyze the whole data and make the alert. We suggest the false positive analyzer (3). We analyze the selected false positive data using incremental mining methods.

3.1 Incremental updating module for the rule : 1 – large items

table. 1. Parameter Table

notation	description	notation	description
DB	original database	db	increased database
C(DB)	Candidate of DB	C(db)	candidate of db
L(DB)	large items of DB	L(db)	large items of db
U: DBUdb	updated database	C': DBUdb	candidate of U
L':DBUdb	large items of U	Support(s)	Support

Table I is a list of the parameters used in this paper. In order to analyze alert pattern, we apply and extend the association mining technique. Incremental data mining maintains the discovered rules. And the features are runs as follows. The large items of the DB can remain the large items in the database U. And the candidate of the DB can regenerate the large items in the database U. To generate the large items in U, we must check whole database U. The process of the incremental mining is an original database DB with increased database db. Therefore this incremental mining technique is more efficient than mining from scratch. But, this method has some shortage. We maintain some additional mined data and continuously use the same support value. To do the efficient mining, we adjust the previous incremental mining methods. So, we can change the support value.

The association rule X → Y holds in DB with confidence if c % of the transactions in DB that contain X also contain Y. The association rules X → Y has support s in DB if s % of the transactions in DB contains X ∪ Y. The problem of mining association rules is to find out all the association rules whose confidence and support are larger than the respective thresholds [9]. The association mining passes through two phrases. At the first step, we find the large itemsets that satisfy the minimum support. At the second step, we find the rules that satisfy the minimum confidence from the large itemsets. Since the second step is generated easily using the large itemsets, most studies concentrated on how to efficiently generate the large itemsets [7]. Incremental association mining technique generates many candidate items, and generates the association rules using the mined data. The process of the incremental association mining has two steps. The first process is a 1- large items generating. The second process is n+1 - large items generating using the n- large items. Following figure describes the 1- large items generate process.

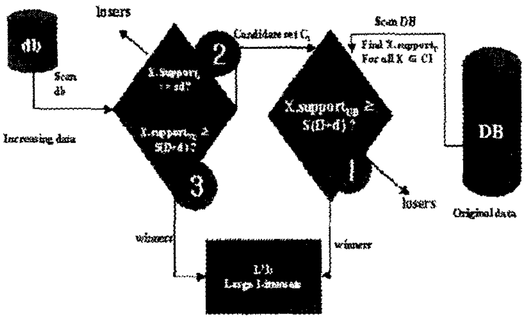

Fig. 2. Incremental rule module 1- large items

The database DB is mined and maintained the large items and candidate. The large items in Updated U satisfied with the support (S (D+d)). Firstly, the large items in the DB can remain the large items in U or can exclude the large items in U. Therefore we check the large items in DB and combined the large items with candidate items in db. And then check the combined support value (1). Next step, we check the candidate of DB (2, 3). The candidate in the DB can remain as the candidate in U or can regenerate the large items in U. But the candidate in db can't generate large items in U when combined any data in DB. So, we only check the large items in the db (2). and combined the data in the DB. And, check the combined support value (3). Then, we can find all large items in U.

3.2 Incremental updating module for the rule : n – large items alarm

The n-large items steps make a candidate and check this candidate satisfied with the support value. If the support of the Cn is satisfied, the Cn can remain L'n. The Cn is generated by previous large items. Firstly, if the Cn is large items in the DB, then it can be remained in U or excluded in U. We check the count in DB and in db. If the value (Ln.support + C'n.support) is bigger than S (D+d), the Cn is the L'n in U (1).

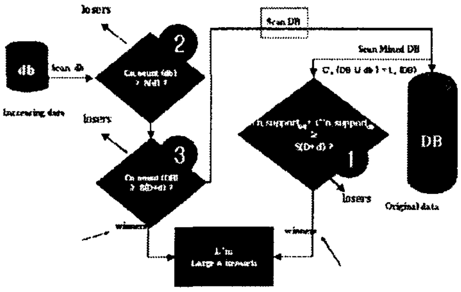

Fig. 3. Incremental rule module n- large items

If the Cn is not large items, we must check the support in the db. And, only the Cn is bigger than S(d) in the db(2), we combined the support and check S (D+d). If the support is bigger that S(D+d), the Cn is the L'n in U(3). The Cn is not large items in the DB, we must check the mined database DB and increased data db. Until the db is empty or k-large items are zero, repeat this process. To efficiently analyze of the incremental data, we must minimize the new scan from the DB in the third process of this figure. Incremental association mining doesn't deal the whole data U, but only uses the mined data and increasing data. And it uses the minimum scans the previous data. So, we can generate the efficient rules. Following figure describes the steps of the incremental association.

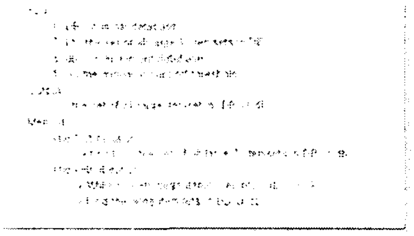
[illegible]

Fig.4. A description of the Incremental association mining

4. Experimental Evaluation

We compared incremental association mining with the previous apriori-based algorithms. Firstly, we classify the alert data using the classifier in the chapter 3. We firstly check the accuracy of the mined rule. So we mined the same data using the apriori-based algorithms from scratch and the incremental mining methods. As in the picture, the incremental mining technique is faster than the previous data mining technique. And the mined rules are the same.

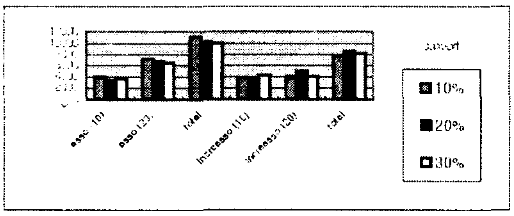

Fig. 5. Performance Ratio

The incremental mining only mined the increasing data and mined data. And then it combined the support and checked the support in the whole database. So these methods are very efficient methods to analyze the incremental data. The gab of the changing data is caused by the scan of the previous data.

5. Conclusions

Intrusion detection systems generate many kinds and volumes of alert data. And these alert data are increasing and changing the characters over the time. It is necessary to do efficient management over the time, so we need the pattern analysis tool. In order to analyze false alarm pattern created in intrusion detection system, we apply incremental mining extended the previous mining method to analyze the alert data. If we use the previous mining technique as before, we must spend much cost and time since the previous mining technique scans the whole data every time. So, we must always repeat the overlapping work. But incremental mining technique mines one previous mined data and increasing data. So this method is more efficient than the previous one. Using the experiment of this framework, we improved the efficiency of the suggested framework

References

[1] W. Lee, S. J. Stolfo, K. W. Mok " A Data Mining Framework for Building Intrusion Detection Models" 2001

[2] M J Lee, M S Shin, H S Moon, K H Ryu, K Y Kim, "Design and Implementation of Alert Analyzer with Data Mining Engine " International Conference on Data Engineering and Automated Learning,2003

[3] K. Julisch. " Dealing with False Positives in intrusion Detection " In 3nd Workshop on Recent Advances in intrusion Detection, 2000

[4] Cuppens. F, Miege. A. " Alert Correlation in a cooperative Intrusion detection framework " In proceedings of the IEEE Symposium on Security and Privacy, 2002

[5] H. Debar, A. Wespi " Aggregation and Correlation of Intrusion-Detection Alerts " In Recent Advances in Intrusion Detection, number 2212 in Lecture Notes in computer Science, P 85-103, 2001

[6] M. Lin, S. Lee " Incremental Update on Sequential Patterns in Large databases " Proceedings of the Tools for Artificial Intelligence Conference, 1998

[7] R. Agrawal, R. Srikand " Fast Algorithms for Mining Association Rules in Large Databases " In Proc, 20th Int. Conf. On Very Large Databases, 1994

[8] N. L. Sarda, N. V. Srinivas " An Adaptive Algorithm for Incremental Mining of Association Rules " In Proc, 9th international Workshop on Database and Expert Systems Applications, 1998

[9] David W. Cheung, Jiawei Han, Vincent T. Ng, C. Y. Wong " Maintenance of Discovered Association Rules in Large Databases : An Incremental Updating Technique " In Proc, 12th International Conference on Data Engineering, 1996