Intrusion Alert Correlation Based on UFP-Growth & Genetic Algorithm

Anand Jawdekar **Vineet Richariya**

LNCT, Bhopal, India Department of Computer Science

Abstract
Intrusion alert correlation is an important factor for network security assessment. In the current scenario various security assessment algorithm are available for risk calculation. These algorithms were qualitative in nature. It is difficult for security managers to configure security mechanisms. The paper discuss the problem of managing alerts. A novel approach for intrusion alert correlation using UFP-Growth and Genetic Algorithm is presented in this paper. UFP-Growth is used for association rule mining and genetic algorithm is used for finding optimal pattern. The proposed method implement in MATLAB 7.8.0. For implement purpose various function and script file were written for implementation of model. For the test of our hybrid method, we used DARPA KDDCUP99 dataset. Our proposed method compare with existing ACR (assessment of credibility and risk) technique and getting better result such as risk calculation and minimized alert correlation rate.

Keywords
Alert correlation rate, Intrusion alert correlation, Kdd, risk calculation, etc.

1. Introduction

Study on the network intrusion detection system began in the 20th century, 80's, has been more than 20 years, however, the development of intrusion detection technology, is far from the objectives and expectations[1]. The current intrusion detection system still face the following problems: (1) lack of capacity for the accurate description of the attack (2) The IDS does not recognize the relevance of alert information, (3) lack of visualization for the study of alert information.

Therefore, it is necessary to make correlation and adjustment analysis on IDS alert information, to reduce false alert rate and omission rate; to concise alert information, and provide these higher-level security information such as the strategy and process of the intrusion in a visual form to the security administrator, and predict the next possible attack, so that administrators may take timely measures to put an end to the ongoing and imminent attacks. The security of our computer systems and data is always at risk. The extensive growth of the internet has prompted network intrusion detection to become a critical component of infrastructure protection mechanisms. Network intrusion detection can be defined as identifying a set of malicious actions that threaten the integrity, confidentiality, and availability of a network resource. The conventional main intrusion detection techniques are misuse detection and anomaly detection. All the two aspects may lead to high false alarm rates[2][3].

In recent years, risk arises from information system and networks from time to time. As its impact is expanding, people realize it is time to seek a valid solution to these threats and risks for providing information security and society security as well. Information security risk assessment (risk assessment for short) refers to assessment process of information system and confidentiality, integrity and continuity in transmission, processing and storage, according to system vulnerability, threat and actual negative impact caused by threat source, then identify risk of information security based on possibility of threats and the extent of negative impact [4].

Alert correlation is a promising technique in intrusion detection. It analyzes the alerts from one or more intrusion detection system and provides a compact summarized report and high-level view of attempted intrusions which highly improves security effectiveness. Alert correlation [5-8] is used to (1) reduce the number of alerts that an IDS would generate to more manageable levels while still retaining strong detection capacities, (2) improve IDS correctness by reducing the false positives and negatives in the alerts generated by the IDS sensors, and (3) unveil an intruder's intrusion strategy after the attack has happened. alert correlation problem is to extract "true" alerts (or filter out the false alerts) from the raw alerts generated by the IDS sensors by utilizing relationships (e.g., similarities, sequential relationships, etc.) among alerts.

A novel approach of intrusion alert correlation based on UFP-growth and genetic algorithm is presented in this paper. It minimizes the false alarm rate in IDS and give the quantitative values of security parameters (risk calculation, alert correlation rate, positive correlation, negative correlation) to the network manager, it can help security managers adjust the corresponding security mechanism and choose the response method against attack in detail.

The remainder of this paper is organized as follows. The next section discussed related work. Section 3 describes proposed algorithm in detail. An implementation and evaluation is described in Section 4. The conclusion is in Section 5.

Manuscript received September 5, 2015
Manuscript revised September 20, 2015

2. Related work

In the following we summarize some of the recent research works in the area of alert correlation and alert mining. Li Yang and Dong Xinfa [9] describe Alert Correlation Model. In this title authors describe method and solution as the multi-step attack is one of the primary forms of the current network intrusions. Through the study on patterns of the multi-step attack, a model of alert correlation which is based on self-regulate is designed. It describes the definition and classification of alert correlation. Also it introduces the association rules. To improve efficiency of IDS, we applies data mining technology to IDS. We present a method of how to acquire the intrusion knowledge from the logs and detect the intrusion behaviors based on the improved Apriori algorithm.

Jin SHI, Guangwei HU, Mingxin LU and Li XIE[10] describe ACRL approach(assessment of credibility, risk and loss). It assesses attack sequences from credibility, risk and the loss of system and provides the assessment values to security managers. It can assess the network security mechanisms and measures in position and can help security managers adjust the corresponding security mechanisms and choose the response methods against attacks in detail.

Lu Simei, Zhang Jianlin, Sun Hao, Luo Liming[11] describe Security Risk Assessment Model Based on AHP/D-S Evidence Theory. proposed AHP/D-S evidence theory handle the uncertainty of the system. Compared with other methods, the analysis of hierarchy process (AHP) method has been widely used in security risk assessment, for this method can change from the qualitative index into quantitative index. Realistic risk assessment involves many uncertainty factors, some of which are even unknown. proposed a risk assessment model which is generated by combining AHP method with D-S method to solve these problems. Not only does the AHP/D-S method combine the advantages of both, but also can solve uncertain problems more scientifically. Finally, a sample of how to use AHP/D-S method in security risk assessment is given to prove our method. Previous research works on risk assessment methods [12,13] almost all qualitative and to system overall. Our approach focuses on assessment in position and in the network system. It can help system managers to quantitatively adjust the configuration of network system in detail.

3. Proposed work

3.1 UFP-Growth Method For Finding Frequent Itemsets

One of the most popular data mining approaches is to find frequent itemsets from a transaction dataset and derive association rules. Finding frequent itemsets (itemsets with frequency larger than or equal to a user specified minimum support) is not trivial because of its combinatorial explosion. Once frequent itemsets are obtained, it is straightforward to generate association rules with confidence larger than or equal to a user specified minimum confidence.

Frequent pattern growth, or simply FP-growth, which adopts a divide-and-conquer strategy as follows. First, it compress the database representing frequent items into a frequent pattern tree, or FP-tree, which retains the itemset association information. It then divides the compressed database into a set of conditional database (a special kind of database), each associated with one frequent item or “pattern fragment”, and mines each database separately. For each “pattern fragment”, only its associated data sets need to be examined. Therefore, this approach may substantially reduce the size of the data sets to be searched, along with the “growth” of patterns being examined.

Procedure UFP_growth(Tree,α)

1. If tree contains a single path P then
2. For each combination (denoted as β) of the nodes in the path P
3. Generate pattern β U α with support_count=minimum support count of nodes in β;
4. Else for each a_i in the header of Tree{
5. Generate pattern β= a_i U α with support_count= a_i . support_count;
6. Construct β’s conditional pattern base and then β’s conditional FP_tree $Tree_\beta$;
7. If $Tree_\beta \neq \Phi$ then
8. Call FP_growth($Tree_\beta$, β);}

3.2. Genetic Algorithm

Using UFP-growth algorithm frequent itemsets can be found, for improving the efficiency of the algorithm we used genetic algorithm for obtaining optimal pattern. The application of the genetic algorithm in the context of data mining is generally for the tasks of hypothesis testing and refinement, where the user poses some hypothesis and the system first evaluates the hypothesis and then seeks to refine it. Hypothesis refinement is achieved by “seeding” the system with the hypothesis and the evaluation function for fitness. The algorithm operates through a simple cycle:

1. Creation of a population of strings.
2. Evaluation of each string.

3. Selection of the best strings.
4. Genetic manipulation to create a new population of strings.

4. Implementation and evaluation

To investigate the effectiveness of the proposed method for alert correlation of intrusion and risk assessment of the system. We perform some experimental task, all these tasks perform in MATLAB7.8.0 software and well famous intrusion data set kddcup99 provided by DARPA agency.
UFP-Growth and Genetic Algorithm have been proposed for risk calculation and alert correlation rate. UFP-Growth algorithm is used for association rule mining and Genetic Algorithm is used for finding optimal pattern. Using UFP-Growth algorithm correlation is performed on different attack. Once correlation is performed risk level is calculated for each intrusion. After that alert correlation rate is determined, there is also possibility of positive correlation rate and negative correlation rate, and these parameter values are also computed. Here we have taken KDDCUP 99 dataset for experimental process, which contain four different types of attack and one normal. These dataset is used for simulation process and results should be computed.
In the form of results we calculate the four parameter:
Risk Calculation: This is the quantitative value of risk which shows the risk factor of intrusion. On the basis of the risk factor alert messages are generated.
Alert Correlation Rate: This is rate of correlation when risk factor is computed the alert message should be generated.
Positive Correlation Rate: when normal data is treated as intrusion and risk level is computed of that data this is called as positive correlation rate.
Negative Correlation Rate: When intrusion or malicious data can be treated as normal data and the risk level is computed of this data, called as negative correlation rate.
Result should be calculated using both methods one was existing method (ACR) and other one is proposed (UFP-GA). Our method shows the promising results. As shown in table 1 and table 2.

Table 1- Simulation results based on ACR

Method	ACR			
Risk Level Probability	Risk Calculation	Alert Correlation Rate	Positive Correlation Rate	Negative Correlation Rate
0.1	89.27	3.85	2.45	3.25
0.2	90.96	5.55	4.15	4.95
0.3	89.27	3.85	2.45	3.25
0.4	91.00	5.59	4.19	4.99
0.5	91.19	5.77	4.37	5.17
0.6	89.27	3.85	2.45	3.25
0.7	91.24	5.83	4.43	5.23
0.8	91.33	5.91	4.51	5.31
0.9	89.27	3.85	2.45	3.25

Table 2- Simulation results based on UFP-Growth & Genetic Algorithm

Method	UFP-Growth & Genetic Algorithm			
Risk Level Probability	Risk Calculation	Alert Correlation Rate	Positive Correlation Rate	Negative Correlation Rate
0.1	95.27	2.85	1.95	2.45
0.2	96.96	4.55	3.65	4.15
0.3	95.27	2.85	1.95	2.45
0.4	97.00	4.59	3.69	4.19
0.5	97.19	4.77	3.87	4.37
0.6	95.27	2.85	1.95	2.45
0.7	97.24	4.83	3.93	4.43
0.8	97.33	4.91	4.01	4.51
0.9	95.27	2.85	1.95	2.45

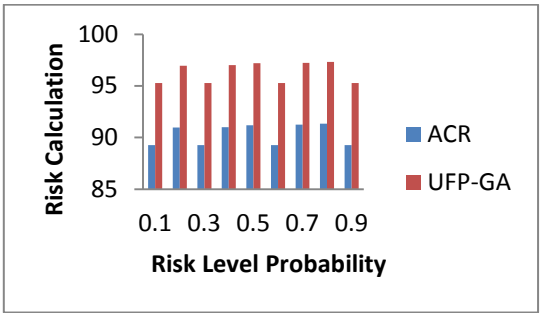

Fig.1 Risk Calculation using ACR and UFP-GA

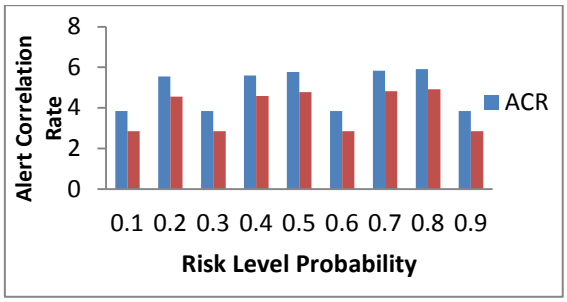

Fig.2 Alert Correlation Rate using ACR and UFP-GA

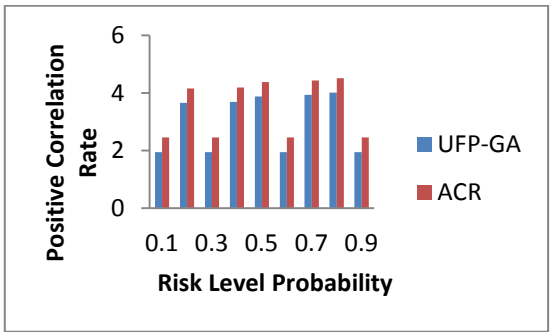

Fig.3 Positive Correlation Rate using ACR and UFP-GA

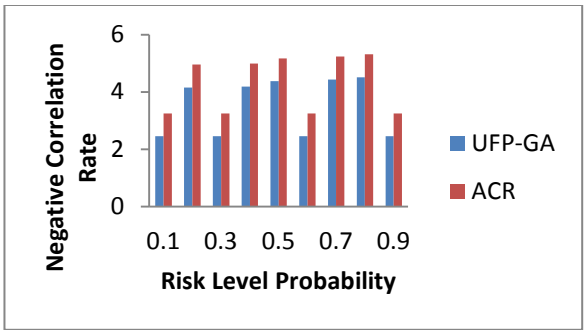

Fig.4 Negative Correlation Rate using ACR and UFP-GA

Our experiment shows the promising result as compare with earlier approaches.

5. Conclusion

We proposed a new method for extracting the most critical and useful information in alert sequences that can be generated by an IDS. We used novel approach to obtain all of different combinations of alerts then correlate the attacks for minimizing the false alarm rate. Proposed alert correlation approach assists administrators to come aware of the most critical alerts in the alert set, based on the specification of their network ie. Risk calculation and alert correlation rate. So, administrators will be able to discriminate critical attack patterns that are almost 10 through 30 percent of all of patterns. In future we improve the value of risk calculation and minimized alert correlation rate.

References

[1] Fayyad U,Piatesky-Shapiro G,Smyth P. The KDD Process for Extracting Useful Knowledge Form Volumes of Data Communications of the ACM,1996.

[2] W. Lee, S. J. Hershkop, P. K. Chan, E. Eskin, W. Fan, M. Miller, S. Hershkop and J. Zhang, "Real Time Data Mining-based Intrusion Detection", In Proc. of the DISCEX II 2001. Anaheim, Vol. 1, pp. 89-100, 2001.

[3] D. Parikh and T. Chen, "Data fusion and cost minimization for intrusion detection", IEEE Trans. on Information Forensics and Security, Vol. 3, No. 3, pp. 381-389, 2008.

[4] Wang Yingmei, Wang Shengkai and Cheng Xiangyun, Security Risk Assessment of Information System, Publishing House of Electronic Industry, Beijing, 2007.

[5] K. Julisch and M. Dacier, "Mining intrusion detection alarms for actionable knowledge", Proceedings of the 8th ACM International Conference on Knowledge Discovery and Data Mining, July 2002, pp. 366-375.

[6] A. Valdes and K. Skinner, "Probabilistic alert correlation", Proceedings of the 4th International Symposium on Recent Advances in Intrusion Detection (RAID 2001), 2001, pp. 54-68.

[7] F. Cuppens, F. Autrel, A. Miège, and S. Benferhat "Correlation in an intrusion detection process", Internet Security Communication Workshop (SECI'02), September 2002, pp. 153-172.

[8] F. Cuppens and A. Miège, "Alert correlation in a cooperative intrusion detection framework", 2002 IEEE Symposium on Security and Privacy, May 2002, pp.202-215.

[9] Li Yang and Dong Xinfa "Alert Correlation Model Design based on Self-regulate" in Second International Conference on MultiMedia and Information Technology IEEE, 2010.

[10] Jin SHI, Guangwei HU, Mingxin LU and Li XIE "Intrusion Alerts Correlation Based Assessment of Network Security" in International Conference of Information Science and Management Engineering IEEE, 2010.

[11] Lu Simei, Zhang Jianlin, Sun Hao, Luo Liming "Security Risk Assessment Model Based on AHP/D-S Evidence Theory" in International Forum on Information Technology and Applications IEEE, 2009.

[12] Alter, S., Sherer, S.: A general, but readily adaptable model of information system risk. Communications of Association for Information Systems,14 (2004), 1-28.

[13] Sun, L., Srivastava, R. P., Mock, T. J.: An Information Systems Security Risk Assessment Model under Dempster-Shafer Theory of Belief Functions Journal of Managem.

Anand Jawdekar receive BE degree in the discipline of omputer Science & Engineering from Maharana Pratap College of Technology, under RGPV Bhopal in 2007. He currently pursuing Masters from LNCT, Bhopal.