# Intrusion Alert Correlation based on UFP-Growth and Genetic Algorithm

Anand Jawdekar
PG Research Scholar, RGPV
Bhopal, India

Vineet Richariya
Professor, LNCT
Bhopal, India

## ABSTRACT
Intrusion alert correlation is subject to assessment of security and risk level of quantitative analysis of security threats. Intrusion alerts correlation, especially the quantitative characterization of network security and the approach of the build and update of network security scenario and measurement, is one of the important basic approaches of building security services based on the correlation. Various author proposed a model for security analysis of intrusion alert correlation such as Assessment of Credibility, Risk and the Loss of system (ACRL). In this method the correlation value of intrusion find the way of credibility and risk. Some another approach are also used such as graph theory approach for the analysis of node behavior in attack scenario. In this paper we proposed a new algorithm for intrusion alert correlation based on uncertain FP-growth and genetic algorithm. Uncertain FP-growth finds the possibility of probability in attacks occurred before events and mange by the security policy manger. In the process of correlation various value of quantitative are generated some value are exactly correlated and some are low value of quantitative. For the measurement of low value of risk correlation we used genetic algorithm for the optimization process of risk level.

## General Terms
Intrusion, genetic algorithm, frequent pattern growth.

## Keywords
Intrusion Alert correlation; Uncertain FP-growth & Genetic Algorithm; KDDCUP1999; risk calculation; Intrusion detection; etc.

## 1. INTRODUCTION
Study about intrusion detection system started earlier.[1] However current intrusion detection system faces some problems. Such as quality of attack, relevance of alert information,conceivedof the study of alert information. According to it is required to make correlation and adjustment analysis on IDS alert information, to reduce false alert rate and minimized risk level of security.

Using alert correlation technique security manager maintain the security of the system. In the current scenario security of computer system is always at risk. With growing internet network intrusion has became a critical component of infrastructure protection mechanism. Intrusion detection system can be defined as a set of action taken against any malicious code effected security parameter such as confidentiality, authentication, integrity etc. The conventional intrusion detection mechanism can be classified as misuse and anomaly detection, but these method tends to high false alarm rate.[2][3]

In recent years, risk arises from information system and networks from time to time. As its impact is increasing, persons realizes the valid single step solution to these threats.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].

Intrusion alert correlation is a network security assessment method. Alert correlation is a important technique in intrusion detection. It analyzes the alerts from one or more intrusion detection system and provide the succinct view of the system, and reinforce the present state of the system. Intrusion alert correlation [5-8] is basically used to reduce the overhead increases in intrusion detection system. It improves preciseness of the intrusion detection system for reducing positive as well as negative false alarm rate. Intrusion alert correlation is extracting the true alerts from meta alerts using establishing relationship among each 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 gives the quantitative values of security parameters (risk calculation, alert correlation rate, positive correlation, and 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.

## 2. RELATED WORK
In the following we summarize some of the recent research works in the area of alert correlation and risk assessment.Li Yang and Dong Xinfa [9] describe Alert Correlation Model. In this work author focus on multi-step attack, this is primary form of network intrusion. Here pattern recognition is used to recognize the patterns. Alert correlation model is proposed which was based on self regulate. This model classifies the alert information and derives the association rules. Apriori algorithm used for predicting the behavior of the data and detect intrusions.

Jin SHI, Guangwei HU, Mingxin LU and Li XIE [10] describe ACRL approach (assessment of credibility, risk and loss). Here correlation can be done on the basis of the credibility of the attacks. Here ACRL correlate the attack

group on the basis of the credibility and find out the risk factor of the system and loss. Here some quantitative values are generated which was useful for security manager to prevent the state of the system. 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 existing 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. AHP and D-S used in combined to solve these issues.Jung Huang, Chin-Fa Lin, Ching-Yu Li, Jia-Jian Liao, Yu-Wu Wang, Kai-Wen Hu[12] describe intrusion detection system is a security layer that is used to prevent intrusive activities. Here intrusion detection module can be tuned continuously to provide the accurate and succinct view of the system.

# 3. PROPOSED WORK

## 3.1 UFP-Growth and Genetic Algorithm for intrusion alert correlation

We proposed a new algorithm for intrusion alert correlation based on uncertain FP-growth and genetic algorithm. Uncertain FP-growth finds the possibility of probability in attacks occurred before events and mange by the security policy manger. In the process of correlation various value of quantitative are generated some value are exactly correlated and some are low value of quantitative. For the measurement of low value of risk correlation we used genetic algorithm for the optimization process of risk level.

The alerts fulfilling some situations are aggregated into one Meta alert. For example, a situation can be defined as: the alerts which come from the same source and with the same attack class within a period of time. This process can largely reduce the amount of alerts without any loss of security information. We employ an algorithm to generate correlation sequences. A genetic algorithm has been to optimized the alerts and eliminate the alerts which are irrelevant to the desired attack patterns. First the alerts are modeled by a tree using UFP-growth, where each node represents an IP address, and each directed edge represents the alerts from the source IP to the destination IP. Then according to the different sources and destinations, the alerts are divided into several group. Each possible attack would fall into certain group such as g1,g2,…..gn. According to the algorithm, one can choose the alerts in a connected-tree within every period of time to be a candidate sequence for rule mining. Meanwhile, one can also delete some alerts to improve the efficiency of the mining algorithm, e.g., discard the alerts in a tree which only corresponds to certain attack rule if the correlation is considered to be irrelative value of risk, or just select the alerts in the tree which are related to risk level.

The proposed algorithm is combination of UFP-tree algorithm and genetic algorithm. Here UFP-tree algorithm generates a input sequence for genetic algorithm. Basically input sequence is nothing it's a group value of attack predicted by our process. Now for the better prediction of risk assessment we optimized a group value of risk correlation. For the optimization of group correlation value we used genetic algorithm. Here genetic algorithm minimized a expected value of support and confidence. Some steps of algorithm is discuss here.

Input: A set of attack group sequence of data

Output: The rule set of attack scenario U kLk, and U k( H (Lk)) (k=1,2,...,n- 1) and input of genetic algorithm for optimization of rule set.

**Algorithm:**

1. Find the value of expected support and confidence such as Since existence of each attack Xin a dataset Di is captured by its existence probability P(X, Di )Support of an attack group S in a dataset Di is the expected of coexistence of all the items in S, i.e.

$$\prod x \in S \ P(X, Di)$$

The value of expected support of S in Di is (0, 1)Support of an attack S in a dataset can be obtained by summing over all dataset the expected probability of S, i.e.,

$$\sum i[\prod x \in S \ P(x, Di)]$$

2. Find all the large 1-attack: Lj=tlarge 1-attacks}; once getting the large (k-l)-attack Lk 1, get the large k-attack Lk;

3. Given 1i E Lk-l, find the sequence set Pk={P2-P2. I C Lk, where Vp EE Pk 1i is the beginning subsequence of pj; zT(l ) = Count(Pk )Icount(li) H(Lk 1) = {zr(li),li E Lk 1};

4. finally, input of genetic U kLk, and U k(f (Lk)) (k=1,2,...,n-1)

5. Step 1: Input group of attack sequence of data X1,X2…….Xn , the rule number Rn , population scale XN, crossover probability cP , mutation probability mP , vaccination probability vP , stop conditions cS ;

Step 2: Code the chromosome in real number and initialize population A(i),i = 0 at random;

Step 3: Calculate the fitness of each individual in the current instant;

UFP-tree needs optimization of expected support and confidence for generation of rule based tree, Hence the fitness function of algorithm is determined by f(x).

F(x) = {(α +2β)-αi, αi<β+2α

0, αi≥αi+2β

I=1,2,……..,N

Step 4: Judge the termination conditions. If the termination conditions are satisfied, then turn to step 5, otherwise, turn to step 6;

Step 5: Decode to find and calculate the optimal correlated risk matrix. And set the optimal risk according to maximum output the results.

Step 6: Exit

# 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 |

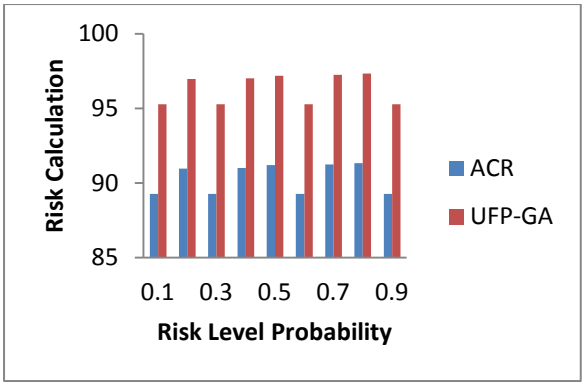


**Fig.1: Risk Calculation using ACR and UFP-GA**

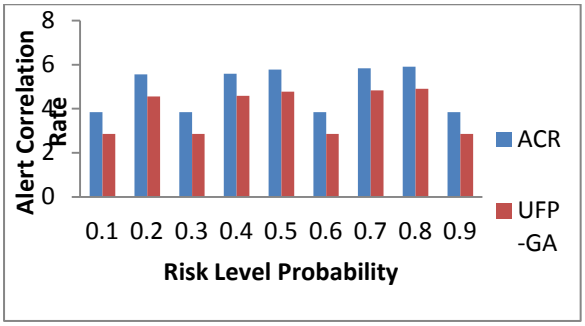


**Fig.2: Alert Correlation Rate using ACR and UFP-GA**

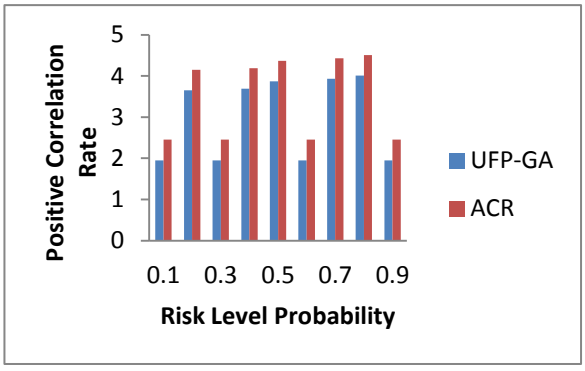


**Fig.3: Positive Correlation Rate using ACR and UFP-GA**

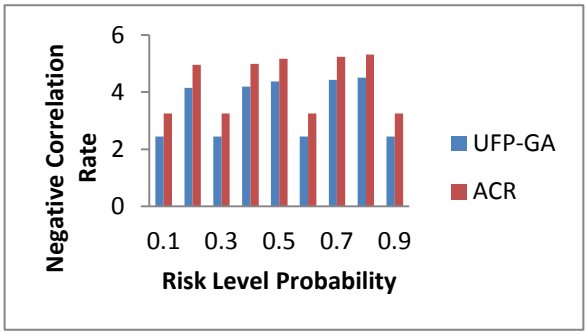


**Fig.4: Negative Correlation Rate using ACR and UFP-GA**

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

## 5. CONCLUSION

This work processes a model of intrusion alert correlation based on uncertain FP-growth and genetic algorithm. Uncertain FP-growth finds the possibility of probability in attacks occurred before events and mange by the security policy manger. Forthe measurement of low value of risk correlation we used genetic algorithm for the optimization process of risk level. In our approach the attack group is generated by UFP-growth and during the process of mining the predictability of every security state in the attack group can be estimated.. We employ an algorithm to generate correlation sequences. A genetic algorithm has been to optimized the alerts and eliminate the alerts which are irrelevant to the desired attack patterns. Here genetic algorithm was used to optimize the low value risk this algorithm suffer from larger values of data, so in future we use multi objective function algorithm such as swarm intelligence.

## 6. REFERERNCES

[1] Fayyad U,Piatesky-Shapiro G,Smyth P. The KDD Process for Extracting Useful Knowledge Form Volumes of Data ommunications 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 SecurityCommunication 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] "An Adaptive Rule-Based Intrusion Alert Correlation Detection Method" in First International Conference on Networking and Distributed Computing IEEE, 2010.

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

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