Network Security Risk Assessment Based on Node Correlation

Zengguang Wang 1, Yu Lu 1 and Jindong Li 2
1 Shijiazhuang Campus of Army Engineering University, Shijiazhuang 050003, China
2 The 69225 Army, Hejing 841300, China
w1223797579@163.com

Abstract. A network security risk assessment method based on node correlation is proposed in order to solve the problem that little consideration is given to node correlation and diversity in tradition network security risk assessments. This method which is based on Hidden Markov model quantifies security risk of the node through indirect risks caused by the direct risks and correlations of nodes. Combined the security risk and the importance of the node, the overall risk of the target network is calculated. This method can evaluate network security risk more accurately and provide basis for the formulation of network security policy.

1. Introduction
With the continuous development of information technology, the Internet has gradually become an indispensable part of people's life and work. With the continuous development of information technology, the Internet has gradually become an indispensable part of people's life and work. An endless stream of attacks poses a great threat to network security. Traditional network security protection based on detection can only be passive defense after attack, and can not solve the network security problem from the root [1]. Network security risk assessment can evaluate the risk state of the target network before the threat occurs, and provide support for implementing the network security control strategy.

As one of the research hotspots in the field of network security, more achievements have been made in the field of network security risk assessment. The previous research on the risk of network security mainly focuses on the study of the risk of a single node [2]. To a certain extent, the quantitative problem of network security risk is solved. However, in the process of risk assessment, the importance of the network node correlation (network node correlation, NNC) and the importance of the node to the network security is not considered. The network security risk assessment result is not accurate. To solve this problem, a network security risk assessment method based on node correlation is proposed in this paper. This method is based on Hidden Markov model to evaluate the security risk of independent network nodes. Then, we calculate the correlation between nodes and the importance of nodes by probability. The overall risk of the target network is calculated through the security risk and the importance of the node.

2. Relevant Theoretical Knowledge

2.1. Definition of Node Correlation
Define 1: parts. The hosts in the target network are recorded as nodes in the network. The services and applications running on the host are recorded as the main body on the nodes. The component on the node which is called C_{Ai} is consist of the number symmetry between the host A and the main body i.
Define 2: NNC. NNC is a special access relationship based on physical connection. Attacker gains access to node A through network attack, and gets the permission of component C through attack of component C docking point B. The access relationship that can be used is called NNC from component C to component C. The access relationship that can be used is called NNC from component C to component C. NNC can be represented by an ordered five tuple <A, B, i, j and W>. Among them, A and B represent nodes in the network, and i and j represent the main body on the node. W represents the specific quantized value of the NNC relationship, which can determine the different values according to the specific network environment and practical experience [3].

Define 3: the importance of node. The importance of nodes depends mainly on the type of service, storage, and data flow. In short, the more services provided by network nodes, the more important it is. If the number of hosts in the network is L, the importance of nodes is quantified as Ps, and the importance of node A in the network is quantified as PA, and then the relative importance weight value of node A in the network can be defined as:

$$V_A = \frac{PA}{\sum_{i=1}^{L} P_i}$$

2.2. Necessity Elaboration
In the process of network security risk assessment, the correlation between nodes will affect the accuracy of the evaluation process. In the assessment of network security risk, if there is a NNC relationship between two nodes, when the control party is attacked and the prosecution is not attacked, according to the traditional risk assessment method, only the risk of the control party is increased and the prosecution is affected by the prosecution. Obviously, this is not in line with the actual situation of network operation [4]. Therefore, considering the correlation of nodes in the process of network risk assessment can improve the accuracy of evaluation results.

The location and service provided by network nodes lead to different importance of nodes relative to the network. The importance of nodes also affects the results of risk assessment. The network nodes in the core position and the network nodes at the edge position are affected by the attack, and the impact on the overall network risk is different. Therefore, in the process of network security risk assessment, we should consider the importance of nodes and the impact on network security risks.

3. Hidden Markov Model Based on Node Correlation
3.1. Hidden Markov Model
The security situation of the target network can be represented by the hidden Markov model, and the hidden Markov model can be represented by five tuples. Among them:

(1) $S = \{s_1, s_2, \ldots, s_M\}$, which represents the security state of a network node. Because the network is a changing complex system, the security state of nodes will change with time. Therefore, a stochastic process $\{X_t \mid X_t \in S\}$ is used to represent the security state of nodes at t time.

(2) $Y = \{y_1, y_2, \ldots, y_N\}$, which represents a collection of different types of attacks that can be detected in a network. Attacks in the network will change with time. Therefore, a random process $\{Z_t \mid Z_t \in Y\}$ is used to represent the types of attacks detected at time t.

(3) $A = \{a_{ij}\}_{i=1}^{M} \times \{j=1}^{M}$, which represents the transfer matrix of the node security state. In the formula, $a_{ij} = P(X_{t+1} = s_j \mid X_t = s_i), 1 \leq i, j \leq M$.

(4) $B = \{b_{k}(j)\}$, which represents the observation matrix, which indicates the probability of $y_k$ attack when node $s_j$ is in state. In the formula, $b_{k}(j) = P(y_k \mid s_j), 1 \leq k \leq N, 1 \leq j \leq M$.

(5) $\lambda = \{\lambda_i\}$, which represents initial probability distribution of security state. In the formula, $\lambda_i = P(X_0 = s_i), 1 \leq i \leq M$, which indicates the probability that nodes are at a safe state at time 0.
3.2. Node State Probability Calculation
The probability that the node K is in the state Sj at time t is $\lambda_k(i,j) = \{\lambda_k(i,j), 1 \leq i \leq M, 1 \leq j \leq N\}$, which can be calculated by the following formula:

$$\lambda_k(i,j) = \frac{P(X_i = s_i, Z_i = y_j)}{P(Z_i = y_j)} = \frac{P(Z_i = y_j | X_i = s_i)P(X_i = s_i)}{\sum_{i=1}^{M} P(Z_i = y_j | X_i = s_i)P(X_i = s_i)}$$

(2)

In the formula (2),

$$P(X_i = s_i) = \sum_{j=1}^{N} P(X_i = s_i, X_j = s_j) = \sum_{j=1}^{N} P(X_i = s_i | X_j = s_j)P(X_j = s_j) = \sum_{j=1}^{N} a_{ij}^{0} \lambda_i$$

(3)

Therefore,

$$\lambda_k(i,j) = \frac{P(Z_i = y_j | X_i = s_i)\sum_{j=1}^{N} a_{ij}^{0} \lambda_i}{\sum_{j=1}^{N} P(Z_i = y_j | X_i = s_i)\sum_{j=1}^{N} a_{ij}^{0} \lambda_i}$$

(4)

In the formula (4), $a_{ij}^{0}$ represents the transfer probability of step t of node K.

4. Risk Assessment Method Based on Hidden Markov Model
The risk value of nodes in the network can be defined by the specific environment of network operation. Vector $\text{Risk} = (r_1, r_2, \ldots, r_N)$ is used to represent the risk value corresponding to each state.

4.1. Calculation of Node Risk Value
Because of the special access relations among nodes in the network, the risk of nodes is influenced by the risk of the associated nodes in addition to their own security risk. The risks of nodes can be divided into two parts: direct risks $DR_k(t,j)$ and indirect risks $IR_k(t,j)$. $DR_k(t,j)$ refers to the risk brought by the state of the node itself; $IR_k(t,j)$ refers to the influence of the neighboring nodes on the node risk value.

The computation of $DR_k(t,j)$ does not take into account the correlation of nodes. It can be calculated by the probability distribution of T nodes and the corresponding risk vectors. The concrete formula is as follows:

$$DR_k(t,j) = \sum_{i=1}^{M} \lambda_k(i,j)r_i$$

(5)

The computation of $IR_k(t,j)$ needs to consider the special access relationship between nodes. Therefore, the calculation of $IR_k(t,j)$. It is necessary to consider the NNC relationship between node K and adjacent nodes. The definition of node correlation in reference [5] is defined [5]. There are M nodes that have NNC relations with node K, which is called $k_1, k_2, \ldots, k_M$. According to the specific network environment, the quantitative value of NNC relationship is determined as $W_{k_i,k_j}$. The value of $W_{k_i,k_j}$ is $[0, 1]$, and the greater the value is, the greater the risk of node K will be affected by related nodes. The risk value of the node with K node NNC at t time is recorded as $r_k(t,j)$. The impact of nodes with NNC relationships on the risk of node K is recorded as $\Delta r_k(t,j)$. It can be calculated through the quantized value of NNC relationship and the risk value of the related nodes. The specific formula is as follows:

$$\Delta r_{k_i}(t,j) = W_{k_i,k_j} \times r_{k_i}(t,j)$$

(6)
The indirect risk of node K is caused by the special access relationship between nodes. Therefore, the value of the indirect risk should be taken as the maximum of the risk of the related nodes.

\[ IR_k(t, j) = \max_{1 \leq l \leq M} \Delta r_k, \quad IR_k(t, j) = \max_{1 \leq l \leq M} \left\{ W_{k,l} \times r_k(t, j) \right\} \]  

(7)

The risk of the node is determined by \( DR_k(t, j) \) and \( IR_k(t, j) \). It can be calculated by the following formula:

\[ R_k(t) = \left[1 - f(x)\right] DR_k(t, j) + f(x) IR_k(t, j) \]  

(8)

\( f(x) \) is the weight function of indirect risk, representing the proportion of indirect risk in node risk, and \( x \) corresponds to the quantized value of NNC corresponding to \( IR_k(t, j) \). The selection of weight function should be carried out according to the specific conditions of the network, so as to obtain more risk values that are more in line with the actual situation of the network. The following characteristics should be met: (1) related to the quantization value of NNC relation; (2) monotonic increasing function on \([0, 1]\) interval [6].

4.2. Calculation of Network Risk Value

When there are M nodes in the target network, the corresponding risk values of nodes are recorded as \( R_k(t, j) \) through the calculation method of the node risk in 4.1. In order to be more realistic in the risk value of the network, the traditional single host risk is not directly added and then the network risk value is calculated on average. Considering different nodes' influence on network security, the importance parameters of nodes are set. The network risk value \( R(t, j) \) is calculated through node importance and node risk value. The specific calculation formula is as follows:

\[ R(t, j) = \sum_{m=1}^{M} VR_m(t, j), 1 \leq j \leq N \]  

(9)

5. Conclusion

In this paper, a node correlation based network security risk assessment method is proposed by introducing node relevance, which solves the problem that the traditional network security risk assessment is ignored by the neglect of node correlation. At the same time, considering the relative importance of nodes in the evaluation process, it is closer to the actual situation of network operation. This method can evaluate network security risk more accurately and provide support for the formulation of network security policy.

6. References

[1] Zhou Wei, Zhang Hong, Li Bohan. Network risk assessment method based on attack-defense graph model [J]. Journal of Southeast University (Natural Science Edition), 2016, 46 (4): 688-694.

[2] Li Weiming, Lei Jie, Dong Jing, et al. An optimized method for real-time network security quantification [J]. Chinese Journal of Computers, 2009, 32 (4):793-804.

[3] Poolsappasit N, Dewri R, Ray I. Dynamic security risk management using Bayesian attack graphs [J]. IEEE Trans on Dependable and Secure Computing, 2012, 9(1):61-74.

[4] Zhang Shuwei, LIU Wen-fen, WANG Jun. Network Security Quantification Assessment Based on the Correlations of Nodes [J]. Journal of Information Engineering University, 2015, 16 (2):145-151.

[5] Zhang Yongzheng, Xia Jingbo, Fang Ziyang, et al. Network security assessment based on node quantification [J]. Chinese Journal of Computers, 2007, 30 (2): 234-240.

[6] Xi Rongrong, Yun Xiaochun, Zhang Yongzheng, Hao Zhiyu. An Improved Quantitative Evaluation Method for Network Security [J]. Chinese Journal of Computers, 38 (4), 2015, 749–758.