A New Method of Network Risk Assessment Based on Bayesian Model

Kunfu Wang\textsuperscript{1}, Wei Feng\textsuperscript{1}, and Xing Li\textsuperscript{1}

\textsuperscript{1}System Engineering Research Institute of China State Shipbuilding Corporation, Beijing, China

Abstract. In order to assist network administrators to assess network security risks, a new Bayesian model of network risk assessment method is proposed. Firstly, the model designs the quantitative method of attack revenue and attack cost index, introduces the atomic attack efficiency variable, and integrates the variable into the calculation of probability, obtains the prior risk probability of each node in the network, so as to carry out the static evaluation of network risk. Secondly, DNO_Alg of deleting node order is proposed to determine the order of eliminating elements, so that Bayesian model can be transformed into cluster tree. Finally, combined with the detected attacks, the cluster tree propagation algorithm is used to dynamically calculate the posterior risk probability of nodes, so as to evaluate the network risk in real time.

1 Introduction

With the rapid development of computer network, network attacks occur more and more frequently, resulting in more and more serious network security problems. The main reason why network attacks continue to occur is that there are loopholes in the computer system\textsuperscript{[1]}. For example, the well-known "wannacry blackmail virus" attack event is to exploit the loopholes in different versions of windows to carry out attacks, which brings different degrees of economic losses to the country, enterprises and individuals\textsuperscript{[2]}. According to the data released by China Information Security Center, up to 2017, 15955 software and hardware vulnerabilities have been added to the network, and the increase in the number of vulnerabilities further aggravates the hidden danger of network security\textsuperscript{[3]}. As one of the active defense technologies, network security risk assessment can help network administrators to find potential network vulnerabilities and threats, so as to reduce network risk.

2 Related research

In recent years, scholars at home and abroad have studied network security risk assessment, and put forward some assessment methods based on different views. Wu Hongrun et al\textsuperscript{[4]} designed a network attack selection model based on the attack cost. Through the analysis of the attack cost, we found out the network attack with high probability of success, and gave the corresponding protective measures to reduce the risk of network security, but this method ignored that the attacker would also consider the attack benefit in the actual
application. Fuxionsun et al. [5] put forward to quantify the success probability of atomic attack according to whether the vulnerability information is published, whether there are corresponding attack methods and tools when they combine Markov chain and attack graph to evaluate the risk of network, and then calculate the risk value of nodes on this basis; Hu Hao et al.[6] proposed to transform the attack graph into an Absorbing Markov chain, use the CVS vulnerability scoring standard to quantify the node state transition probability in the Markov chain, and then calculate the probability of each node; the above method has certain feasibility, but when calculating the node probability, it ignores that the attacker is an intelligent decision-maker, and also considers the attack cost and attack revenue. At the same time, literature does not consider the impact of detected attacks on node probability, so it can only evaluate network risk statically, not dynamically.

Xie et al. [7] used Bayesian network to analyze network security, they considered the attack events, but this method did not further calculate the posterior probability of the nodes observing the attack events. Poolsappasit et al. [8] take the node state observed in Bayesian network as the attack evidence, and then calculate the posteriori probability of nodes in the network, so as to dynamically assess the network risk, which provides convenience for network administrators to grasp the network security situation in real time.

Chen Xiaojun et al. [9] used the observed attack events to deduce the probability of attack at a certain step, but only discussed the impact of the observed attack events at a certain node on the posteriori probability of the node, ignoring that the posteriori probability of other nodes associated with the node would be affected by the posteriori probability of the node; Gao Ni et al. [10] comprehensively considered the impact of attack events on the posterior probability of relevant nodes in the risk assessment of network through Bayesian attack graph, but did not further analyze the impact of attack cost and attack revenue on the probability in the calculation of node probability. Ma Chunguang et al. [11] quantified the node probability from attack pressure, unknown threat and other aspects when using Bayesian attack graph to dynamically evaluate the network, which improved the accuracy of the evaluation results, but ignored the situation that the relationship between nodes was "or" when calculating the node probability. With the increase of resource nodes in the network, the time complexity when Chen Xiaojun, Gao Ni and Ma Chunguang use Bayesian inference to calculate the posterior probability may increase exponentially, which is not conducive to the timely assessment of network risk.

Based on the above analysis, this paper does the following work:

1. analyze the attack cost and attack revenue, introduce the attack efficiency variable, and combine the vulnerability utilization success rate to give the formula of atomic attack success probability, so as to further calculate the node probability, so as to carry out the static evaluation of the network risk;

2. give the node deletion order algorithm (ndo_alg) to determine the elimination order, and then carry out the Bayesian attack The graph is transformed into the corresponding group tree, and then combined with the detected attack events, the group tree propagation algorithm is used to calculate the posteriori probability, and the network risk is dynamically assessed.

3 Bayesian attack graph model

The attack graph can clearly describe the potential association between vulnerabilities in the network, which provides convenience for network risk assessment. When using attack graph for risk assessment, because of the uncertainty of attack behavior, it increases the difficulty of network security risk assessment and affects the accuracy of risk assessment. However, Bayesian network has a great advantage in solving problems caused by relevance.
and uncertainty\textsuperscript{[13-14]}. Therefore, this paper constructs Bayesian attack graph to model the target network by combining Bayesian and attack graph.

**Definition 1 (Bayesian attack graph):**

\[
BSNAG = \{S, E, P, F, P_{pr}, O, P_{pos}\}
\]

is a directed acyclic graph. The parameters are defined as follows:

a) \(S = \{s_i | i = 1, 2, \ldots, N\}\) represents a collection of resource state nodes.

Such as, \(s_i\) has the property of Bernoulli random variable, \(s_i = 0\) means the resource is not occupied by the attacker, \(s_i = 1\) means the resource is occupied by the attacker.

\(S_b = \{s_j | s_j \in S, s_j \in Pa(s_i)\}\) is the initial set of resource state nodes, Parent node set with \(Pa(s_i)\) as \(s_j\),

\(S_m = \{s_j | s_j \in S, s_j \notin Pa(s_i)\}\) is the set of intermediate resource state nodes,

\(S_i = \{s_i | s_i \in S, s_i \in Pa(s_i)\}\) is the collection of target resource state nodes;

b) \(E = \{e_{i,j} | i = 1, 2, \ldots, N, j = 1, 2, \ldots, N\}\) indicates that the directed edge set between resource state nodes is also an atomic attack set. \(E \subseteq S \times S\), if \(\exists s_i, s_j \in S, s_i \in Pa(s_j)\), then \(e_{i,j} = s_i \rightarrow s_j\);

c) \(P = \{P(e_{i,j}) | i = 1, 2, \ldots, N, j = 1, 2, \ldots, N\}\) represents the probability set of atomic attack success. \(\forall P(e_{i,j}) \in P\) is the weight on the directed edge \(E\);

d) \(F = \{And, Or\}\) represents the set of dependencies between the parent nodes.

\(\forall s_i \in S, s_j \notin S, \exists f_{i,j} \in F\) represents the dependency between \(s_i\) parent nodes;

e) \(P_{pr}\) is the prior probability of resource state node;

f) \(O = \{o_{i,j} | i = 1, 2, \ldots, N, j = 1, 2, \ldots, N\}\) represents the set of attack evidence detected.

Such as, attack evidence \(o_{i,j}\) indicates that the intrusion detection system has detected atomic attack \(e_{i,j}\);

g) \(P_{pos}\) represents a posteriori probability of resource state node. That is, combining the detected attack evidence, the probability of dynamically updated resource state nodes.

### 4 Probability analysis of atomic attack success

The occurrence of atomic attack is related to the resources (security defects) in the network. When most scholars quantify the success probability of atomic attack, they usually use the success probability of vulnerability self-utilization to express it. They often ignore that the attacker is a rational decision-maker. When attacking a vulnerability in the network, they will not only consider the success probability of vulnerability utilization, but also consider the benefits and Therefore, we use the above three indicators to calculate the success probability of atomic attack.

#### 4.1 Analysis of the probability of successful exploit

The success probability of vulnerable exploitation can use Attack Vector (\(AV\)) in the basic evaluation part of CVSS scoring system, Attack Complexity (\(AC\)), Authentication (\(AU\)), three measures\textsuperscript{[15]}, the attribute scoring standards corresponding to indicators are shown in Table 1:
According to the CVSS scoring system, the quantification formula of vulnerability utilization success probability is as follows:

$$P_0(v_i) = 2 \times AV \times AC \times AU$$

(1)

### 4.2 Analysis of the benefit of atomic attack

**Definition 2 AttackProfit (AP rofit):** It refers to the interests gained by an attacker when he carries out an atomic attack. In this paper, we use resource loss to quantify the gain of atomic attack.

**Definition 3 Resource Loss (RL):** It represents the loss of a resource after being attacked by an atom. This paper describes the loss of a resource from three aspects: attack threat, resource importance, and resource security attribute, referring to the idea of quantifying the loss of a resource in reference[16].

**Definition 4 Attack Threat (AT):** It indicates the degree of damage to the target resources caused by the attack carried out by the attacker. The attack threat quantification is shown in Table 2:

| level | Attack classification | AT |
|-------|-----------------------|----|
| L₁    | Information           | 0.3|
| L₂    | Remote login          | 0.5|
| L₃    | Get user permission   | 0.8|
| L₄    | Get root permission   | 1  |

The security attributes of resources are usually represented by three indicators in the CVSS scoring system: Confidentiality (C), Integrity (I), and Availability (A). Different attacks do different harm to the three indexes in the security attributes. \((L_c, L_I, L_A)\) is used to express the different emphasis on the three indexes, and \(L_c + L_I + L_A = 1\). For example, \((0, 1, 0)\) indicates an attack against the integrity of security attributes.

**Definition 5 Resource Importance (RI):** It indicates the importance of the target node in the network. It is expressed in three levels: high, mid and low. In this paper, the importance quantification standard of typical resources is given, as shown in Table 3:

| Importance | Val | Resource description                      |
|------------|-----|------------------------------------------|
| high       | 1/2 | Hosts that hold critical information, such as database services |
| mid        | 1/3 | Mail Sever, FTP Sever, Web               |
| low        | 1/6 | General host                             |
The importance of resources has different emphasis on $(R_c, R_i, R_a)$ for the three indicators of security attributes, and $R_c + R_i + R_a = 1$, the emphasis on quantitative criteria is shown in Table 4:

**Table 4. Safety attribute partial weighting standard.**

| Applications | category          | Attribute relationship | Biased          | Biased value $(R_c, R_i, R_a)$ |
|--------------|-------------------|------------------------|-----------------|-------------------------------|
| General host | $R_i$             | C=I=A                  | Null            | $(1/3, 1/3, 1/3)$             |
| Information system | $R_i$               | I>C=A                  | Integrity       | $(1/4, 1/2, 1/4)$             |
|                |                   | I>C>A                  |                 |                               |
|                |                   | I>A>C                  |                 |                               |
| Data storage system | $R_i$               | C>I=A                  | Confidentiality | $(1/2, 1/4, 1/4)$             |
|                |                   | C>I>A                  |                 |                               |
|                |                   | C>A>I                  |                 |                               |
| Mail Sever    | $R_i$             | A>C=I                  | usability       | $(1/4, 1/4, 1/2)$             |
| Web Sever     |                   | A>C>l                  |                 |                               |
|                |                   | A>I>C                  |                 |                               |

Combined with the analysis of attack threat, resource importance and security attributes, the calculation formula of resource loss is given:

$$RL(e_{j\rightarrow i}) = AT \times RL_i \times (L_i \times R_c + L_i \times R_i + L_i \times R_a)$$  
(2)

According to definition 2, the atomic attack benefit $AProfit(e_{j\rightarrow i})$ is:

$$AProfit(e_{j\rightarrow i}) = RL(e_{j\rightarrow i})$$  
(3)

### 4.3 Cost analysis of atomic attack

**Definition 6 Attack Cost (ACost):** It indicates the price that an attacker needs to pay when carrying out an atomic attack. The atomic attack cost consists of operation cost and risk cost.

In reference\[17\], the calculation formula of the operation cost of atomic attack a is given:

$$Cost(e_{j\rightarrow i})_{operation} = a \times Cost(meta-operations)$$  
$$+ \beta \times Cost(sequence)$$  
(4)

Such as, $Cost(meta-operations)$ stands for yuan operation cost, $Cost(sequence)$ is the cost of the operation sequence.

In the process of attack, the attacker not only pays the operation cost, but also bears the risk cost, which is determined by the risk coefficient and the attacker's attack experience. The risk coefficient $\theta$ indicates the possibility that the attacker is discovered by the network security officer during the attack. The key $M(v_j)$ of the target node in the network has an impact on $\theta$, the more critical the target node is in the network, the more likely the attacker is to launch an attack on the target node, and the more likely the attack behavior is to be found by the administrator.

Generally, the more influence the target node has on the network, the more critical the position of the target node in the network. Use CVSS\[18\] scoring standard to calculate the impact of the target node:
According to formula (5), the criticality corresponding to the calculation target node is further given as follows:

\[ M(v_j) = \frac{\text{Im}(pact(v_j))}{\sum_{i=1}^{N} \text{Im}(pact(v_i))} \]  

Such as, \( N \) represents the number of target nodes in the network.

4.4 Analysis of the effectiveness of atomic attack

**Definition 7 Attack Efficacy (AE):** It indicates the attacker's recognition of atomic attack. The higher the recognition, the more likely the attacker is to carry out this attack.

Before the attack, the attacker will judge the cost and gain of the attack, and judge the attack efficiency according to the cost-benefit ratio. The higher the ratio is, the higher the cost or the lower the profit is, the lower the corresponding attack efficiency is; on the contrary, the higher the corresponding attack efficiency is.

According to formulas (3), the cost-benefit ratio \( \varepsilon(e_{i\rightarrow j}) \) of the atomic attack is:

\[ \varepsilon(e_{i\rightarrow j}) = \frac{\text{ACost}(e_{i\rightarrow j})}{\text{AProfit}(e_{i\rightarrow j})} \]  

According to formula (7), the recognition degree of attacker to atomic attack \( e_{i\rightarrow j} \) is given, that is, attack efficiency \( AE(e_{i\rightarrow j}) \) is:

\[
AE(e_{i\rightarrow j}) = \begin{cases} 
0 & \varepsilon(e_{i\rightarrow j}) \geq 1 \\
1 - \varepsilon(e_{i\rightarrow j}) & 0 < \varepsilon(e_{i\rightarrow j}) < 1 \\
1 & \varepsilon(e_{i\rightarrow j}) = 0 
\end{cases}
\]  

5 Risk assessment based on BSNAG model

5.1 Static risk assessment

Static risk assessment can find the potential danger in the target network and help the network security officers to understand the network status. In Bayesian attack graph, node risk is usually evaluated statically according to prior probability. The prior probability of a node is the joint probability of the local conditional probability of the node and its parent node. Therefore, in order to calculate the prior probability of the node, the local conditional probability of the node needs to be calculated first.

Local conditional probability reflects the risk that a resource state node may suffer. For any \( s_j \in S_m \cup S_r \), the local conditional probability of \( s_j \) is related to the atomic attack from its parent node \( Pa(s_j) \) to this node. There are two kinds of dependence relations between the parent nodes in the Bayesian attack graph: \{And, Or\}, the calculation formula of the local conditional probability of state node \( s_j \) is as follows:

(1) When the dependency between the parent nodes is \( f_j = \text{And} \):

\[
P(s_j \mid Pa(s_j)) = \begin{cases} 
0 & \exists s_i \in Pa(s_j) | s_i = 0 \\
\prod_{j=1}^{n} P(e_{i\rightarrow j}), \text{otherwise} 
\end{cases}
\]  

(2) When the dependency between the parent nodes is \( f_j = \text{Or} \):

\[ 329 \]
\[ P(s_j \mid Pa(s_j)) = \begin{cases} 0, & \forall s_i \in Pa(s_j) | s_i = 0 \\ 1 - \prod_{i=1}^{s_j} [1 - P(e_{i,n})], & \text{otherwise} \end{cases} \] \hspace{1cm} (10)

\forall s_j \in S_m \cup S_p, the prior probability of \( s_j \) is further calculated from the local conditional probability formula:

\[ P_{pre}(s_r, s_2, \ldots, s_j) = \prod_{j=1}^{j} P(s_j \mid Pa(s_j)) \] \hspace{1cm} (11)

5.2 Dynamic risk assessment

With the increase of resource state nodes, the time complexity may increase exponentially when the posterior probability is calculated by Bayesian inference, and the network status cannot be evaluated in time. The cluster tree propagation algorithm can reduce the time complexity when calculating the posterior probability. From the point of view of the cost of elimination, this paper proposes a new algorithm: the algorithm of deleting node order.

5.3 Delete node order algorithm

In the Bayesian attack graph and its corresponding rightness graph, the connection between nodes constitutes the edge. The less the number of edges, the less the connections between nodes, the less the complexity of the graph, and the less the calculation complexity and the corresponding cost of elimination. Therefore, according to the complexity of the graph, this paper proposes an algorithm of deleting node order, in which the complexity of the graph is measured by the number of edges, and the specific steps are as follows:

(a) When deleting a node in the righting graph, the edge related to the node will also be deleted. The more edges are deleted, the less the number of edges in the remaining righting graph will be, and the lower the corresponding complexity will be. DE is used to represent the number of deleted edges, which is negatively related to the complexity of a positive graph.

(b) After deleting the nodes in the alignment graph, in order to make all the neighboring nodes of the node in a group, some edges may be added between the neighboring nodes. The number of edges and the corresponding complexity of the remaining regular graphs will increase. AE is used to represent the increased number of edges, which is positively related to the complexity of the end graph.

(c) From the analysis of (a) and (b), it can be seen that the deleted edge number DE and the increased edge number AE will affect the complexity IC of the residual end normal graph, so the complexity can be quantified by IC=AE/DE. In this paper, according to the complexity of the residual correction graph after deleting nodes, the order of node deletion in the correction graph is determined. When deleting a node, the complexity of its corresponding residual end graph is smaller than that of other nodes, so the node is deleted first. And so on to determine the order in which nodes are deleted.

Based on the above analysis, this paper proposes the algorithm of deleting node order, DNO_Alg, and the specific steps of DNO_Alg are as follows:

Input: the correct graph corresponding to BSNAG,
Output: node removal order,
1:Initialize queue Q to store node deletion order ;
2: Determine whether the number of nodes is greater than 3; if yes, execute step 3; otherwise, execute step 5 ;
3: Calculate the complexity of each node IC=AE/DE;
4: Find out the node corresponding to the minimum complexity IC, store the node in the queue Q and delete the node from the graph, the number of nodes decreases, and execute step 2;
5: Output node deletion order queue Q, algorithm terminated.

The common feature of DNO_Alg algorithmic and minimum edge missing search algorithm: when calculating the number of deleted and increased edges of a node, all the neighboring nodes of the node need to be searched. Therefore, the time complexity of DNO_Alg algorithm is similar to that of the minimum edge missing search algorithm.

6 Conclusion

Network security risk assessment has always been a hot topic at home and abroad. In this paper, a new method of network risk assessment based on Bayesian model is proposed. First of all, according to the current network attackers are rational, the attack efficiency and vulnerability utilization success probability quantified by the atomic attack cost and the atomic attack revenue are combined to calculate the atomic attack success probability, and further calculate the prior risk probability of nodes to evaluate the network statically; secondly, the node deletion order algorithm is propose NDO_Alg, transform Bayesian model into cluster tree.

References

1. Xing Xu-jia, Lin Chuang, Jiang Yi-xin. A Survey of Computer Vulnerability Assessment[J]. Journal of Computer, 2004, 27(1): 1-11.
2. Akbanov M, Vassilakis V G, Logothetis M D. Ransomwa-re detection and mitigation using software-defined networking: The case of WannaCry[J]. Computers & Electrical Engineering, 2019, 76: 111-121.
3. Zhao Shang-ru, Li Xue-jun, Fang Yue, et al. A Survey on Automated Exploit Generation[J]. Journal of Computer Research and Development, 2019, 56(10):2097-2111.
4. Wu Hong-run, Qin Jun, Zheng Bo-jin. Anti-attack Ability Based on Costs in Complex Networks[J]. Computer Science, 2012, 39(S):224-227.
5. Sun F, Pi J, Lv J, et al. Network Security Risk Assessment System Based on Attack Graph and Markov Chain[J]. Journal of Physics: Conference Series, 2017, 910(1):012005.
6. Hu Hao, Liu Yu-ling, Zhang Hong-qi, et al. Route Prediction Method for Network Intrusion Using Absorbing Markov Chain[J]. Journal of Computer Research and Development, 2018, 55(4):831-845.
7. Xie P, Li J H, Ou X, et al. Using Bayesian networks for cyber security analysis [C] //Proceedings of 2010 IEEE/IFIP International Conference on Dependable Systems and Networks( DSN) . Chicago: IEEE, 2010: 211-220.
8. Poolsappasit N, Dewri R, Ray I. Dynamic Security Risk Management Using Bayesian Attack Graphs[J] . IEEE Transactions on Dependable & Secure Computing, 2012, 9(1):61-74.
9. Chen Xiao-jun, Fang Bin-xing, Tan Qing-feng, et al. Inferring Attack Intent of Malicious Insider Based on Probabilistic Attack Graph Model[J]. Journal of Computer, 2014, 37(1): 62-72.
10. Gao Ni, Gao Ling, He Yi-yue, et al. Dynamic Security Risk Assessment Model Based on Bayesian Attack Graph[J]. Journal of Sichuan University, 2016, 48(1):111-118.

11. Ma Chun-guang, Wang Cheng-hong, Zhang Dong-hong, et al. A Dynamic Network Risk Assessment Model Based on Attacker’s Inclination[J]. Journal of Computer Research and Development, 2015, 52(9):2056-2068.

12. Kaynar K. A taxonomy for attack graph generation and us age in network security[J]. Journal of Information Security and Applications, 2016, 29(C):27-56.

13. Li S, Tryfonas T, Russell G, Andriotis P. Risk assessment for mobile systems through a multilayered hierarchical Bayesian network[J]. IEEE Transaction on Cybernetics, 2016, 46(8):1749-1759.

14. Hu Hao, Ye Run-guo, Zhang Hong-qi, et al. Vulnerability life cycle oriented security risk metric method[J]. Journal of Software, 2018, 29(5):1213–1229.

15. Lei Ke-nan, Zhang Yu-qing, Wu Chen-si, et al. A System for Scoring the Exploitability of Vulnerability Based Types[J]. Journal of Computer Research and Development, 2017, 54(10):2296-2309.

16. Jiang Wei. Research on Key Technologies of Active Defense Based on Attack and Defense Game Model[D]. Harbin Harbin Institute of Technology, 2010.

17. Wang Hui, Liu Shu-fen. A Scalable Predicting Model for Insider Threat[J]. Journal of Computer, 2006, 29(8): 1346-1355.

18. Tripathi A, Singh U K. Analyzing trends in vulnerability classes across CVSS metrics[J]. International Journal of Computer Applications, 2011, 36(3): 38-44.