4th WSEAS Int. Conf. on COMPUTATIONAL INTELLIGENCE, MAN-MACHINE SYSTEMS and CYBERNETICS Miami, Florida, USA, November 17-19, 2005 (pp130-133)

Knowledge base based Analysis of Security Situation of Networks

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

Abstract: Analysis of security situation of networks is a hotspot in information security research field. First of all, this paper describes the model of networks security situation knowledge refining of networks, then related knowledge base theoretic is given to analyze, finally a unique efficient method is given to analyze the whole networks' security situation.

Keyword: Network Security; Situation Analysis; Knowledge Base

1 Introduction

Today, technology of networks attacks becomes more and more complex and developed, and the security of network research becomes a hotspot too[1]. In order to defend the nation from the information security threat and prepare for future information war, many countries make great efforts to do this task. After the 911 event, American Native Security Department and Nation Key Base Establishment Protection Committee come into existence the Write House support. The vital function of this organization is to protect the nation from networks security attack by terrorist. Although such this huge and particular information security architecture is constructed, the status of information security of America is not improved. In fact, sensitive and security information exists on thousands and millions computer of America government and military, but part of them connect with Internet, and easy to be attacked by hackers. Another example to show the security situation is the Israel networks security system is not the best in the world, but the research of information security attack and defend technology is very developed, and keeps ahead in recent years. After the cosmically attacking by hackers in 1999, Japan government recognized the importance of information security. All the examples show us the networks security situation is not optimism.

2 Networks security situation

The Internet is mainly constructed with two infrastructures: Routers construct the backbone of Internet, DNS servers parse the domain name as IP address. The mainly direct attack method of Internet backbone networks is to attack the routers and DNS servers on the backbone. To solve these problems in the next generation cyberspace security solution, the conception of cyberspace situational awareness is introduced, the security data will be fused from several parts in the distribute networks system based on security cyberspace intrusion detecting system. But till now, the security evaluation work which is based on the information of intrusion event log is still stagnate on the single event which will bring the threaten to the system security. A good example is Tim's idea, which is to construct the cyberspace security situational infrastructure with the application of multi sensors data fusion[4], but it has

shortage that it can't give us the useful security situation information of whole cyberspace, and can't give help to networks managers to decide how to solve the actual security problems.

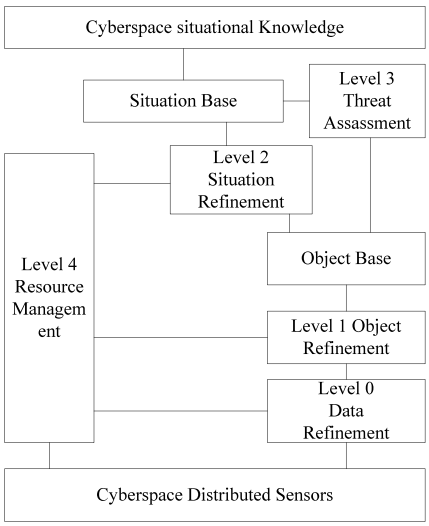

Fig.1 Cyberspace security situation knowledge distilling model

In the Cyberspace security situation analysis model above, the cyberspace security situation data is collected from base metadata, identifier, counts or attribute through the networks. The original data need to be filtered and corrected; this will process in the level 0 refinement according to the figure 1. Objects distilling in the level 1 are associated in the time (or space), data is written to show the weight. Observation data can be associated or classified according to the intrusion detection primitive base metadata[2]. Attributes of objects is checked out by using their action, association, common origin, collective protocol, correlated attractions ratio or other higher level attributes, and form a object based mass set. After objects array, associate, set on the context of such object base, situation refinement provides the situation knowledge and reorganization. The higher level attributes which readers care will be discussed in the next step. And the first step will begin at refine process of sensor, the 0 level data and 1 level object.

3 Knowledge based analysis and security cyberspace data mining

In common application, the 0 level data and the 1 level 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：

$$f(x) = \sum_k w_k \Psi_k(x) \tag{1}$$

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

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

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_1x_2; any way, if x_2 and x_3 belong to the same data, the primary function calculation of x_2x_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 $j\neq k$,

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

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

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

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 all the cyberspace security situation is considered, the result is still right in the whole cyberspace security context.

4 Conclusion

Cyberspace security situation research is a hotspot in information security research field. This paper describes the model of cyberspace security situation first. And analyzed it based on knowledge science.

The complexion of cyberspace security makes us can not solve the problems exist in the whole cyberspace security situation learning. This is the shortage of security 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, the only error item left is $\sum_{j \notin F_a} w_j^2$, it caused by primary set definition of character variable which can not be observed at A. We plan to continue developing knowledge grid computing as well by exploring issues such as data grid computing, knowledge mine, design of distributed KDD applications on the knowledge grid and knowledge discovery, and we hope that our results encourage further research by other as well[5].

The arithmetic of solution of cyberspace security situation is not based on the suppose that all the data set can be disaggrega-tired, but we can get the whole distributed security situation data model, and keep the traffic in minimal level, this is the advantage of this technique.

References:

[1] XIAO Haidong, Li Jianhua,Analysis and Research of Grid Secutrity Based on Knowledge Base *ACTA Scientiarum Naturalium Universitatis Sunyatsen* Vol.43 Nov.2004

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

[3] Liu Tongming, *Data Mining Techniques and Its Application*, National Defense Industry Press,2001, P194-218.

[4] Tim Bass. Multisensor data fusion for next generation distributed intrusion detection systems [A] . *1999 IRIS National Sympo2 sium on Sensor and Data Fusion* , Laurel ,USA ,1999.

[5] Haidong Xiao, Jianhua Li, Analysis of Grid Security Authorization Police Based on Knowledge Mining *WSEAS Transactions on Computer* p1105-p1107.