The Information System Security Situational Awareness Based On Cloud Computing

Ma Zhicheng[1]
[1]Information and Telecommunication Company
State Grid Gansu Electric Power Company
Lanzhou, China
email: mazc@gs.sgcc.com.cn

Jin Lin[2],Yang Peng[1]
[2]New Energy Power System State Key Laboratory
North China Electric Power University
Beijing, China
email: linjinhpns@gmail.com

***Abstract*—With the increasing amount of data in the information network, various network threats are growing. Designing efficient and reliable security situational awareness methods becomes one of the main tasks of information security. By using the Apriori algorithm based on MapReduce in the cloud computing environment to knowledge discovery in the network security situational awareness, we can realize rapid security modeling and security situational generation of huge amounts of data. Firstly, the system uses D-S evidence theory to model security situation, streaming, filtering and integrating various types of security events. Secondly, we introduce MP-Apriori algorithm to knowledge discovery, from which association rules of security alarm events generate. Lastly, network security situation generates based on security situation generation algorithm. Then, we use KDD Cup 99 data sets to simulate experiment. We input equal amount of security warning events, verify the accuracy of the CC-SSA and compare the time of using four different computing nodes. The experiment shows that CC-SSA method for a network security situational awareness of huge amounts of data is superior in speed.**

Keywords-Cloud Computing, Apriori Algorithm, Security Situational Awareness, Data Dining, Knowledge Discovery

I. INTRODUCTION

Our network is facing the changing worms, large-scale attacks and other security threats. Traditional network security devices usually work in an independent manner. It can't find and use the relationship between the events accurately and efficiently, which leads to many uncertainties. The device has a high rate of false positives and false negatives. Network security situation awareness is an effective way to solve these problems. It provides network security's "Global View", assesses the overall security status of the network system and speculates future security trends.

The critical work of network security situational awareness is capturing and analyzing a large amount of security events which provided by distributed and heterogeneous sensors. The information is presented in an appropriate manner, so that managers can grasp complex and dynamic security posture quickly. However, the real network security situational awareness still has many difficulties, such as various alarm sensor generates excessive events and the false alarm rate is too high. How to improve the speed of analyzing warning message is the key issue of network security situational awareness.

This paper summarizes the recent progress1-8 and presents a security situational awareness method which is based on the cloud computing (CC-SSA, cloud computing - security situational awareness). This method meets the increasing data trend of the network. The network security situational awareness can complete more efficiently by this way.

II. NETWORK SECURITY SITUATION MODELING AND GENERATION FRAMEWORK

The network security situation modeling and generation framework has two parts, shown in Fig.1. First, the security situation modeling builds a formal model to measure network security situation. It is based on the D-S evidence theory and supports integration and processing of various types of security events from the security situation sensor. Then, we introduce the MR-Apriori algorithm (Apriori algorithm based on MapReduce) for the knowledge discovery. MR-Apriori algorithm can help us access security alarm event's association rules. Last, the network security situation graph generates dynamically based on the security situation generation algorithm.

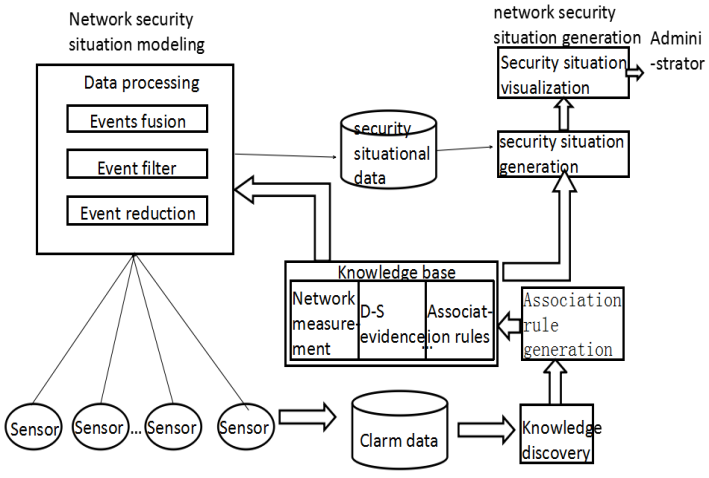

Fig.1 Network security situation modeling and generation framework based on cloud computing.

A. *Network Security Situation Modeling*

The main purpose of security situation modeling is to construct a data model which adapts to measure the network security situation. In the initial pre-processing stage, by normalizing the alarm event, we transfer all of the received security events into a standard data format that can be understood by the processing module .In the stage of processing the security situational data, we input the normalized sensor alarm events and streamline, filter and fusion them. The goal of event streamlining is to combine a series of redundant events detected by the sensor of the same attack. The target of event filter is to delete the event that does not meet the constraint requirements. Event fusion is based on the D-S evidence theory18 .It introduces different levels of confidence for the pretreated events. Then it integrates multiple properties to quantified evaluate the network alarm events.

B. *Network Security Situation Genaration*

a) *Association Rules Extraction Based on Knowledge Discovery*

We use the Apriori algorithm in the field of knowledge discovery. Apply it to the cloud computing for distributed computing, which can greatly improve the speed of analysis and get the situation knowledge from the security alarm events. The purpose of using the Apriori algorithm is to get the regularity among the event attributes. The regularity can be converted to the filtering rules associated with associated actions.

The core content of realizing Apriori algorithm based on MapReduce model is to find the key data of the original algorithm, then map these data to the Key's value and the Value's value of the MapReduce.

1) The improvement of frequent item sets statistics

In the process of MapReduce calculation, select item sets as the Key's value of this phase, the value is 1.By using the Map function, Hadoop framework divides the data set into several subsets. Distribute them to run and count on all nodes. Then use the Reduce function to count words and choose k frequent item sets. By this way, we can realize the parallel improvement of the transaction set's scanning process. The scan time will be greatly shorten.

2) Generate k items superset by grouping

In the process of generating a superset of k items, we definite the same k-1 items modes as the master mode. The frequent items with the same master mode were sent to the same Reduce as the Key's value. Each sub-model produced by the last model arrangement is defined as the generation mode base. We use the generation mode base as the Value's value. Then generate the generation mode by the Reduce.

3）Read the frequent item sets and the speed-up of set tailoring

When cutting the k+1 items superset for the k+1 candidate sets, it is needed to match k items for each item of the k+1 items superset. In the process of generating k items superset by using MapReduce model, every k items superset is all combined by the master mode and production mode. Therefore, if we gather all the superset of the same production pattern to one Reduce, we only need to read the subset ended by the production mode in the k items frequent set and tailor. The time and space required by reading k items frequent set can be greatly reduced. At the same time, since the read set is a subset of the frequent set, the compared tailoring time is greatly reduced.

The Apriori algorithm's Program flow chart based on MapReduce model is shown in fig. 2.

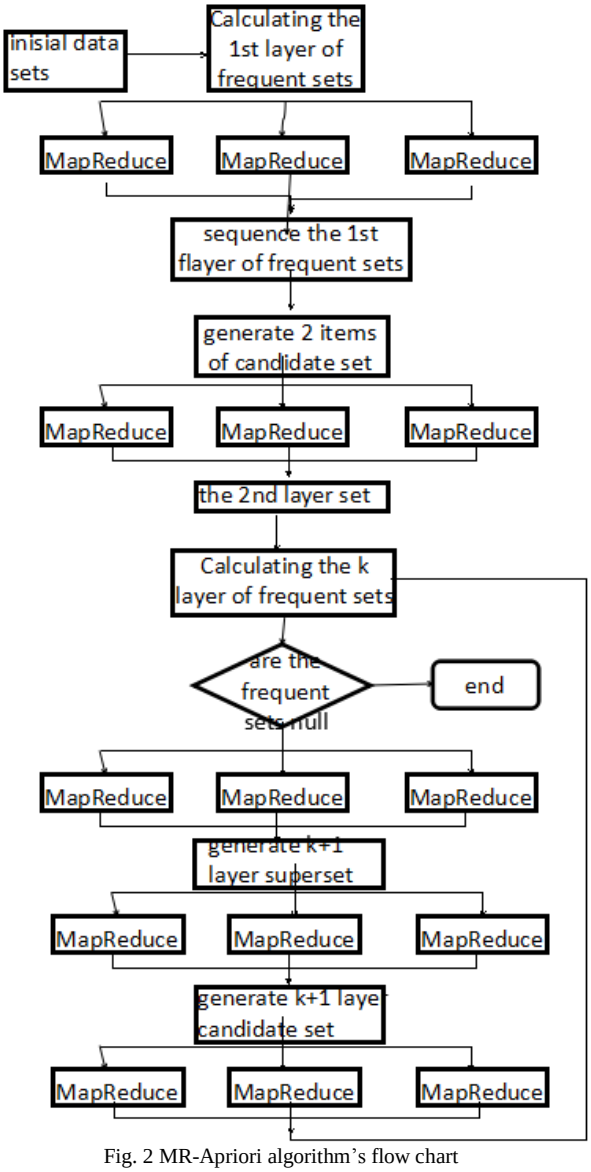

Fig. 2 MR-Apriori algorithm's flow chart

4) Produce association rules

Firstly, check each nonempty set d of the frequent item sets f and generate the corresponding rule d→ (f-d); then, calculating the degree of confidence support (f) ÷ support (d). When the ratio is greater than the minimum confidence, association rules generate.

Below is a parallel association rules' implementation process based on MapReduce model:

Mapper {

```
Map () {
//Each record is a frequent item sets/
d=f-1;
i=1;
While ( confidence(f,d)≥minisupport &&f>i+1){
i=i++;
Emit (d→f-d, confidence(f,d);
d=f-i
}}}
Reducer {
Reduce () {
Emit (key, value);
}}
```

b) Network Security Situation

We calculate according the network node's risk level according to the fused alarm event. We mainly consider the following factors: the alarm confidence level c , the alarm severity level s, influence of resources m, the security protection grade of nodes p_n 、 the alarm recovery coefficient r_n and so on.

The security situation assessment value's calculation of a single node n is as follows:

$$S_n(t) = \frac{\sum_{t_i \in T_t} c_i s_i m_i}{p_n r_n} \quad (1)$$

Since the positions and effect of the different nodes in the network are different, their important degree is different. So we must consider the node weights w_n.The calculation of entire network's security risk value is as follows:

$$S_N(t) = \sum_{n \in N} w_n S_n(t) \quad (2)$$

According to the above calculation, we can draw the curve of the security situation assessment values changing over time.

III. SIMULATION EXPERIMENT

A. *The Experimental Data Set*

The experiment using KDD Cup 99 data sets. It consists of 5 million records, containing the "Normal" type behavior records and "DoS" and "R2L", "U2L" and "Probe" four kinds of threats.

B. *Experimental Results and Analysis*

Firstly, we fuse and associate the security alarm events detected by the security situation sensor by using CC-SSA and generate the network security situation, which verify the security situational ability of CC-SSA. Then, in the case of a consistent input data sets, using 5, 10, 15 computing nodes, regardless of time window, test the time of the whole process of safety awareness, which verify the efficiency of CC-SSA.

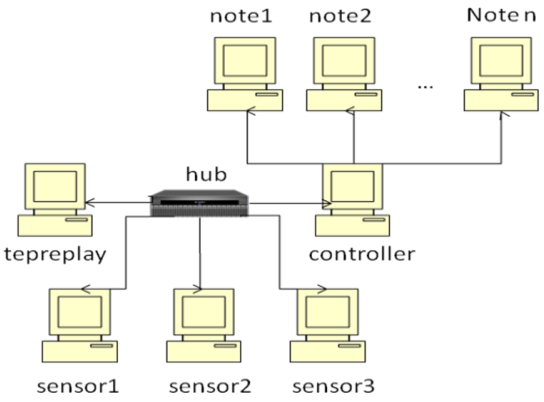

Fig.3 Laboratory network environment configuration diagram

The network environment configuration is shown in Fig. 3. The sensor1, sensor2 and sensor3 are the security situation sensors, used to collect the security alarm event in the network experiment. The controller and note1 to note n compose a small cloud computing environment. The controller is the master side. It is responsible for assigning tasks of correlation analysis security alarm events reported by sensors to every note, merging of the results obtained from each node for the relationship and showing the network security situation with the graphical interface. The tepreplay is used for the playback of the test data set.

When receiving an important alarm event, the system do integration analysis. According to the algorithm steps in section 2.2, the system analysis the alarm events and generate the security situation.

Security situation assessment can objectively reflect the value of the security status of the host and the host under attack. For each host under attack of the data set, the security situation evaluation value is calculated as follows:

$$S_{host_t}(t) = \frac{\sum_{t_i \in T_t} c_i s_i m_i}{p_n r_n}$$

Achieved the security situation assessment value of all the target host, we can calculate the entire network's security situation assessment value with the node weights:

$$S_N(t) = \sum_{n \in N} w_n S_n(t)$$

Usually, the higher the network security situation assessment value, the lower the entire network's security is, the more serious be attacked. We implement simulation attack many times in the experiment environment. By timing calculate S_N using CC-SSA, it accurately reflects the dynamic change of network security situation, as shown in Fig. 4.

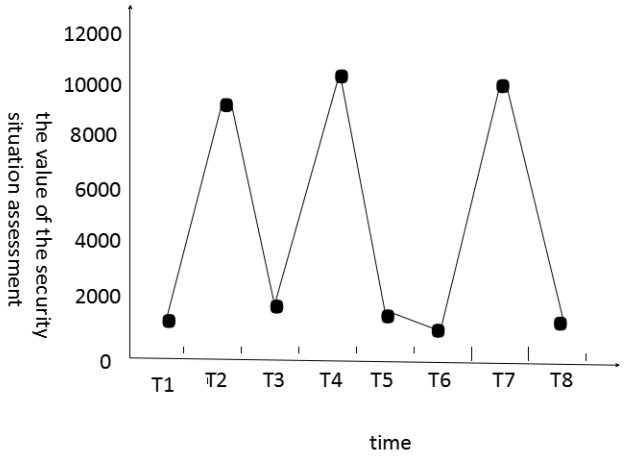

Fig.4 Network security situational change curve

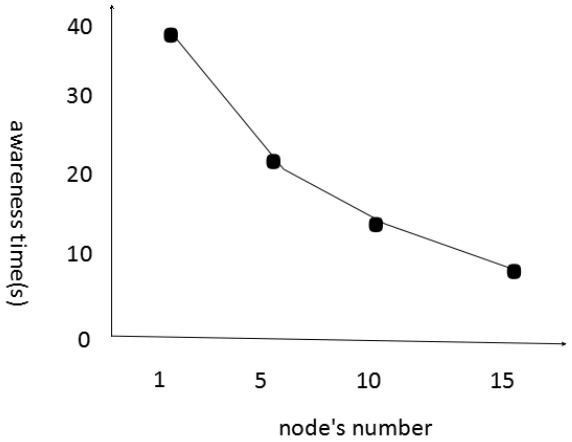

Fig.5 awareness time-node's number curve

According to the above methods, input the KDD Cup 99 data set, respectively use 5, 10, 15 compute nodes for correlation analysis， calculate the security situation after all data sets has been input in one time, then record the time used, as shown in Fig. 5. It can be seen that with the increasing number of computing nodes, the time of security situational awareness will be greatly reduced. Cloud computing makes situational awareness efficiency greatly improved in the case of a large amount of data.

IV. CONCLUSIONS

In this paper, we proposed the network security situational modeling and generation framework. It supports accurate modeling and efficient generation of the network security situation. Experiments show that this system possesses the advantages of high efficiency. With the rapid increasing of the amount of data in the future, its advantage will be greatly highlighted. The next step of the research work is the study of real-time prediction technology of major security attack.

REFERENCES

[1] Lai J B, Wang H Q, Zhu L. Study of Network Security Situation Awareness Model Based on Simple Additive Weight and Grey Theory[A]//Proceeding of 2006 International Conference on Computational Intelligence and Security[C].vol2,Guangzhou, China,2006:1545-1548

[2] Liu M X,Zhang Q Y,Zhao H,et al.Network Security Stuation Assessment Based on Data Futon[A]//2008 Workshop on Knowledge Discovery and Data Mining[C]. Adeaide, Austrilia, 2008:542-545

[3] Wang H Q, Lai J B, Ying L. Network Security Situation Awareness Based on Heterogeneous Multi-sensor Data Fusion and Neural Network[A]//Second International Multisymposium on Computer and Computational Science[C]. Iowa, USA, 2007: 352-359

[4] Siraj A.A Unified Alert Fusion Model for Intelligent Analysis of Sensor Data in an Intrusion Detection Environment[D].Mississippi, USA: Faculty of Mississippi State University, August 2006

[5] Hu W, Li J H, Shi J J. A Novel Approach to Cyberspace Security Situation Based on the Vulnerabilities Analysis [A]// Proceedings of the 6th World Congress on Intelligent Control and Automation[C]. vol 1,Dalian,China, 2006:4747-4751

[6] Zhang Y, Tan X B, Xi H S. A Novel Approach to Network Security Situation Awareness Based on Multi-perspective Analysis [A] //IEEE 2007 International Conference on Computational Intelligence and Security[C]. Harbin,China,2007:768-772

[7] Chen X Z, Zheng Q H, Guan X H, et al. Quantitative hierarchical threat evaluation model for network security [J]. Journal of Software,2006,17(4):885-897

[8] Han J W, Pei J, Yin Y W. Mining frequent patterns without candidate generation[J].ACM SIGMOD Record,2000,29(2):1-12

[9] Mika K. A knowledge discovery methodology for telecommunication network alarm databases [D]. Finland, University of Helsinki,1999

[10] Hettich S , Bay S D . KDD Cup 99 task description[EB/OL]. 1999. http://kdd.ics.uci.edu/databases/kddcup99/task.html.