Proceedings of the 4th WSEAS Int. Conf. on Information Security, Communications and Computers, Tenerife, Spain, December 16-18, 2005 (pp59-62)

Arithmetic of Security Data Fusion under High Performance Computing Environment

XIAO HAIDONG, LI JIANHUA
School of Electronic, Information and Electrical Engineering
Shanghai Jiaotong University,
Shanghai, 200030
CHINA.

Abstract: - Research of computing security is a hotspot recently. In this paper, data fusion technology is used to analyze the security situation under the high performance computing environment. A unique efficient method is given to analyze the whole systems' security situation. With the result, complex security data fusion under high performance computing can be solved in an easy way.

Key-Words: - high performance computing, data fusion, security

1 Introduction

The computing requirement of human is developed towards the new targets of high performance, multiformity, multi function, many cosmically science application is not only constructed with one high performance computer, but also need more virtual super computers based on grid technology. They are consisted by many machines; cooperate with each other, connected with many science instruments. The target is to solve the computing problems we faced. The best advantage of these system is that we can share the computing resource such as computing resource and data resource on the every node[1], but the base of sharing system is that must be constructed on the safely access.

To solve the security problems under the high performance computing environment, IDS (Intruding Detecting System) will process the data fusion in the isomerous distributing networks, and form the security situational awareness.

2 Security problem under the high performance computing environment

The difference between high performance computing application and traditional client/server application is: it needs substantive resource, and the resource requirement is dynamic, its commutation architecture is more complex, and the performance requirement is stricter, etc.[2]

The high performance computing system and application require all of the stand security function, include authentication, access control, integrality, privacy and resist deny. We often discuss the authentication and access control. Especially include: (1) provide the authentication solve scheme, which allow users, users' computing process included also, to testify the identity each other by the resource they used in these process. (2) Never change the control mechanism at any time. Based on this authentication, the security policy is constructed, and forms all the local security policy into a whole security framework.

To develop this security architecture, the follow limitations should be satisfied:

Credence protect: the credence of users (such as pass words, private key, etc.) must be protected.

Real time security situation awareness: as we know now, security situation knowledge distilling model is constructed with hierarchical security risk knowledge with three levels: object refinement, situation refinement, and threat assessment. How to fusion the security data to satisfy the real time security situation awareness requirement, we will try as follows.

3 Analyze the high performance computing security situational data based on Knowledge

In common application, the data and object refinement often be processed in the complex distribute data circumstance, and can not be accessed freely, Kargupta give a influx data mining structure, it analyze the part data with the orthogonal basis function [3], it provides solution to the problem which is we can't construct the whole situation data model only with the part data analysis. We will use this method to mine the cyberspace security data.

The follow example is supervised inductive learning, it shows the distribute knowledge discovery process of the security situation.

Basement function $f: X^n \to Y$ generate the dataset $\Omega = \{(x_1, y_1), (x_2, y_2) \cdots (x_k, y_k)\}$, learning function is $\hat{f}: X^n \to Y$, $\hat{f}$ is approximate function to f .Any item of range set $X = \{x_1, x_2, \cdots, x_n\}$ is a n N-tuple, x_j corresponding to the dimension of this region. A set of primary functions is constructed below:

$$f(x) = \sum_k w_k \Psi_k(x) \quad (1)$$

$\Psi_k(x)$ is the number k primary function, w_k is the corresponding coefficient, the purpose of arithmetic is to generate the approximate function.

$$\hat{f}(x) = \sum_k \hat{w}_k \Psi_k(x) \quad (2)$$

In above function, $\hat{w}_k$ is the estimate value of coefficient w_k. Different learning arithmetic uses different primary function.

When the primary functions are selected, we can evaluate their coefficient. If the training data set is S, we use the distribute data in the S to evaluate. If the security situation data character space is divided orthogonally, suppose the two data point A and B, corresponding to the set $S = \{S_a, S_b\}$, there, $S_a = \{(x_{(a,1)}, y_{(1)}), (x_{(a,2)}, y_{(2)}), \cdots, (x_{(a,k)}, y_{(k)})\}$. Data set $x_{(a,i)}$ is the number is situation data set, and only has a situation character x_a. Suppose that y_i is the number I classified situation dataset identifier, and this method can be used to all the data. Because dataset S_b is not the local data of data A, checking their primary functions and evaluating their coefficients only can be done based on the local situation characters of data A. for the same reason, data B has the same problem. If A and B don't exchange their data, their characters about primary function can't be evaluated. With the example of orthogonal basis function below, it form is:

$$\hat{f}(x) = \sum_i w_i x_i + \sum_{i,j} w_{i,j} x_i x_j + \sum_{i,j,k} w_{i,j,k} x_i x_j x_k + \cdots$$

If x_1 and x_2 belong to the data A and data B separately, if the common information of these two data can't be provided, we can't calculate $x_1 x_2$; any way, if x_2 and x_3 belong to the same data, the primary function calculation of $x_2 x_3$ will be done. This example shows that the primary function and their coefficient are decided by the whole situation character space divide method.

According to the formula (1) and (2):

$$f - \hat{f} = \sum_k (w_k - \hat{w}_k) \Psi_k(x)$$

That is

$$(f - \hat{f})^2 = \sum_{j,k} (w_j - \hat{w}_j)(w_k - \hat{w}_k) \Psi_j(x) \Psi_k(x)$$

Sum them all on the training dataset Ω:

$$\sum_{x \in \Omega} (f - \hat{f})^2 = \sum_{j,k} (w_j - \hat{w}_j)(w_k - \hat{w}_k) \sum_{x \in \Omega} \Psi_j(x) \Psi_k(x)$$

In according to that all the primary functions is orthogonal, when all the x are considered in the cyberspace, another condition is $j \neq k$,

$$\sum_{x \in \Omega} \Psi_j(x) \Psi_k(x) = 0$$

We get

$$\sum_{x \in \Omega} \Psi_j(x) \Psi_j(x) = 1$$

and .random variable is defined as $Z_i = \Psi_j(x_i) \Psi_k(x_i)$, so when $j \neq k$, $E[Z_i] = \sum x_i \Psi_j(x_i) \mathrm{A} \Psi_k(x_i) = 0$, according to the law of great number. When n is very big:

$$\sum_{x \in \Omega} (f - \hat{f})^2 = \sum_j (w_j - \hat{w}_j)^2 \quad (3)$$

We can get that with all the j, when $\hat{w}_j = w_j$, the sum of square difference is minimum.

Towards those cyberspace security situation characters divided orthogonal, solution is different, suppose the characters space can be divided into two parts as A and B, named S_a and S_b. Suppose F is the set of all primary function, F_a and F_b is the primary function set defined with the security situation character variables in the set of S_a and S_b, F_{ab} is the set in the F, and it use the primary functions which are defined in the S_a and S_b contemporary. So, $F = F_a \cup F_b \cup F_{ab}$. Below we consider complexion that every data set only uses its local character variable to construct the learning function $f(x)$:

$$\hat{f}_a(x) = \sum_{j \in F_a} \hat{w}_j \Psi_j(x) \quad (4)$$

According to the equations of (1) to (4)

$$(f(x)-\hat{f}(x))^2 = \sum_{i,j\in F_a}(w_j-\hat{w}_j)(w_i-\hat{w}_i)\Psi_i(x)\Psi_j(x) + \sum_{i\notin F_a, j\in F_a} w_i(w_j-\hat{w}_j)\Psi_i\Psi_j + \sum_{i\notin F_a, j\notin F_a} w_f(w_i-\hat{w}_i)\Psi_i\Psi_j + \sum_{i\notin F_a, j\notin F_a} w_j w_i \Psi_i\Psi_j$$

With the support of law of great number, we get:

$$\sum_{x\in\Omega}(f(x)-\hat{f}(x))^2 = \sum_{i\in F_a}(w_i-\hat{w}_i)^2 + \sum_{j\notin F_a} w_j^2 \quad (5)$$

When $\hat{w}_j = w_j$, the equations above get the minimal value $\sum_{j\notin F_a} w_j^2$. Although the error is not zero, the w_i is the optimize solution. Even if the entire cyberspace security situation is considered, the result is still right in the whole cyberspace security context.

4 Construction of security situation apperceiving model

Computer networks are usually protected against attacks by a number of access restriction policies that act as a coarse grain filter. Intrusion detection systems (IDS) are the fine grain filter placed inside the protected network, looking for known or potential threats in network traffic and/or audit data recorded by hosts.[4]The typic security situation apperceiving model should be constructed as follow figure shows:

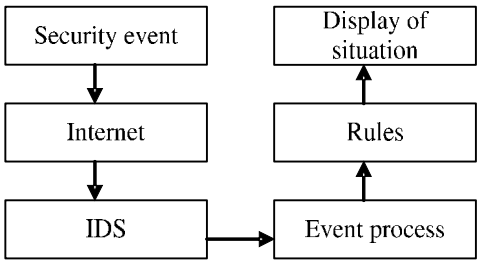

Fig. 1 security situation apperceiving model

The model has the function of detecting the networks' intrusion and collecting the data of networks security situational estimating.

Data of networks security event attrack is transfered to the acceptor of IDS through the networks, because the exist of the false alarm of IDS, the resource data is disturbed, this radom security false alarm can be described as a noise distrbution to the security situation system, and it makes the characters of security situational signal distorted.

How to solve this problem? We can do as follow steps:

1. Processing the security event association and security characters distilling, transforming;
2. Switch the situational variable to the other observation space, in this space, the security characters will be defined as observation vectors R(t), all attributes of security characters can be denoted as every dimensions of this vector;
3. To inspect the R(t) and evaluate the parameters, the results will be used to display the security situation or do the futrue data fusion processing.

To process this mathatic model, the classic static methold is helpful to describle the inspectability of IDS security event and base characters of processing evaluation.

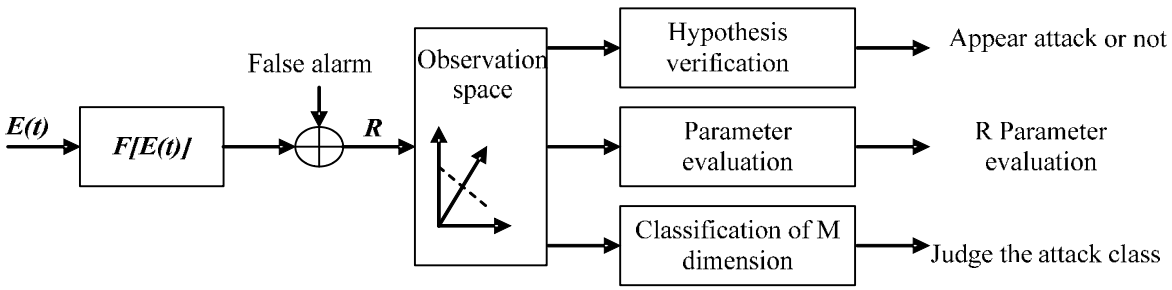

Fig. 2 mathatic model to describle the inspectability of IDS security event and base characters of processing evaluation.

F[E(t)]transfer the event E(t)from K dimension event detecting space to the N dimension observation space, together with the false noise to form the inspectablity data resource parameter, by processing the data with static methold, the secuity situation measurement of observation system is got:

1. Probability of false alarm;
2. Error distributing of state evaluation;
3. Precision of attack classification.

And network situational measurement:

1. Virus situation parameters;
2. Attack threaten parameters;
3. Host abnormal situation parameters;
4. Intensity of security threaten;
5. Frangibility of system;
6. Major threaten of attack and type of virus

4 Conclusion

The security requirements of grid computing system presented by I. Foster and other people in recent years,[5] but security data fusion under high performance computing environment not mentioned, so the problem of security situation evaluation not be solved, The complexion of security problems of high performance computing environment make us can not solve the problems exist in the whole cyberspace security situation learning. This is the shortage of security situation evaluation system in existence, but with the technique discussed in this paper, optimized solution of situation observation characters can be got by knowledge learning of part cyberspace security situation indirectly. In the future, we will design partition of computing granularity of security data fusion under high performance computing environment.

References:

[1] Haidong Xiao, Xinghao Jiang, Jianhua Li, *Grid Node Computing pool Security Analysis based on Knowledge Base*, Proceeding of CCICS'05

[2] Xu Defa, *The Security Problems and Solution of Grid Computing*, http://www.ssc.net.cn/xslw/aqwt.htm

[3] Mukherjee,., Heberlein,L., and Levitt, K., Network intrusion Detection, *IEEE Network Magazine*, Vol.8. No.3, pp26-41, May/June 1994.

[4] Giorgio Giacinto and Fabio Roll, Intrusion Detection in Computer Networks by Multiple Classifier Systems, Pattern Recognition Proceedings. 16th International Conference ,2002, pp:390 - 393 vol.2

[5] I. Foster, C. Kesselman, G. Tsudik, S. Tuecke. A Security Architecture for Computational Grids. *Proc. 5th ACM Conference on Computer and Communications Security Conference*, pp. 83-92, 1998.