A Improved Network Risk Assessment Model Based on HMM

Yong Yan Chen
Computer Center. Kunming University of Science and Technology, KUST
Kunming, China
dawanmitang@163.com

Wei Dai
Computer Technology Application Key Lab of Yunnan Province. Kunming University of Science and Technology, KUST
Kunming, China
dw@cnlab.net

***Abstract*—In HMM risk assessment of network, relationship between the nodes is the key part to compute the state transition matrix. The change of IDS alarm level has property of randomness. To calculate the value of the IDS alarm as the input of HMM is based on the Bayesian algorithm model. Using HMM model, probability of attack successful is computed when length of attack sequence is varied. The experimental evidences show new method is veracity and validity.**

Keywords-HMM model; relationship of nodes; bayesian algorithm

I. INTRODUCTION

In the process of the information system of IDS Hazard Assessment, when an increase in attack sequence, the risk of the servers in the network is increased. To explore a scientific, rational, practical quantitative risk assessment approach is the hot spot of the current research in the field of network security. In these methods, HMM is a typical quantitative assessment method, which is popular in many scholars around world. Unfortunately, the use of pure HMM assessment methods are to ignore the interaction between network nodes. The network node correlation (NNC) analysis method is not a simple group of data. IDS alarm will change with the degree of attack and harm. And degree of attack and harm has kind of random. The method of use of NNC can be linked to a number of isolated vulnerability, which brings more accurate analysis of the overall network security risk. NNC introduced HMM, meanwhile IDS alarm level defined the state transition matrix will compute by Bayesian algorithm. To solve the HMM application problem associated widespread neglect between network nodes.

II. NETWORK GRAPH THEORY MODEL

According to the graph theory, the entire network is abstracted as a directed graph $G=(V,\{R\})$. Vertices represent each node of the network, such as servers, PC host and routers. V is a finite non-empty set of vertices. Connected to the arc adjacent vertices have access relations. R is a set of relationships between the two vertices. Shown in Figure 1, $V=\{a,b,c,d\}$, and $R=\{<ab><bc><ac><cd><ad><da>\}$.

The definition of the correlation of the network node is as follow. The correlation between nodes NNC represented as an ordered pentad <A,i,B,j,P>. The A and B denote two nodes. i and j denote the user, the operating system, applications or services on A and B node. The main P represents the probability of successful attack by the NNC, which means the attacking probability of successful exploitation of B from A. This presents as shown in Figure 1, combined graph theory. $A\in V$, $B\in V$, $<A,B>\in R$ and $<B,A>\in R$.

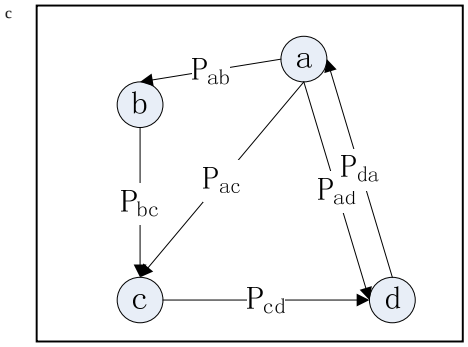

Figure 1. Network graph theory model.

III. RELATIONSHIP PROBABILITY BETWEEN NODES

In order to improve the accuracy of network risk assessment, risk of contact of the host in the network. Through the collection and analysis of network traffic data, and configuration files on the focus of analysis switches, routers, firewalls and other network and security equipment, NNC is divided into seven categories, as shown in table 1.

NNC and corresponding relationship between the node invasions can also be for seven categories. The invasion probability P may take corresponding NNC weights. Hypothesis between nodes NNC belongs to the category 1, namely A node as system administrator identity to execute orders. Its right value is 1. When the attacker A successful attack A node, and attack node B through the path < A, B >, the attack probability of success is for 1. Between node invasion probability is 1. The invasion probability P responses classification and right value, which is shown as in table 1.

TABLE I. ATTACK PROBABILITY OF TYPE AND WEIGHTS

Type	Weights	Descriptions
P_1	0.8	This is a special network association. It embodies the listener is listening relationship between access nodes and service nodes. This relationship is an important means for an attacker to obtain information of the attack.
P_2	0.1	Attacker can only access the server at the data link layer. This relationship reflects the connectivity of the data link layer. It is in order to solve the two designed for Ethernet ARP attack.
P_3	0.2	Attacker can only access server in the IP layer. This relationship reflected IP layer connectivity
P_4	0.3	This relationship only shows that the transport layer connectivity. For example, through the firewall to access internal services components
P_5	0.5	Attacker to the server components as a registered user to access to or release of public and personal information, but can not perform a system command.
P_6	0.7	Attacker access to the server as common user who executive orders, and completely control part system resources and resources of himself privileges.
P_7	1	Attacker access to the server as administrator who executive orders, and full control of all system resources.

IV. HMM-BASED EVALUATION ALGORITHM

Suppose to N nodes in the network, $V=\{v_1,v_2,...,v_N\}$. The nodes can be observed to attack there are M kinds, in this threat degree algorithm used in the literature [5], the IDS alarm is divided into 10 levels, respectively from 1 to 10. The higher the level, the greater the threat degree, $A=\{a_1,a_2,...,a_{10}\}$. A attack sequence $Y=\{y_1,y_2,...,y_t\}, y_t\in A$. HMM contains a triple $\lambda=\{T,O,I\}$. T is the transformation matrix, O is observation matrix and I is initial matrix. T represents a transformation matrix between each node.

$$T=\begin{bmatrix} 1 & p_{12} & \cdots & p_{1N} \\ \vdots & \vdots & \vdots & \vdots \\ p_{N1} & p_{N2} & \cdots & 1 \end{bmatrix}$$, where p_{ij} represents the transition probabilities from node i to node j. When $i=j$, $p_{ij}=1$. In this paper, a node transition probability values is decided by the intrusion probability p between the nodes, as $p_{ij}\in P$. That is said observed probability of certain type of attack at a node $t_{v_i}(a_j)$, where $v_i\in V, 1\le i\le N, a_j=A$ and $1\le j\le 10$.

V. BAYESIAN ALGORITHM MODEL

Depending on the degree of attack and harm, IDS alarm can be divided into five levels: no permissions, registry permissions, read access, write access and full control permissions. More serious the harm caused by, the higher the level of warning. IDS alarm level show as follow:

TABLE II. IDS ALARM LEVEL

Harm Degree	IDS Alarm
no permissions a_1	$t(a_1)$
registry permissions a_2	$t(a_2)$
read access a_3	$t(a_3)$
write access a_4	$t(a_4)$
full control permissions a_5	$t(a_5)$

The change of the harm degree has property of randomness in a degree. The traditional manual method of setting has arbitrary large. It should be consider of the characteristics of the harm degree. When approaching attack efforts in one state to the next state, it is a weak sign of performance and non-stationary. Several efforts did not result in a substantive breakthrough, the detected attack are "invalid" attack. Therefore less potential breakthrough feature information, no failure information is typical of the prior distribution. In this case, only through historical samples to get IDS alarm value. Bayesian structure can deal with variable random uncertainty and correlation which usually applies to the expression and analysis uncertainties things. This method has the ability to describe events in logic relationship of polymorphism and non-deterministic, meanwhile it can Inference mechanism and state description. The method is suitable for multi-state system reliability analysis. Bayesian method can use in the multi-state system reliability assessment, which can be a good way to make up for the deficiencies of the traditional reliability assessment methods.

After N groups of historical data will be selected, the five kinds of state probability value of the harm degree $\{t(a_1),t(a_2),t(a_3),t(a_4),t(a_5)\}$ can be calculated as follows.

$$t(a_1)=\prod_{i=1}^{n}p(a_i=1)$$

$$t(a_2)=\prod_{i=1}^{n}[p(a_i=2)+p(a_i=1)]-\prod_{i=0}^{n}p(a_i=1)$$

$$t(a_3)=\prod_{i=1}^{n}[p(a_i=3)+p(a_i=2)+p(a_i=1)]-\prod_{i=0}^{n}p(a_i=2)-\prod_{i=0}^{n}p(a_i=1)$$

and so on. The 5 types IDS alarm is shown as.

$$O=\begin{bmatrix} t_{v_1}(a_1) & t_{v_1}(a_2) & \cdots & t_{v_1}(a_5) \\ \vdots & \vdots & \vdots & \vdots \\ t_{v_N}(a_1) & t_{v_N}(a_2) & \cdots & t_{v_N}(a_5) \end{bmatrix}$$

The initial I state j is a vector which represents the probability calculated from the beginning of each node. That means the probability of attack start from each node. Thus, when the network suffered a series of attacks, IDS alarm is

for input. To use HMM builds the state transition matrix to determine the nodes of the network vulnerable to attack. At last, the most likely attack path will be calculated.

VI. EXAMPLE ANALYSIS

In this example, a power supply bureau's product network is given, shown as figure 2. The right side of the firewall is the main production server zone. Only authorized users is permitted access this zone. The left side of the firewall is external network. It could be public network or extranet. The servers in main server section are as from A to E.

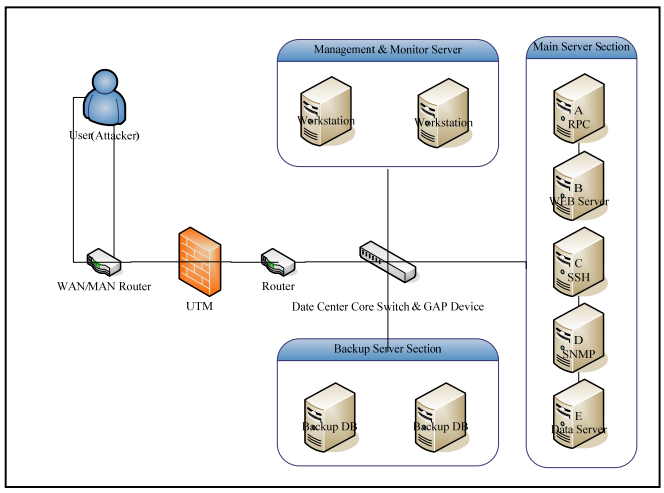

Figure 2. Topology of production department's information network

The rules descript as follow.

- The target of the strike by the Oracle database node E.
- Firewall security policy: the www service allows external network computer s access node c and node A and B ports to block all access to other nodes and ports without limiting outward from the intranet network access.
- The user on node A node access to service parts as registered user who can access to or release of public and personal information.
- The user on Node C can execute shell command in the node B without identity verification. Node B can access to node C as normal user.
- Web services on Node C can read and write data to node E database information, but can manage database on E. Node B can manage database on node E through SNMP service on node D.

According to the above rules, the network model simplified into graph theory model, as shown in figure 3 shows.

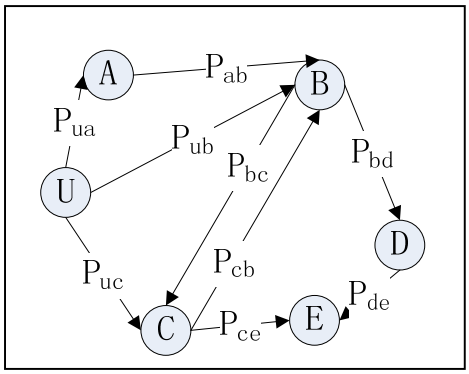

Figure 3. Network graph theory model

According to Figure 3, transition probabilities will be given as follow.

TABLE III. VALUES OF TRANSITION PROBABILITIES

Transition Probabilities.	Values
p_{ua}	0.3
p_{ub}	0.3
p_{uc}	0.2
p_{ab}	0.5
p_{cb}	0.5
p_{ce}	0.5
p_{bd}	0.7
p_{bc}	0.7
p_{de}	0.7

According to HMM algorithm, $V = \{U, A, B, C, D, E\}$.

$$T = \begin{bmatrix} 1 & 0.3 & 0.3 & 0.2 & 0 & 0 \\ 0 & 1 & 0.5 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0.7 & 0.7 & 0 \\ 0 & 0 & 0.5 & 1 & 0 & 0.5 \\ 0 & 0 & 0 & 0 & 1 & 0.7 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

To Select the 20 historical data, observation matrix O be computed , $\{t(a_1), t(a_2), t(a_3), t(a_4), t(a_5)\}$ is as follows:

$$O = \begin{bmatrix} 0.41 & 0.52 & 0.02 & 0.02 & 0.03 \\ 0.09 & 0.10 & 0.72 & 0.08 & 0.01 \\ 0.05 & 0.29 & 0.10 & 0.52 & 0.04 \\ 0.01 & 0.19 & 0.04 & 0.32 & 0.44 \\ 0.01 & 0.09 & 0.03 & 0.27 & 0.60 \\ 0.01 & 0.02 & 0.04 & 0.30 & 0.63 \end{bmatrix}$$

According to historical data of network operation, the initial state is $[0.22,0.24.0.21,0.19,0.11,0.03]$. The IDS alarm given attack sequence $[1,2,4]$ or $[1,2,4,5]$. $[U,C,E]$ or $[U,B,D,E]$ is the most possible attacker path with using HMM model. Node B and C are the key point and need protect more. Once finding attacking, the administrator should adjust its security policy and ensure the security of the database.

For single attack sequence, the likelihood of a successful attack database is relatively small. Single attack will not pose a threat to the database directly. But the longer the attack sequence length is getting, more dangerous database will be. For example, when the attack sequence come up to 4, there will be more than 30 kinds of attack sequence to mount the Web database. And the main attack routes are $[U,C,E]$ and $[U,B,D,E]$.

The attack sequence length and attack success probability are proportional. The change is relatively flat when the attack sequence length increases at the beginning. Next, there is a rising trend. Finally, the trend is to reach a steady state. On this stage, the attack sequence continues to grow, but the success rate of attacks has been under control in a certain proportion.

VII. CONCLUSION

Improved HMM network risk assessment model solve the problem of the relationship between the nodes. For the IDS alarm level randomness, Bayesian algorithm is given. Observing by different lengths of attack sequence, the attack paths are solved. Then a key node in the network is found. The further work is the check the dependencies between the nodes, which make the results of the assessment more accurate and credible.

ACKNOWLEDGMENT

This paper is based upon work supported by the National Natural Science Foundation of China under Grant No. 10878009 and No.11103005. Opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors would like to express their gratitude to some anonymous reviewers of this paper for their thoughtful comments and suggestions.

REFERENCES

[1] Swiler L P, Phillips C, Ellis D, Computer-attack Graph Generation Tool[C]//Proceedings of the 2nd DARPA Information Survivability Conference & Exposition. Los Alamitos, California, USA:IEEE Computer Society,2001:307-321

[2] Sheyner O, Haines J, Jha S, Lippmann R, Wing J. Automated generation and analysis of attack graphs. In: Hinton H, Blackley B, Abadi M,Bellovin S, eds. Proc. Of the IEEE Symp. On Security and Privacy. Oakland: IEEE Computer Society Press, 2002. 254-265.

[3] Wang Ju An, Guo Wang Hao, Guo Minzhe, Zhou Linfeng, Camargo Jairo. "Ranking Attacks Based on Vulnerability Analysis".Proceedings of the 2010 43rd Hawaii International Conference on System Sciences, 2010 1-10

[4] Cheng Qiansheng. "Rough Sets Model Based on Attribute Measure Space", Fuzzy Systems and Mathematics,vol 23, pp. 162-166,May, 2009

[5] Mell Peter and Scarfone Karen and Romanosky Sasha. "Common Vulnerability Scoring System".IEE Security and Privary,Vol. 4, pp. 85-89,June 2006.

[6] Feng Guodeng, Zhang Yang, Zhang Yuqing. "Survey of information security risk assessment". Jornal of Communications,vol. 25, pp. 10-18, July 2004.

[7] Qian Meng, Mao Han dong, Yao Li, Zhang Wei ming, "Network Security Analysis Model Based on Logic Exploitation Graph". Computer Engineering, vol. 35, pp. 147-152, September 2009.

[8] Chen XZ, Zhen QH, Guan XH, Lin CG. "Quantitative hierarchical threat evaluation model for network security". Journal of Software, vol. 17,pp. 885-897, May 2006.

[9] YANG Xiao—feng , SUN Ming—ming , HU Xue-lei , YANG Jing—yu. "Improved HMM model based method for detecting cyber attacks". Journal On Communications, vol. 31,pp. 95-101, May 2010.

[10] HUANG Guangqiu, WANG Xiaohai. Research on Network Intrusion Detection Method Based on BP-HMM. Computer Engineering, vol 33, pp. 131-133, May 2007

[11] Huang Jingde，Hao Xueliang，Huang Yi. Potential fault recognition based on improved hidden Markov model. Chinese Journal of Scientific Instrument, vol 32,pp 2481-2486,Nov 2011.

[12] Qiao Pei-li, Zhang Hai-xiao, Wang Yan-li. A Real—time Security Assessment AIgorithm Based On HMM. JOURNAL HARBIN UNIV. SCI. &TECH, vol 13, pp. 42-45,