A Dynamic Risk Assessment Method Based on Bayesian Attack Graph

Zhiyong Luo, Rui Xu, Jianming Wang, and Weicheng Zhu
(Corresponding author: Zhiyong Luo)
School of Computer Science and Technology, Harbin University of Science and Technology
Harbin 150080, China
Email: luozhiyongemail@sina.com
(Received Mar. 8, 2021; Revised and Accepted Apr. 5, 2022; First Online July 3, 2022)

Abstract

This paper proposes a network intrusion intention analysis method based on Bayesian attack graphs, aiming to improve the incomplete atomic attack probability common in current network risk assessment models. Most evaluation models ignore the impact of intrusion intentions on security risks. First, based on the Bayesian belief network, a quantitative attack graph of the atomic attack probability is calculated by quantifying the vulnerability value, attack cost, and attack benefit. Next, the dynamic transition probability is proposed to describe the relationship between nodes in different network states, and a risk assessment model based on the Bayesian belief network is established. Finally, dynamically update the node's state to predict the attacker's intention. The experimental comparison shows that the model can dynamically assess the network intrusion risk and predict the attack path more accurately.

Keywords: Attack Graph; Bayesian Belief Network; Invasion of Intentions; Network Security; Risk Assessment

1 Introduction

With the development of network attack technology, network intrusion becomes easier and more concealed, and the multi-step nature of network attack becomes one of the difficulties in the study of security events [13]. When security personnel detect an intrusion, the permissions on the system may have been compromised. Because of the multi-step nature of the attack, it is difficult for the defenders to determine the attack path [4]. Without proper risk assessment, intrusion response systems can degrade network performance, mistakenly disconnect users from the network, or cause administrators to re-establish services at great cost [15].

The contribution of this research has the following three points:

1) Most models use atomic attack probability to evaluate simply from the vulnerability availability, and lack comprehensiveness. In order to obtain the most primitive data accuracy, this article calculates the probability of atomic attacks from three indicators: vulnerability availability probability, attack cost and attack revenue. More indicators have more accurate advantages in calculating probability, and truly reflect the use of vulnerabilities in the actual network;

2) In order to improve the accuracy of the dynamic network risk assessment, the Bayesian belief network is combined with the attack graph, the dynamic transition probability is innovatively proposed, and the dynamic risk assessment model is established, which can improve the overall assessment efficiency of the dynamic network risk;

3) Using the dynamic transition probability proposed in this paper to generate the attack path and calculate the overall reachable probability of the path can avoid the influence of the vulnerability of a single network node on the path selection, so as to realize the prediction of the attack path and improve the accuracy of the prediction.

2 Related Work

In the 1990s, Phillips et al. first proposed the concept of attack graph, which used the configuration information of the attacked nodes, the causal relationship between nodes and the attacker's ability to generate attack graph and applied it to the analysis of network vulnerability [14]. Attack graph is a directed graph composed of vertices and directed edges. According to different models, vertices can represent elements of host, service, vulnerability, permission and network security status, directed edges indicate the path and order of attackers [12]. Using the attack graph, we can model the path that an attacker may invade the target network [10]. and graphically display the details of the attack behavior, such as the target network, vulnerability, attack path, etc. [7]. so as to provide

support for predicting the attacker's subsequent attack behavior and facilitate administrators to timely respond to unexpected network intrusion events [18].

Harjinder [9] proposes a threat prediction algorithm based on Bayesian attack graph, which can provide complete prediction information with threat scenarios, and then quantify the threats in the threat prediction algorithm into security risks from two levels of host and network. In order to deal with the uncertainty of attack probability, Mohammad [6] optimized the attack graph, analyzed IDS alarm and intrusion response data to update the attack probability, and finally generated the prediction attack graph to gain insight into network security. Based on the traditional attack graph, WANG [16] adds the weight of attack distance, calculates the possible attack path, considers the attack cost of different nodes, and used the estimation function to judge the preferred attack path. Ahmadianramaki and Rasoolzadegan [1] first extract the causal relationship between intrusion alarms, then use Bayesian network to construct attack scenarios, and finally predict the subsequent attacked nodes. Hu [7] used alarm information and real-time attack behaviors from different dimensions to calculate vulnerability utilization rate, assess the attacker's ability, and put forward a threat prediction algorithm based on dynamic Bayesian attack graph to quantify network threats and the risk of sustained attacks. Fan [3] proposed an attack graph construction method based on Rete. The Rete algorithm was added in the process of attack graph construction to transform the constructed attack graph into pattern matching between threat action attributes. In order to defend the moving targets in the network and calculate the costs and benefits, LEI [11] used the attack graph to build a hierarchical network resource graph. Combined with the variable point detection method, a defense effectiveness evaluation method based on the variable point detection is proposed, which can effectively improve the construction efficiency of the network resource graph.

The above research establishes different network security risk assessment models based on attack graph, However, the evaluation index of atomic attack probability defined in CVSS is relatively single, it cannot quantify the risk of network nodes against the attacker's intention. In Section 3.1 of this article, a dynamic network intrusion intention analysis model is established based on Bayesian attack graphs.

3 Bayesian Attack Graph Establishment

It may not be easy to directly perceive network attacks, and effective means are needed to help perceive network attacks. Attack modeling techniques, such as attack graph and attack tree, are commonly used mathematical models [8], which can intuitively represent the sequence of network nodes, which may lead to a successful attack on a given network [2].Attack graph can be divided into state attack graph and attribute attack graph. In the state attack graph, vertices represent the state information of the network, while edges represent the migration direction and process of the state. However, the state attack graph cannot deal with the rapidly growing state nodes, and its structure is not intuitive enough, so it is not suitable for large-scale networks. Each attribute vertex in the attribute attack graph represents an independent security element, avoiding the state explosion problem of the state attack graph [1]. Therefore, the attribute attack graph has better scalability for complex large-scale networks. The Bayesian attack graph proposed in this paper is a kind of attribute attack graph, which combines Bayesian theory and attribute attack graph, and uses Bayesian belief network to describe the dynamic relationship between attacks. Aiming at the complexity and variability of network status in reality, this paper proposes the concept of dynamic transition probability based on the traditional Bayesian attack graph, and calculates the transition probability based on the connections between network nodes. It can not only calculate the probability of reaching each node in the attack graph more accurately, but also has a better effect in predicting the possible attack path.

3.1 Bayesian Attack Graph Definition

Bayesian attack graph is a directed acyclic graph and can be expressed as BAG=(S,A,E,R, P),the definition is as follows:

1) S is the set of attribute nodes, which is divided into three categories, namely $S = S_{start} \cup S_{transition} \cup S_{target}$,$S_{target}$ represents the originating node of a network attack, $S_{transition}$ is the Intermediate node of attack behavior, S_{target} is the target node of this attack. Among them, $S_i = \{0, 1\}$.1 means that the attacker has successfully exploited the node vulnerability of this attribute to possess the node; otherwise, 0 means that the node is not occupied;

2) $A = \{A_i | i = 1, 2, \ldots, n\}$ is the atomic attack set, which refers to the attack behavior of the attacker against the node vulnerability. In other words, the migration mode of attribute nodes can be expressed as $A_i : S_{pre} \rightarrow S_{next}$;

3) $E = \{E_i | i = 1, 2, \ldots, n\}$ is the set of directed edges in the attack graph, which represents the causal relationship between the attack behavior between attribute nodes,$(S_{pre}, S_{next}) \in E_i$ represents a directed edge that attacks S_{next} from the S_{pre};

4) R represents the parent-child attribute node involvement. Can be represented by binary group $< S_j, d_j >$, $d_j \in \{AND, OR\}$,AND means that the attack can only be completed if all the parent nodes that reach S_j have true states. Similarly, OR means as long as one of the parent nodes is true;

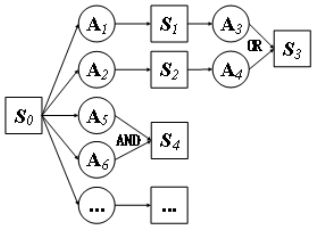

Figure 1: Bayesian attack graph example diagram

5) P represents the accessibility probability of attribute nodes in the attack graph, and P_1 represents the static reachability probability of attribute nodes in the attack graph. P_2 is the dynamic reachability probability of node attributes in the attack graph.

3.2 Build Bayesian Attack Graph

3.2.1 Bayesian Attack Graph

The Bayesian network is a directed acyclic graph, in Bayesian networks, the state and occurrence probability of nodes are only related to the parent node. In the attack graph, whether the network vulnerability is used is also related to the parent node in the attack path. This kind of node relationship is related to the Bayesian network and Corresponding to the attack diagram. Since both Bayesian network and attack graph are a kind of directed acyclic graph, and directed edges represent a kind of causal relationship, Bayesian network and attack graph can be combined to predict the network security situation.

As shown in Figure 1: S_0 is the originating node of the attack, S_3 and S_4 are the target network nodes of the attacker, S_1 and S_2 are intermediate transition attribute nodes. A_1,A_2,A_3,A_4,A_5,A_6 represent atomic attacks. AND means that the attack strategies of atomic attack A_5,A_6 reaching S_4 are all true, the attack can only be realized; OR means that the attack strategy of atomic attack A_3,A_4 reaching S_3 can be implemented as long as one of them is true, namely, the attack on target node S_3 can be completed by completing either of the two attack paths shown in the legend.

3.2.2 Vulnerability Utilization Probability

The exploitable probability of the vulnerability is related to the vulnerability of the attribute node. Generally, the Common Vulnerability Scoring System (CVSS) [17] provided by the national vulnerability database (NVD) of the United States is used for quantification. CVSS can provide complete scoring parameters, an open scoring framework, a combination of dynamic assessment and vulnerability dependencies between attribute nodes, and quantification of vulnerability utilization. According to the CVSS quantification standard, this paper quantifies the usability from four indexes: Access Vector (AV), Access Complexity (AC), Privileges Required (PR) and User Interaction (UI). The measurement given in the CVSS quantification standard covers all aspects of vulnerability value measurement, which makes the measurement results more accurate and avoids the bias of the prediction results due to the lack of measurement indicators. Therefore, in order to better highlight the degree of impact of different impact indicators on network security, this article innovatively uses the classification of different impact indicators. The method is to assign a lower level of registration to indicators with a relatively small degree of influence, and assign a higher level to indicators with a relatively large degree of influence.The specific scores are shown in Table 1:

Table 1: CVSS indicator score

Indexes	Measurements	Score
AV	Network(N)	0.85
	Adjacent(A)	0.62
	Local (L)	0.55
	Physical(P)	0.20
AC	Low (L)	0.77
	High (H)	0.44
PR	None (N)	0.85
	Low (L)	0.62
	High (H)	0.07
UI	None (N)	0.85
	Required(R)	0.62

In order to quantify the Vulnerability utilization probability, the score of vulnerability needs to be calculated first, and the calculation formula is shown in Equation (1):

$$Score = 8.22 * AV * AC * PR * UI \tag{1}$$

Definition 1. *Since CVSS standard vulnerability score range is* $[0, 10]$*, for vulnerability* v_i*,* P_{v_i} *is used to quantify its vulnerability utilization probability, and the calculation formula is shown in Equation (2):*

$$P(v_i) = \frac{Score}{10} * 100\% \tag{2}$$

(Note: The parameters in this article are for reference only, and the value of each parameter is modified and set by the administrator according to the specific network environment.)

3.2.3 Conditional Probability

In the attack graph, the vulnerability is not independent, and whether it can be exploited is also affected by its parent node.

Definition 2. *Conditional probability means the possibility of a certain attribute node vulnerability being exploited under the influence of its parent node vulnerability. For attribute node* S_j*, conditional probability is expressed by* $P(S_j|P_{ar(}S_j))$*, and* $P_{ar(}S_j))$ *means the set of its parent nodes. According to* d_j*, the calculation formula of conditional probability is shown in the following equations:*

Table 2: Attack cost index score

Cost	Measurements	Score
SI	Complete/Function/Null	0.1/0.3/0.7
SP	Common/Special/Particular	0.15/0.35/0.6
Or	Tool/Script/Manual/Corporation	0.1/0.25/0.45/0.7
IR	Null/Regular/Configuration/Critical	0/0.2/0.55/0.8

Table 3: Attack benefits index score

Measurements	Score
Information Leakage	0.3-0.55
Remote Register	0.55-0.7
Authentication Bypass	0.7-0.8
Limited Access	0.85-0.95
Root Access	1.0

1) When d_j=AND,

$$P(S_j|P_{ar}(S_j)) = \begin{cases} 0, (\exists S_i \in P_{ar}(S_j), S_i = 0) \\ \prod_{i=1}^{n} P_{ar}(v_i), (others) \end{cases} \quad (3)$$

2) When d_j=OR,

$$P(S_j|P_{ar}(S_j)) = \begin{cases} 0, (\forall S_i \in P_{ar}(S_j), S_i = 0) \\ 1 - \prod_{j=1}^{n}(1 - P_{ar}(v_i)), (others) \end{cases}$$

3.2.4 Attack Costs and Benefits

When an attacker attacks a network node, he will not only consider the available probability of the vulnerability of the node, but also consider the cost of attacking the node and the benefit after the attack. The cost and benefit of attack will not affect the original state transition between nodes, but will affect the choice of attack nodes. A rational attacker will choose nodes with low cost and high benefit.

Definition 3. *When an attacker initiates an attack, he will invest necessary costs such as human resources, material resources, and attack cost. For atomic attack A_i, $cost(A_i)$ is used to represent the cost of the attack.*

In this paper, the attack cost is evaluated from four indexes: Shellcode Information (SI), Shellcode Platform (SP), Operation Requirement (OR), and Information Requirement (IR). The specific score is shown in Table 2:

The attack cost can be quantified by the scores of SI, SP, Or and IR. The calculation formula is shown as follows:

$$cost(A_i) = 1 - ((1 - SI) * (1 - SP) * (1 - Or) * (1 - IR))$$

Definition 4. *For an atomic attack A_i, when an attacker completes an attack on a node through the attack, the proceeds that can be obtained are called attack proceeds, which are expressed by benefit (A_i). The specific score is shown in Table 3.*

The final state value of the attribute node after the attack is equal to the benefit(A_i) of the attack. Each final state value score given is a range value.

3.3 Node Reachability Probability

3.3.1 Atomic Attack Probability

Based on the quantification of the exploitable probability, attack cost and benefit of node vulnerabilities, the attack probability of an attacker against its child nodes in the current attribute node can be calculated, that is, the probability of an atomic attack, with a value range of $[0, 1]$. When the attack probability is 0, it means that the attack has no benefit to the attacker, and the attacker will not launch the attack. When the value is 1, it means that the gain of the attack is far greater than the cost, and the attacker must launch the attack.

Definition 5. *The probability that an attacker completes an atomic attack A_j through vulnerability v_i is called the atomic attack probability, which is represented by $P(A_j)$, and According to the indicator definition, the calculation formula is shown in Equation (4):*

$$P(A_j) = min\left(\frac{P(v_i) * benefit(A_j)}{cost(A_j)}, 1\right) \quad (4)$$

3.3.2 Static Reachability Probability

By using the conditional probability of all attribute nodes in the Bayesian attack graph, the reachable probability of each node, namely the static reachability probability, can be calculated. Static reachability probability can be used for static evaluation of network risk to show the static risk of the network.

Definition 6. *The static reachability probability represents the reachable probability of each attribute node in the static network and is the joint conditional probability of the current node and its ancestor node. That is, for $S_j \in S_{transition} \cup S_{target}$, the calculation formula of the node S_j static reachability probability is shown in Equation (5):*

$$P_1(S_j) = \prod_{j=1}^{n} P(S_j|P(S_j)) \quad (5)$$

The static reachability probability of attribute nodes S_1 and S_2 is calculated by combining their conditional probability with S_0 static reachability probability, and the static reachability probability of attribute node S_3 also depends on the static reachability probability of S_1 and S_2. See Figure 2 for details:

Then the prior probability of S_1, S_2 and S_3 is:

$$P_1(S_1) = P(S_1|S_0 = 1) * P(S_0) = 0.7 * 0.9 = 0.63$$

$$P_1(S_2) = P(S_2|S_0 = 1) * P(S_0) = 0.7 * 0.4 = 0.28$$

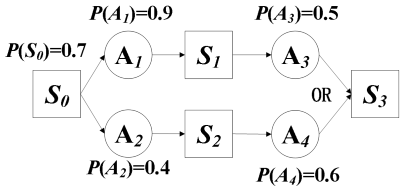

Figure 2: Bayesian attack graph exploit probability of vulnerability

$$P_1(S_3) = \begin{cases} P(S_3|S_1=1,S_2=1)*P(S_1|S_0=1)* \\ P(S_2|S_0=1)*P(S_0)+ \\ P(S_3|S_1=1,S_2=0)*P(S_1|S_0=1)* \\ P(S_2=0|S_0=1)*P(S_0)+ \\ P(S_3|S_1=0,S_2=1)*P(S_1=0|S_0=1)* \\ P(S_2|S_0=1)*P(S_0) \end{cases}$$

$$= \begin{cases} 0.5*0.9* \\ 0.4*0.7+ \\ 0.5*0.9* \\ 0.6*0.7+ \\ 0.6*0.1* \\ 0.4*0.7 \end{cases} = 0.3318$$

3.3.3 Dynamic Reachability Probability

In fact, the network is not static. When the attacker's intention is known, in order to quantify the network risk according to the attacker's intention, it is necessary to update the reachability probability of other nodes in combination with the known target attribute node.

Definition 7. *The attribute node set* $S = \{S_i|i=1,2,...,n\}$ *is divided into the node set detected to be attacked* $S_{compromised} = \{S_j \in S|S_j = 1\}$ *and the node set to be updated* $S_{update} = S - S_{compromised}$. *The dynamic reachability probability represents the probability of dynamically updating the reachability probability of the node* $S_b(S_b \in S_{update})$ *in the update set after capturing the node* $S_a(S_a \in S_{compromised})$ *to be attacked, the calculation equation is shown as follows:*

$$\begin{aligned} P_2(S_b|S_{compromised}) &= \frac{P(S_{compromised}|S_b * P_1(S_b))}{P_1(S_{compromised})} \\ P(S_{compromised}|S_b) &= \prod_b P(S_b=1|S_a) \\ P_1(S_{compromised}) &= \prod_b P(S_b=1). \end{aligned} \quad (6)$$

In Figure 1 of Section 3.2, assuming that the target attribute node of the attacker is S_3, the dynamic reachability probability of attribute node S_2 is calculated:

$$\begin{aligned} P_2(S_2|S_3) &= \frac{P(S_2)|S_3 * P_1(S_2)}{P_1(S_3)} \\ &= \frac{\sum_{S_1=0,1}[P(S_3|S_1,S_2)*P(S_1)]*P_1(S_2)}{P_1(S_3} \\ &= \frac{0.5*0.28}{0.3318} = 0.4219. \end{aligned}$$

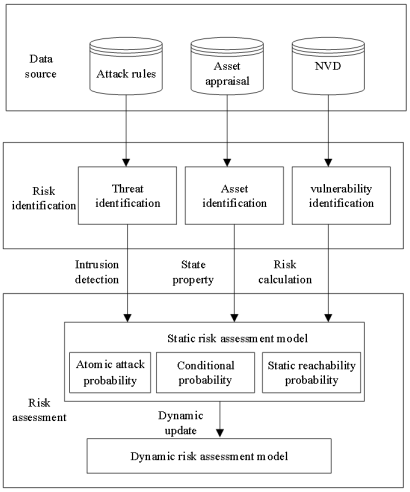

Figure 3: Bayesian attack graph risk assessment system framework

4 Risk Analysis Method of Bayesian Attack Graph

4.1 Model Design

The process of constructing Bayesian attack graph is to connect the security elements of network risk, namely threats, resources and vulnerabilities, find out the attack path intended by the attacker and calculate its reachable probability, and establish the intrusion risk assessment system. The risk assessment system framework based on Bayesian attack graph proposed in this paper is shown in Figure 3: There are three stages:

1) Data acquisition phase:There are three types of data collection: attack rules, asset evaluation and NVD. Attack rules construct a labeled directed graph semantic rule model, knowledge description of the attack technology described in natural language text, definition of semantic rules, and formal description methods to explain the attributes of network entities and their logical operation relationships; asset evaluation is an evaluation system The key assets of the system are the business and data of the system, including core business components, user data, passwords and keys used for authentication and authentication; NVD is the US National Vulnerability Database.

2) Risk identification stage: Use Snort intrusion detection system to identify network threats and detect attack events; OVAL vulnerability scanning technology is used to identify host vulnerability information, and CVSS standard is used to quantify the possibility of successful exploitation of vulnerabilities. Network resources are identified and associated with vulnerability.

3) Risk analysis stage, vulnerability probability and local conditional probability are used to calculate the accessibility probability of each attribute node, and a static risk assessment attack graph is established. Based on the intention of attackers, the reachability probability of nodes is updated and used to generate a dynamic risk assessment model.

4.1.1 Static Risk Assessment

After calculating the conditional probability and static reachability probability of attribute nodes, a static risk assessment model can be built on the basis of the original attack graph. The static risk assessment can evaluate the potential risks in the network. The construction algorithm is shown in Algorithm 1.

Algorithm 1 Static risk assessment attack graph $S_{BAG} = (S, A, E, R, P_1)$

Begin
Initialize parameters in S_{BAG}; Attribute nodes, atomic attack, directed edges and dependencies in AG were copied to S_{BAG};
for (each directed edges E_iE in S_{BAG}) **do**
  Calculate P_{A_i} using formula (4);
end for
for (each attribute node S_i in S_{BAG}) **do**
  if i=1 **then**
    $P_1(S_1 = 1) = P$;
  else
    Calculate $P(S_i|P_{ar}(S_i))$ using formula (5-6);
    Calculate $P_1(S_i)$using formula (7);
  end if
  Copy $P_1(S_i)$ into the parameter P_1;
  return Static Bayesian attack graph; $S_{BAG} = (S, A, E, R, P_1)$
end for
End

4.1.2 Dynamic Risk Assessment

In the real complex network, the elements of network security will change with the operation of the network, and the accuracy of static risk assessment will be reduced when the attacker's attack target is known. Therefore, it is necessary to build a dynamic risk assessment model based on the dynamic reachability probability calculated by Bayesian theory. The construction algorithm is shown in Algorithm 2.

Algorithm 2 $DYNAMIC_BAG(S_{BAG})$

Begin
Initialize parameters in D_{BAG}; Attribute nodes, atomic attack, directed edges, dependencies and Static reachability probability in S_{BAG} were copied to D_{BAG};
for (each attribute node S_i in S_{BAG}) **do**
  if S_i=0 **then**
    $S_i \in S_{update}$;
  end if
end for
for (each attribute node S_i in S_{BAG}) **do**
  if S_i=0 **then**
    for (each parent node $S_k \in P(S_j)$) **do**
      Calculate $P_2(S_i)$ of S_k using formula (8);
    end for
  end if
end for
Copy $P_2(S_i)$ into the parameter P_2;
return Dynamic bayesian attack graph $D_{BAG} = (S, A, E, R, P_1)$ graph $D_{BAG} = (S, A, E, R, P_2)$;
End;

4.1.3 Attack Path Generation

Definition 8. *The attack path indicates that in the generated Bayesian attack graph, the intruder can invade the target node S_{target} from the initial attribute node S_{start} along a group of attribute nodes, then the path composed of the group of nodes is an attack path AP_i of the Bayesian attack graph, and the Attack Path set in the attack graph is recorded as attack path. The specific algorithm is shown in Algorithm 3).*

Definition 9. *In order to compare the attack probability of different paths, the product of the reachability probability of all nodes in a path is called the total reachability probability of the path, that is, for AP_i, the calculation equation of the total reachability probability is shown in Equation (7):*

$$P(AP_i) = \prod P(S_i), S_i \in AP_i \tag{7}$$

5 Experimental Analysis and Optimization Evaluation

5.1 The Experimental Setup

In order to verify the accuracy of the intrusion intention analysis model based on Bayesian attack graph, this paper establishes the network topology as shown in Figure 4. The structure mainly includes D_1 domain, D_2 domain, D_3 domain and DMZ domain. Through the installation of firewall, the network area is divided and the communication rules between subnets are formulated to ensure that external access cannot reach the internal network area. The specific visiting rules are as follows:

Algorithm 3 Attack Path(S_{BAG}, D_{BAG})

```
Begin
Initialize parameters in Attack Path;
for (for (each target node S_i ∈ S_target) do
    Add S_i to AP_i;
    if P_ar(S_i) ≠ None then
        if d_j == OR then
            n = len(P(S_i));
            copy AP_i to (AP_{i-1}, AP_{i-n});
            for (each node S_j ∈ P(S_i)) do
                Add S_j to AP_{i-j};
            end for
        else
            Add P(S_j) to AP_i;
            S_i = P(S_i);
        end if
    else
        return AP_i;
    end if
    Add AP_i to Attack Path;
end for
return Attack Path;
End
```

1) Only host H_4 in D_1 domain can access SQL server;

2) Only host H_9 in D_2 domain can access SQL server;

3) The hosts of D_1 domain and D_2 domain can access each other with servers in DMZ domain;

4) When D_1 domain accesses D_2 domain, it can only access host H_7 through host H_4;

5) Hosts in the domain can access each other, and other cross domain access is prohibited.

5.2 Attack Graph Generation

Use OVAL vulnerability scanner to scan the experimental network, get the vulnerability information of each host and service, and use Equations (1) and (2) to calculate the vulnerability utilization probability, as shown in Table 4:

In this network, there is important data in SQL server, H10 can be regarded as the attacker's invasion intention, and the scanned vulnerability information, the relationship between vulnerabilities, host and server information, network configuration and other data can be used to generate and output a graphical attack diagram, as shown in Figure 5: In the attack diagram shown in Figure 5, the attribute node represents the host information or vulnerability information, and the atomic attack represents the state migration mode of the attribute node. When a node in the figure has multiple parent nodes, the parent-child nodes can see that the relationship is all OR, that is, $d_j = OR$.

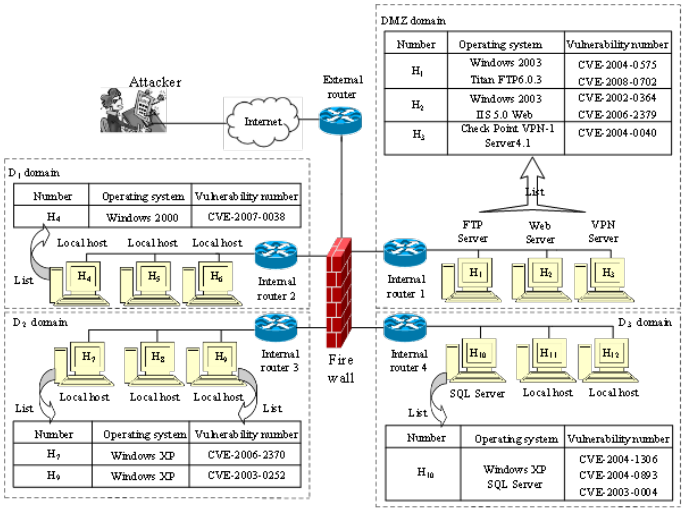

Figure 4: Experimental network topology

Table 4: Vulnerability information and vulnerability utilization probability

Host	CVE number	Vul-number	P(vi)
H1	CVE-2013-4465	v_1	0.46
	CVE-2004-0575	v_2	0.53
H2	CVE-2002-0364	v_3	0.35
	CVE-2006-2379	v_4	0.39
H3	CVE-2009-0241	v_5	0.43
H4	CVE-2007-0038	v_6	0.55
H7	CVE-2006-2370	v_7	0.25
H9	CVE-2003-0252	v_8	0.39
H10	CVE-2004-1306	v_9	0.51
	CVE-2004-0893	v_{10}	0.36
	CVE-2015-1762	v_{11}	0.41

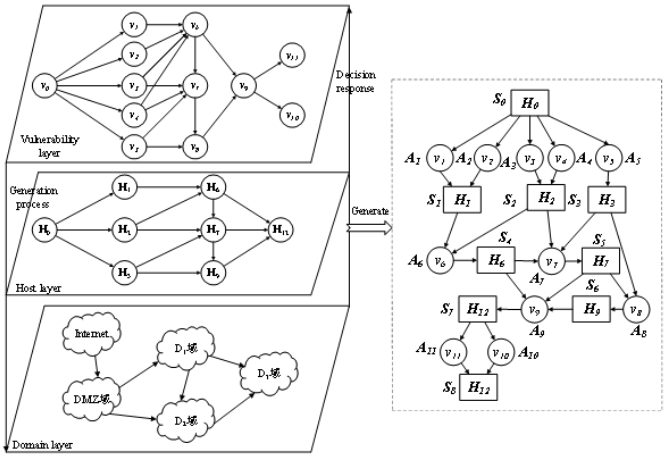

Figure 5: Attack graph in experimental environment

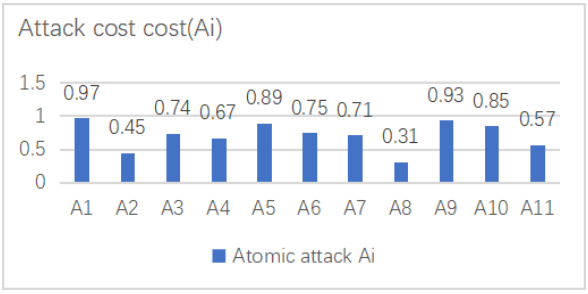

Figure 6: Atomic attack cost

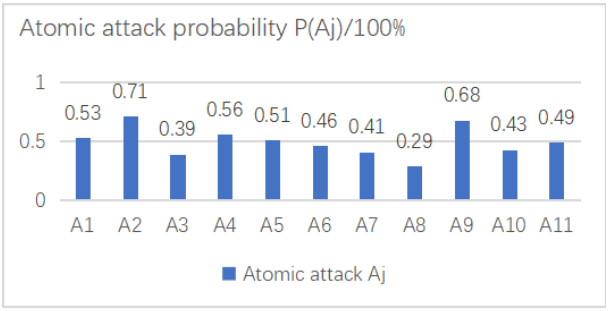

Figure 7: Atomic attack probability

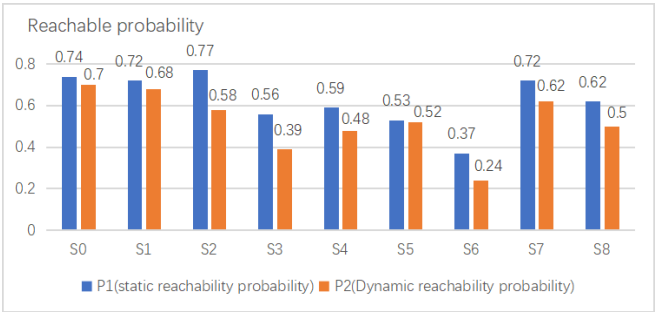

Figure 8: Attribute node accessibility probability

Table 5: Attack paths

Num	Attack path	Num	Attack path
AP_1	S0-S1-S4-S7-S8	AP_5	S0-S3-S5-S7-S8
AP_2	S0-S2-S4-S7-S8	AP_6	S0-S3-S6-S7-S8
AP_3	S0-S2-S4-S5-S7-S8	AP_7	S0-S3-S5-S6-S7-S8
AP_4	S0-S2-S5-S7-S8		

5.3 Risk Assessment

In order to calculate the probability of different atomic attacks, first calculate the corresponding attack cost. According to the scoring standard in Table 2, use Equation (3) to calculate the consumption cost of each atomic attack in the attack graph, as shown in Figure 6: The calculated attack cost, the exploitable probability of vulnerability shown in Table 5 and the benefit of atomic attack in the experimental attack graph are brought into Equation (4), and the attack probability of each atom is calculated, as shown in Figure 7: Combined with the probability of each atom attack on the attack graph obtained in Figure 8, the conditional probability of each attribute node is calculated, and then the static reachable probability of each node is obtained by combining the conditional probability with the attack trajectory according to Algorithm 1, and the static risk assessment of the test network is carried out. The static reachability probability of node S_0 is initialized to $P(S_0) = 0.7$. After the attack target is determined as S8, the reachability probability of each attribute node in the attack graph is updated according to Algorithm 2 and Formula (6) proposed in Section 4.1.3, and the dynamic reachability probability of each node in the dynamic risk assessment attack graph is obtained. The static and dynamic reachability probability distribution of each node is shown in Figure 8: The intrusion risk of the network is significantly increased, and the reachability probability of the target node S_8 is also increased from [0.496 to 0.631], and the intrusion risk of the intermediate attribute nodes S_1 and S_7 is the highest, so measures need to be taken to update the host patch. Therefore, in the real network environment, the accuracy of dynamic evaluation method for network risk assessment is significantly higher than that of static evaluation method, which can provide a good support for administrators to carry out network risk management.

5.4 Attack Path

Use Algorithm 1 to search the attack graph as shown in Figure 5 and get 7 attack paths, as shown in Table 5.

The total reachable probability of each path in static attack graph and dynamic attack graph is calculated by Formula (6), as shown in Figure 9. It can be observed that in both static and dynamic models, the attack path AP_1 is at the highest risk of intrusion. When the target of the attacker is clear, the overall reachability probability of each path is improved, especially the attack path AP_2, through which the risk of invading node S_8 is close to AP_1. The data shows that the network risk has changed after the intention of the attacker is determined, and the dynamic risk assessment can analyze the network risk more accurately.

5.5 Method Comparison

In order to verify the superiority of this model, under the same network environment, the methods proposed by Gao [5] and Zhou [19] are compared with the experimental data of this method.

Figure 10 shows the dynamic reachability probability distribution of different attribute nodes in the three algorithms under the same network environment shown in Figure 6. The evaluation model of literature [5] and literature [19] also used Bayesian belief network to describe

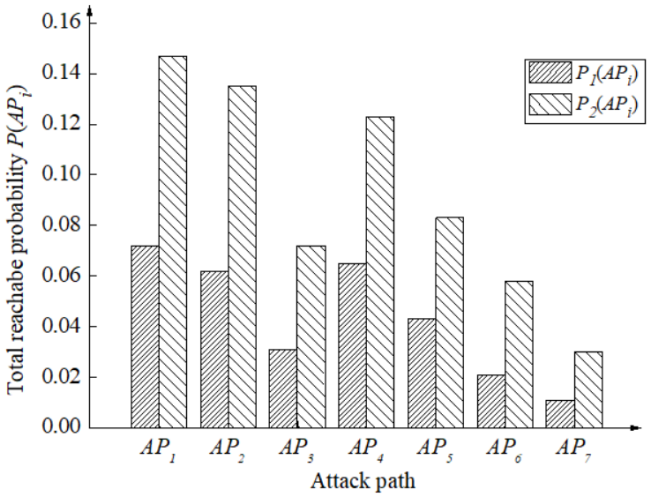

Figure 9: Total reachable probability of attack path

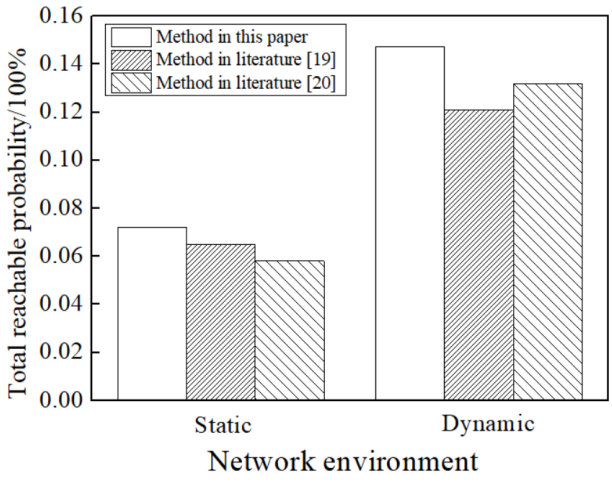

Figure 11: Comparison of predicted path reachability probability

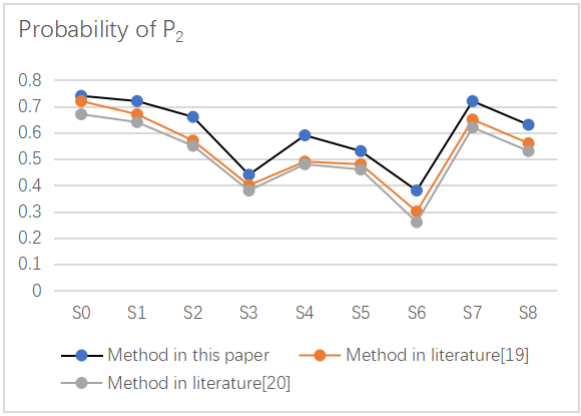

Figure 10: Comparison of dynamic reachability probability

the causal relationship between network attacks. However, due to its single evaluation index of vulnerability, and the cost and benefit of attacks are not considered, the vulnerability utilization ratio of the two does not really reflect the vulnerability in the network. It can be seen from Figure 10 that the accuracy of the evaluation model in this paper is significantly better than the two, because the evaluation is more accurate when the atomic attack probability is calculated from multiple indicators in this paper.

It is a further analysis of network risk to predict the attack path choice of attackers. In this paper, three algorithms are used to predict the attack path. In the same network environment, the attack path chosen by the attacker to the target node is predicted. Figure 11 shows the comparison of the total reachability probability of the three algorithms for the target node S_8 under the static network and the dynamic network respectively.Literature [5] and literature [19] only use the value of the vulnerability as the input parameter of the model, which reduces the valuation of the vulnerability to a certain extent. This article considers more evaluation indicators. In addition to the value of the vulnerability, it also takes into account the cost and benefit of the attack to obtain a more accurate vulnerability assessment probability, which is closer to the actual network attack.

Although all the three algorithms predict the attack path that may be used by the attacker, it is obvious that the overall probability of path reachability predicted by this algorithm is higher than the other two. This is because first of all, the quantification of the atomic attack probability is too single, which leads to the inaccuracy of the reachability probability calculation of the attribute nodes. Secondly, the method of predicting attack path is to start from the target node and continue to look up the attribute node with the highest probability, but ignore the influence of single node on the prediction of attack path.

6 Conclusion

This paper proposes a network intrusion intent analysis method based on Bayesian attack graph. First, the probability of atomic attack is calculated using the three evaluation indicators of vulnerability probability, attack cost and bene t. By using atomic attack probability to build a risk analysis model based on intrusion intent, quantify the static and dynamic probability of atomic attack. Second, use the risk analysis model to calculate the overall probability of each attack path to predict possible attacks. Finally, the reachability probability of the attribute nodes and the total reachability probability of the predicted path verify the superiority of the method. In real networks, the correlation between vulnerabilities also affects the probability of atomic attacks. Next, we will study this and optimize the network intrusion risk assessment model.

Acknowledgments

This study was supported by the Natural Science Foundation of Heilongjiang Province:LH2021F030. The authors gratefully acknowledge the anonymous reviewers for their valuable comments.

References

[1] A. Ahmadianramaki and A. Rasoolzadegan, "Causal knowledge analysis for detecting and modeling multi-step attacks," *Security and Communication Networks*, vol. 9, no. 18, 2017.

[2] K. Durkota, V. Lisý, B. Bošanský, C. Kiekintveld, M. Pěchouček, "Hardening networks against strategic attackers using attack graph games," *Computers & Security*, p. 101578, 2019.

[3] Z. Fan, C. Chang, P. Han, D. Pan, *Generation Method of Attack Graph Based on Rete Algorithm*, Computer Engineering, Information Engineering University, 2018.

[4] L. Gao, F. Wang, N. Gao, Y. Mao, "Security hardening measures selection model based on improved ant colony optimization," *Computer Engineering and Application*, vol. 55, no. 7, pp. 105–112, 2019.

[5] N. Gao, L. Gao, Y. He, Y. Lei, and Q. Gao, "Dynamic security risk assessment model based on bayesian attack graph," *Journal of Sichuan University(Engineering Science Edition)*, 2016.

[6] M. Ghasemigol, A. Ghaemi-Bafghi, and H. Takabi, "A comprehensive approach for network attack forecasting," *Computers & Security*, vol. 58, pp. 83–105, 2016.

[7] H. Hao, H. Zhang, and Y. Yang, "Security risk situation quantification method based on threat prediction for multimedia communication network," *Multimedia Tools and Applications*, vol. 77, no. 11, pp. 21693–21723, 2018.

[8] A. Hsl, B. Kd, and B. Jb, "A review of attack graph and attack tree visual syntax in cyber security - sciencedirect," *Computer Science Review*, vol. 35, 2020.

[9] H. S. Lallie, K. Debattista, J. Bal, "Evaluating practitioner cyber-security attack graph configuration preferences," *Computers & Security*, vol. 79, pp. 117–131, 2018.

[10] J. Lee, D. Moon, I. Kim, and Y. Lee, "A semantic approach to improving machine readability of a large-scale attack graph," *The Journal of Supercomputing*, vol. 75, no. 4, 2018.

[11] C. Lei, D. H. Ma, H. Q. Zhang, Y. J. Yang, and M. Wang, "Performance assessment approach based on change-point detection for network moving target defense," *Journal on Communications*, vol. 38, pp. 126–140, 2017.

[12] Y. Li, C. Z. Wang, G. Q. Huang, X. Zhao, B. Zhang, Y. C. Li, "A survey of architecture and implementation method on cyber security situation awareness analysis," *ACTA Electronica Sinica*, vol. 47, no. 4, pp. 161–179, 2019.

[13] Z. Y. Luo, X. Yang, G. Sun, Z. Q. Xie, J. H. Liu, "Finite automaton intrusion tolerance system model based on Markov," *Journal on Communications*, vol. 40, no. 10, 2019.

[14] C. Phillips, L. P. Swiler, "A graph-based system for network-vulnerability analysis," in *Proceedings of Workshop on New Security Paradigms (NSPW'98)*, pp. 71–79, 1998.

[15] A. Shameli-Sendi, M. Dagenais, L. Wang,, "Realtime intrusion risk assessment model based on attack and service dependency graphs," *Computer Communications*, vol. 116, pp. 253–272, 2018.

[16] H. Wang, T. W. Dai, X. X. Ru, Y. L. Lou, S. Ao, "Prediction method of node attack path based on optimized-ag," *Journal of Jilin University (Science Edition)*, vol. 057, no. 004, pp. 917–926, 2019.

[17] J. Wang, Y. Feng, R. You, "Network security measurment based on dependency relationship graph and common vulnerability scoring system," *Journal of Computer Applications*, vol. 39, no. 6, pp. 1719–1727, 2019.

[18] S. C. Wu, W. Q. Xie, Y. X. Ji, S. Yang, Z. Y. Jia, S. Zhao, Y. Q. Zhang, "Survey on network system security metrics," *Journal on Communication*, vol. v.40;No.386, no. 06, pp. 18–35, 2019.

[19] Y. Zhou, G. Cheng, and C. Guo, "Risk assessment method for network attack surface based on bayesian attack graph," *Chinese Journal of Network & Information Security*, vol. 4, no. 6, 2018.

Biography

Luo Zhiyong biography.Luo Zhiyong, male, was born in Shandong Province, China in July 1978.He is a professor at the School of Computer Science and Technology, Harbin University of Science and Technology. Research direction: computer network and information security, network optimization. words.

Xu Rui biography. Xu Rui, female, was born in Henan, China in January 1997. She is a postgraduate student in the School of Computer Science and Technology, Harbin University of Science and Technology. Research direction: computer network and information security, network optimization, offensive and defensive games. optimization.

Wang Jianming biography. Wang Jianming, male, was born in February 1997 in Jiangsu Province, China. He is a postgraduate student in the School of Computer Science and Technology, Harbin University of Science and Technology, Research direction: network security.

Zhu Weicheng biography.Zhu Weicheng is a lecturer in the School of Computer Science and Technology, Harbin University of Science and Technology. Research direction: network security.