Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume **5**, Number **4**, November **2020**

Node Confidence Calculation Method Based on D-separation of Bayesian Network

Hui Wang*, TianWang Dai, Kun Liu, XinXin Ru and YaLong Lou

School of Computer Science and Technology
Henan Polytechnic University
Jiaozuo, Henan, 454000, China
*Corresponding author:wanghui_jsj@hpu.edu.cn

Received December 2018; revised December 2019

Abstract. *The current node confidence calculation method for Bayesian network has the problem of node confidence miscalculation caused by miscalculation of conditional probability and correlation of nodes. The problem reduces the accuracy of node confidence and impacts the effectiveness of prediction on propagation paths of network threats. Therefore, we present a node confidence calculation method based on d-separation theorem of Bayesian network. First, by analyzing the correlation between attack cost and possibility of attack activity occurrence, we propose an approach for calculating the conditional probability so as to solve the problem of miscalculation in conditional probability. Secondly, by introducing separation theorem, we propose node confidence calculation method to effectively avoid miscalculating node confidence caused by the correlation. Finally, experiment result shows that our method effectively solves the miscalculation problem of node confidence and improves the accuracy of node confidence, consequently it achieves the effective prediction on propagation paths of network threats.*
Keywords: Node confidence; Conditional probability; Correlation; d-separation

1. **Introduction.** In recent years, network security has greatly aroused people's attention. One important reason is that computer networks are at constant risk from cyber attacks, which are becoming increasingly severe and sophisticated[1]. Nowadays, most mature defense technologies, such as firewalls, IDS are passive defense technologies. Based on predetermined solutions, they can protect against known security threats which are detected. To protect target networks, we need actively defend possible unknown threats so that administrators can control security threats before the loss occurred. An important way on preventing unknown security threats is to predict the propagation path of network threats.

Bayesian network uses nodes and directed arcs to describe the dependence of attack activity and attack evidence, and uses probability to describe the uncertainty relationship between nodes. Bayesian network has the ability to deal with the uncertainty relationship [4]. Meanwhile, Bayesian network itself is the kind of causality graph [5].Considering the propagation path of network threats is uncertain, Bayesian network can be used as an effective method for the analysis of network threats. However, this method needs to solve the quantification problem of node probability [6].Currently, the more mature achievement of this standardized work is common vulnerability assessment system(CVSS) [7,8].

At present, the probability calculation of security threats for Bayesian network is mainly based on the dependence and conditional independence assumption. In other words, network node only has the correlation with its parent nodes, while it is independent from

other nodes. Figure 1 is an example of propagation path for network threat based on Bayesian network, the ellipse nodes in the graph represent the resource state nodes of target network, and directed arcs represent attack steps. An attacker obtains the destination host permission by changing the status of a series of resource state nodes.

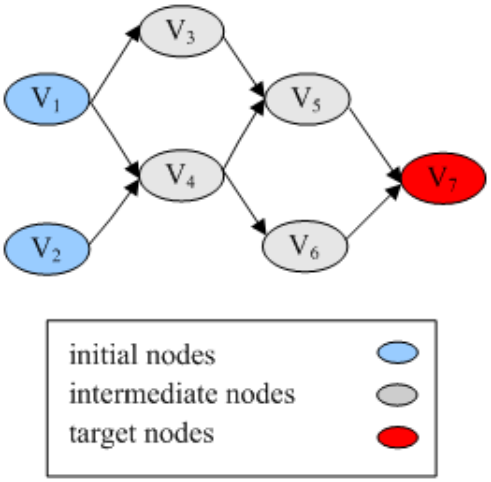

FIGURE 1. An example of propagation path for network threat

In figure 1, V_7 is the target resource state node. In order to occupy V_7, the attacker can use two different paths:(1) $V_1 \to V_4 \to V_5 \to V_7$ (2) $V_1 \to V_4 \to V_6 \to V_7$. In order to predict the propagation path, we need to calculate the probability of each node. The probability of V_7 can be obtained by calculating the probability of V_5 and V_6. Under the assumption of conditional independence, we assume that V_5 and V_6 is completely independent.

In the specific scene of propagation paths of network threats, the cost paid by an attacker is different due to the complexity and importance of the target node. Therefore, the cost will become an important factor for attacker and can change the conditional probability which results the miscalculation of node confidence, and affect the prediction on propagation path of network threats.Accounting for the above problem, the paper proposes a method of predicting the propagation path of network threat. The method can avoid the effect of correlation among nodes by introducing the d-separation theorem of Bayesian network. Besides, it can solve the problem of node confidence miscalculation caused by the attack cost by analyzing the effect of attack cost on attack activity. Consequently, it improves the accuracy of prediction on propagation paths of network threats effectively.

2. **Related Research.** Network security threat propagates in target network mainly through security vulnerability and propagation path has the great uncertainty. We can predict the propagation path by calculating the likelihood of occurrence of attack activity and the probability of security vulnerability utilized successfully. Based on the idea, paper [2,3] tried to calculate the probability of vulnerability which was utilized successfully in view of the Bayesian network and CVSS; Zhang et al.[10] proposed a complex attack prediction method based on fuzzy hidden Markov model. Alhomidi et al.[11] proposed a method to reason and predict intrusion by exploiting vulnerabilities.

In the scenario of specific propagation paths of network threats, the correlation between nodes and the complexity and importance of the target nodes all lead to the miscalculation of node probability and influence the prediction of propagation path. While it

was not considered in paper [9-12]. Homer et al [6]proposed that for the miscalculation of node probability there was a correlation between nodes and gave a solution to this problem based on the d-separation theorem. He did not consider the effect of conditional probability on node probabilities, and the method of adding nodes made the understanding of the attack graph more complicated. Fadlallah et al.[12] developed a tool which interfaces with the SNORT and matches the alerts with an attack graph generated using the NESSUS and the MULVAL attack graph generation library. Li et al.[13] proposed a forward search attack graph generation algorithm, which generating attack graph by searching from the target node to the attacker. Compared with search from attacker to the target node, the additional resources needed to store the former state of attacker are reduced, and the attack map generation efficiency of the large-scale complex network is improved.

Above possible problems, the paper proposes a method to solve miscalculation about node confidence and improve the accuracy of prediction. Therefore, the paper introduces Bayesian network separation theorem and proposes methods to separately calculate node confidence and conditional probability, giving three algorithms.In this paper, we propose an approach for calculating the conditional probability of attack activity occurrence by quantifying attack cost to solve the miscalculation problem of conditional probability caused by attack cost; By introducing the separation theorem and the method of calculating node confidence, making the associated nodes be independent of each other under the condition of having the common d-separation.

3. **Bayesian Attack Graph and Node Probability Calculation of Security Threat.**

3.1. **Bayesian Attack Graph. Definition 1**: Bayesian attack graph G = (V, Q, W, P, Π) is a Bayesian network which contains one or more AND - OR nodes, where:

• $V(G) = \{V_0 \cup V_G\}$ is a set of resource state node which contains non-empty-limited AND-OR nodes. Nodes represent the needing possessed resources when system is attacked. When a resource state is successfully changed, the corresponding value of node is True, otherwise it is False. The initial node set V_0 represents that attacker occupied resources with the certain probability under the consideration of initial state. Target node set V_G represents that attacker can reach the final destination node set after he succeeded in changing a series of resources state.

• $Q(G) \subseteq (V \times V)$ is a directed edge set which nodes are associated with each other in. If $q_{1,2} =<V_1, V_2> \in Q$ represents a directed edge from node V_1 to its children node V_2, then V_1 is called the precursor node of V_2, V_2 is called the successor node of V_1. q is used to represent an attack step. When an attack occurring, the value of q is True, otherwise is False. For $\forall q \in Q(G)$, each implemented attack step consumes some certain attack cost. Normally, this paper uses Prq(V) to represent the precursor node set of V, and Con(V) to represent the successor node set of V.

• W is used to represent the attack weight set. $\forall w \in W$,each node is represented by a 2-tuple (h,m).When V_i is attacked, h is total attack cost on the selected attack path, m represents the total attack cost on all possible attack paths passing the node V_i.

• $P = (P_1 \cup P_2)$, where $P_1$is the conditional probability of attack activity occurrence. For any attack activity, it will happen only the corresponding precursor nodes meet the condition. So, P_1 = (attack activity occurred $|Prq(V_i)$ meets the condition),namely, $P_1 = \{P_2 :(Prq(V_i), V_i)\}\rightarrow$[0,1]. P_2 is the probability of successful attack step. For P_2, only if attack activity occurred, maybe the attack step successful,V_i =True. So,P_2 = (V_i=True—attack activity occurred), namely, $P_2 =\{P_2$:(attack activity occurred,V_i =True)$\}\rightarrow$[0,1].

• $\Pi(G) = \{\pi.\mathrm{V}\times\mathrm{V} \rightarrow [0,1]\}$ represents the distribution of node confidence in Bayesian attack graph.$\pi(V_i)$ represents the probability of successfully occupied node V_i.Due to the node V_0 has been occupied by attacker in initial state,$\pi(V_0)$=1.

Definition 2: $\forall V_i \in V$, AND node means the operation in all precursor nodes of V_i is AND operation, and the value of V_i is True only if all precursor nodes of V_i have successfully implemented attack activity.

Definition 3: $\forall V_i \in V$,OR node means the operation in all precursor nodes of V_i is OR operation, value of V_i is True only if any one precursor node of V_i has successfully implemented attack activity.

According to the above definition, we firstly use vulnerabilities scanning tools (such as X-Scan) to scan the existing security vulnerability of system resources to achieve all resource state nodes. Then, by analyzing dependent relationship among nodes, all directed edge will be identified and marked. Finally, the relationship between AND-OR nodes is shown in figure 2.

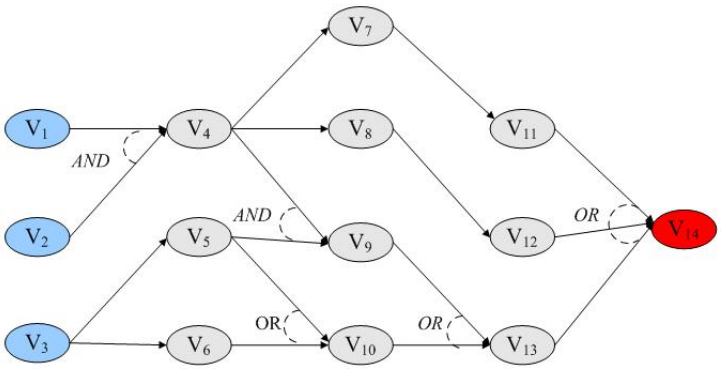

FIGURE 2. Bayesian attack graph

In figure 2, attacker possesses the resource state node V_1,V_2,V_3 respectively with a probability of $\pi(V_1)$, $\pi(V_2)$, $\pi(V_3)$,n the initial state. Assuming that attack activity will occur with the conditional probability distribution of P_1, attacker successively occupies the corresponding nodes $V_4 - V_{13}$, and will finally attain the target node V_{14}.

3.2. **Propagation Path of Security Threat. Definition 4**: Given the Bayesian attack graph, existing a transition sequence set of resource state named $Path$, where V_0 is starting node and V_G is ultimate target node in sequence set. $Path = V_0 \rightarrow V_1 \rightarrow V_2 \rightarrow \ldots V_i \rightarrow \ldots V_n \rightarrow V_G$, and 0≤i ≤ n.We define transition sequence satisfied the following conditions as propagation path of security threat:

1) V_{i+1} must be the successor node of V_i. While, V_i must be the precursor node of V_{i+1}.

2) The intersection of node set V_{i+1} of transition sequence $Path$ with the initial node set and the target node set cannot be empty.

3) The length of transition sequence $Path$ is limited.

In figure 2 , among all transition sequences in accordance with definition 4. Assuming that the network threat will propagate along the following *paths*:

1.$V_3 \rightarrow V_6 \rightarrow V_{10} \rightarrow V_{13} \rightarrow V_{14}$

2.$V_3 \rightarrow V_5 \rightarrow V_{10} \rightarrow V_{13} \rightarrow V_{14}$

3.$(V_1, V_2) \rightarrow V_4 \rightarrow V_8 \rightarrow V_{12} \rightarrow V_{14}$

Attacker occupies each node in *path* 1,2 and 3 successively, then:

1) If correlation between attack cost and node will not be considered, the value of all nodes is simply set to be True, attacker can choose one or more of three paths to carry

out attack theoretically, and the target node will be attained finally.

2) If considering the attack cost, attacker tends to execute the attack step $<V_9, V_{13}>$ when the attack cost is $Cost$ ($q_{10,13}$). That is to say, attack cost improves the possibility of executed attack $path2$ and reduces the possibility of executed attack $path1$ by affecting the occurrence probability of attack activity. **Definition 5**: Given the Bayesian attack graph G , the problem of propagation paths of network threat means that if the attack cost and the relativity of nodes are considered, the corresponding node confidence distribution sequence l in G is $l = \Pi_1, \Pi_1, \Pi_3, \ldots \Pi_i$.

4. The Calculation of Node Confidence.

Essentially, Bayesian network is a causality attack graph, and we can analyze multiple steps in network attacks, thus it can assist us to discover propagation paths of network threat. In order to predict possible propagation paths, all node confidence in Bayesian attack graph need to be calculated.

4.1. The Acquisition of Basic Data.

Bayesian attack graph G = (V, Q, W, P, Π) can present all propagation paths of network security threat. In order to determine the node confidence, the conditional probability of attack activity must be firstly obtained, namely p_1 and p_2. Then the node confidence can be calculated.

For p_2, its value depends on the difficulty level of resource state. So, the access complexity (AC) metric in Common Vulnerability Scoring System(CVSS) which describes how easy or difficult it is to exploit the discovered vulnerability. According to the standard of CVSS the value of p_2 is defined as follows:

Table 1. Basic metric of AC in CVSS

Base metrics	Value	Score(S(a_j))
access complexity (AC)	High (H)	0.35
	Medium (M)	0.61
	Low (L)	0.71

For p_1,the probability value of p_1 will be analyzed in the next section. because p1 will be affected by cost factor.

4.2. The Calculation of Attack Cost.

In the process of attack, the changing of resource state is implemented with a series of commands or operations. The attack cost of commands or operations may be different, but they are likely to be similar in function. In order to analyze the cost of resource state change, according to the similarity of functions, we divide commands and operations into different command set which is defined as Meta-operation.

Definition 6:Meta-operation is a set of commands or operations has similar function. The advantage of using Meta-operation as follows:

- Command and operation cost is easy to be gathered and convenient to calculate the attack cost.
- Selection of command or operation is limited to Meta-operation, it is easy to optimize attack cost.

Definition 7: For $\forall q \in Q(G)$, attack step is defined as $q_i = < Mos_c^i >$. $< Mos_c^i >$ is a set of command or operation constituted of the mapping qi on Meta-operation set Mos.

The definition 7 defines attack step from the view of set ,for $\forall q \in Q(G)$ in G, q is an attack step used to complete one resource state change, and it is a set composed of several commands or operations in accordance with a certain order. Figure 3 shows the mapping relationships:

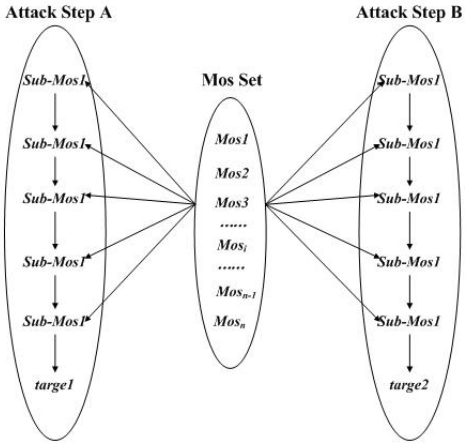

FIGURE 3. The mapping relationship on attack steps and Meta-operations

Sub-Mos is the mapping of Meta-operation set on attack steps. Every subset of Sub-Mos is associated with one function of attack step. The figure 3 shows that the composed attack step is different if operation sequence of the same Meta-operation set is different. Thus, an attack step should include Meta-operation cost and operation sequence cost and it is linear sum:

Definition 8:In Bayesian attack graph, $\forall q \in Q(G)$, the needed consuming cost of each attack step is:

$$Cost(q) = \mu \times Cost(Meta - operation) + \eta \times Cost(Sequence) \tag{1}$$

In definition 8,$Cost(Meta-operation)$ is the execution cost of Meta-operation, $Cost(Sequence)$ is the operation Sequence cost, and μ, η is the corresponding parameter. Because the value of $Cost(Meta - operation)$ depends on the $Meta - operation$ itself and the usage of resource, it is a function of $Meta - operation$ and resource number; the value of $Cost(Sequence)$ depends on order sequence of different $Meta-operation$, it is a function of operation Sequence and $Meta - operation$. But the calculation of $Meta - operation$ and operation sequence cost is still in the theory stage. For the convenience of discussion, the cost value of each attack step is set to 1.

Definition 9: For $\forall q \in Q(G)$, assuming that node d is the attack target, and the attack weight of d is $W(h, m)$, then the occurring conditional probability $P_1(q)$ of attack step $q$as follows:

(1) If $d \in V_0$, then $P_1(q) = 1$.

(2) If $d \in V$, $d \notin V_0$ the number of precursor nodes is s shown by node array $l[i]$, the corresponding attack step is $q[i](0 \leq i \leq s)$,then:

1) If d is neither the AND node nor the OR node, then

$$\mathrm{P_1(q[i])} = \frac{\mathrm{h(q[i])}}{\mathrm{m(q[i])}} \tag{2}$$

2) If d is the AND node, then

$$P_1(q[1]) = P_1(q[2]) = \ldots P_1(q[s]) = \frac{\sum_{i=1}^{s} h(q[i])}{\sum_{i=1}^{s} h(q[i]) + \sum_{i=1}^{s} Cost(q[i])} \tag{3}$$

3) If d is the OR node, Where $l[i]$,...$l[j]$,...$l[s]$ presents the corresponding attack step which is lined according to the order from big to small based on the cost m.

$$P_1(l[1]) = Min\{\frac{h(l[j]) + Cost(l[j])}{\sum_{j=1}^{s} m(q[j])} | j = 1, 2, \ldots s\} \tag{4}$$

Meanwhile, we can calculate the conditional probability of other precursor nodes as $P_1[l[2]]$,$P_1[l[2]]$,...$P_1[l[s]]$.

4.3. **The Calculation of Node Confidence.** Generally, the reason for miscalculation about node confidence in the Bayesian networks is that there is correlation among nodes. In order to solve the problem, the paper presents the following theorem.

Theorem 1: Assuming that there is any node set named $D = \{d_1, d_2, \ldots, d_k\}$ and $N = \{n_1, n_2, \ldots, n_j\}$ in Bayesian attack graph, where $D, N \subseteq V$.Then

$$P[N] = \sum_{D} P[N|D] \cdot P[N] \tag{5}$$

Proof: For any $D, N \subseteq V$, according to the Bayesian theorem. Then:

$$\sum_{D} P[N|D] \cdot P[N] = \sum_{D} P[N, D] = P[N] \tag{6}$$

Theorem 2: Assuming that there is any node set named $D = \{d_1, d_2, \ldots, d_k\}$ and $N = \{n_1, n_2, \ldots, n_j\}$ in Bayesian attack graph, where $D, N \subseteq V$, and each node in set V is independent in the case of all values of nodes in$D$ have been attained. Then

$$P[N|D] = \prod_{n_j \in N} P[n_j|D] \tag{7}$$

In order to calculate node probability $P(n_1, n_2, \ldots, n_j)$ of every node precisely, we find set D first and make nodes in N be independent each other. According to formula (2) and (3), calculates its node probability in the case of conditional independence.

The intersection of any two sets in set X, Y, Z is empty, d-separation in Bayesian network judges the conditional independence of set X and Y when set Z is determined. In order to find the node set D composed of independent nodes each other in set N ,the paper introduces the concept of d-separation.

Definition 10 For the empty intersection of node set X, Y, if there were connect nodes $v \in V$ between nodes $x \in X$ and nodes $y \in Y$, the set $D(D \subseteq V)$ that consists of v and d-separation of v ancestors separate X and Y.

According to definition 10, if the value of all nodes in set D has been known, and all elements in set X and set Y are independent each other, we sigh the situation as $X \perp Y|D$ and the conditional probability as $\varphi(\{X, Y\}, D)$. To narrate it conveniently, we command the formula as $N = X \cup Y$, then $\varphi(\{X, Y\}, D) = \varphi(N, D)$.

Definition 11 For the formula $\forall v \in V(G)$, the node set of d-separation is D, when the value of set D is True, probability is showed by $\lambda(v)$ else $\lambda(\overline{v})$.The single precursor node is z, we sign the following formula $\delta(z) = P_1[q[z]] \times P_2[q[z]] \times \lambda(z)$,

$\varphi_\delta(v, D) = P_1[q[z]] \times P_2[q[z]] \times \varphi(z, D)$,then:

$$\begin{cases} \lambda(v) = 1 - \lambda(\overline{v}) \\ \lambda(v) = \delta(z) = 1 - \delta(\overline{z}) \\ \varphi(v, D) = 1 - \varphi(\overline{z}, D) \\ \varphi_\delta(v, D) = 1 - P_1[q[z]] \cdot P_2[q[z]] \cdot \varphi(\overline{z}, D) \\ D(v) = z \cup D(z) \end{cases} \tag{8}$$

Definition 12 As the definition 11 defined that the logical relation among nodes is w. then,

(1)When w=AND,

$$\begin{cases} \lambda(v) = \delta(z) \\ \varphi_\delta(v, D) = \varphi_\delta(z, D) \\ D(v) = \bigcap\limits_{z \in Z} z \cup D(z) \end{cases} \tag{9}$$

(2)When w=OR,

$$\begin{cases} \lambda(v) = 1 - \delta(\overline{z}) \\ \varphi_\delta(v, D) = 1 - \varphi_\delta(\overline{z}, D) \\ D(v) = \bigcap\limits_{z \in Z} z \cup D(z) \end{cases} \tag{10}$$

The relation AND or OR which is considered among nodes will have effect on the calculated result of confidence, so the definition 12 is extended on the base of definition 11 to calculate node confidence when relation among nodes is AND or OR.

4.4. **The Computing of Node Confidence.** In order to calculate the node confidence of Bayesian attack graph, the paper design the following three algorithms.

Algorithm 1 is a computing algorithm of conditional probability to calculate occurrence conditional probability of attract step. For any node in Bayesian attack graph, the algorithm can calculate total consumptive attack cost h and total needed consumptive attack cost m on the selected attract step, and return the corresponding ratio bases on the node type. The ratio is the occurrence conditional probability of attract step. The specific calculation method refers to definition 9.

Algorithm 2 is the computing of conditional probability too to calculate the probability of conditional independence $\varphi_\delta(v, D)$ when set N is obtained successfully under the consideration of set N happening. Because of the formula $\varphi_\delta(v, D) = P_1[q[z]] \times P_2[q[z]] \times \varphi(z, D)$,therefore the algorithm includes some main steps when it satisfies the formula following as:1 the conditional probability named $\varphi(\{n\}, D)$ of single node n under the consideration of set D having different value; 2)according to the style of different node in set N, transforming format from $\varphi_\delta(v, D)$ to $\varphi(z, D)$ to calculate.(theorem 2 ,definition 11 and definition 12); 3)According to the above steps, returning to the independent conditional probability $\varphi_\delta(v, D)$.The specific calculation method refers to theorem 2 ,definition 11 and definition 12.

Algorithm 3 is the computing of node combined probability to calculate node confidence $\lambda(n)$.For each node is original node, then the value of $\lambda(n)$ is 1,else we divide it into two conditions according to the number precursor nodes: the number is over 1 or equal to 1. For the former condition, it is divided into two conditions to calculate $\lambda(n)$ according to the node set of d-separation is empty or not.

Algorithm 1 ConProb-Computing algorithm

Input: Bayesian attack graph G, attract cost $Cost$, attack weight $W(h,m)$.
Output: Occurrence conditional probability P_1 of attract step.

- IF initial node$\leftarrow v$
- THEN $\{W(h,m)=W(1,1)\}$
- RETURN $P_1(v)=1$
- END IF
- IF not initial node$\leftarrow v$
- $h \leftarrow SUM$
- $m \leftarrow h+SUM$
- END IF
- IF node AND$\leftarrow v$
- THEN $\forall d \in Prq(v)$,$P_1(q_{d,v}) = h/m$
- END IF
- IF node OR$\leftarrow v$
- THEN QUENE(r)=$\{r_1, r_2, \ldots r_n\}\leftarrow$value of each node $\{h(l(i)) + Cost(l(i))\}$
- QUENE(l)=$\{l_1, l_2, \ldots l_n\}\leftarrow$value of each node $\{h(l(i)) + Cost(l(i))/[\sum_{i=1}^{n} h(l(i)) + Cost(l(i))]\}$
- IF minimum value in QUENE(r) $\leftarrow r_1$
- maximum value in QUENE(l) $\leftarrow l_1$
- THEN QUENE$(r) \leftarrow \{r_1, r_2, \ldots r_n\}$
- THEN QUENE$(l) \leftarrow \{l_1, l_2, \ldots l_n\}$
- RETURN $P_1 = q_1$
- For $\forall d \in Prq(v)$, invoking the upper step
- RETURN $P_1 = l_1$
- END IF
- ELSE IF node v is neither node OR nor node AND
- THEN $P_1 = h/m$
- RETURN $P_1 = h/m$

Algorithm 2 InConProb-Computing algorithm

Input:Attack graph G, node set N, all node set D,P_1, P_2in the node set $N$of d-separation.
Output: $\varphi_\delta(v,D)$.

- $N \leftarrow \{n_0, n_1, \ldots n_q\}$
- $D \leftarrow \{d_0, d_1, \ldots d_k\}$
- IF the number of element in set N is more than 1
- THEN $\varphi(N,D) = \prod_{n\in N} \varphi(\{n\},D)$
- IF $N=\{\overline{n}\}$
- THEN $\varphi(\{\overline{n}\},D) = 1-\varphi(\{n\},D)$
- IF the elements in set N and set D are the same
- THEN $\varphi(\{\overline{n}\},D)=1$
- IF the value of element in set D is False
- THEN $\forall d_k \in D, \varphi(\{n\},\overline{d}_k) = 0$
- IF the nodes in set N are original nodes
- THEN RETURN $\varphi_\delta(v,D)=1$
- IF the nodes in set N are incomplete original nodes
- THEN Traversing and searching precursor nodes set Z of each node in set N
- IF node AND $\leftarrow n$
- THEN $\varphi_\delta(v,D) = \prod_{z\in Z} P_1(qz,n) \cdot P_2(qz,n)\cdot\varphi(\{z\},D)$
- RETURN $\varphi_\delta(v,D)$
- IF node OR $\leftarrow n$
- THEN $\varphi_\delta(v,D) = 1-\prod_{z\in Z} P_1(qz,n) \cdot P_2(qz,n)\cdot\varphi(\{\overline{z}\},D)$
- RETURN $\varphi_\delta(v,D)$
- IF the set N has the only one node
- THEN $\varphi_\delta(v,D) = P_1(qz,n)\cdot P_2(qz,n)$
- RETURN $\varphi_\delta(v,D)$

5. **Design and Analysis of Experiment.** In order to verify validity of calculation method about node confidence, the paper design a Small-scale LAN basis for system of Windows. The specific contents of experiment are as follows:

5.1. **Network Configuration of Experiment.** Figure 5 is the topology graph of experimental LAN. The experimental network makes firewall be the boundary and separates Internet from three safe region named M1 M2 M3 in LAN. The attacking host Attack

Algorithm 3 JoProb-Computing algorithm

Input:Attack graph G, node set N, all node set D,$P_1, P_2, \varphi_\delta(v, D)$in the node set $N$of d-separation.

Output: $\lambda(n)$.

- $N \leftarrow \{n_0, n_1, \ldots n_q\}$
- $D \leftarrow \{d_0, d_1, \ldots d_k\}$
- IF initial node $\leftarrow n$
- RETURN $\lambda(n) = 1$
- IF not initial node $\leftarrow n$
- THEN Traversing and searching precursor nodes set Z of each node in set N
- IF the set Z has the only one node
- THEN $\lambda(n) = P_1(qz, n) \cdot P_2(qz, n) \cdot \lambda(z)$
- RETURN $\lambda(n)$
- IF the set Z has many nodes
- THEN Searching the set D of d-separation n
- IF $D(Z) = \emptyset$
- THEN IF node AND $\leftarrow n$
- THEN RETURN $\lambda(n) = \prod_{z \in Z} P_1(qz, n) \cdot P_2(qz, n) \cdot \lambda(z)$
- THEN IF node OR $\leftarrow n$
- THEN RETURN $\lambda(n) = 1 - \prod_{z \in Z} P_1(qz, n) \cdot P_2(qz, n) \cdot \lambda(\overline{z})$
- IF $D(Z)$ is not empty set
- THEN IF node AND $\leftarrow n$
- THEN RETURN $\lambda(n) = \prod_{z \in Z} P_1(qz, n) \cdot P_2(qz, n) \cdot \lambda(\{z\}, D)$
- THEN IF node OR $\leftarrow n$
- THEN RETURN $\lambda(n) = 1 - \prod_{z \in Z} P_1(qz, n) \cdot P_2(qz, n) \cdot \lambda(\{\overline{z}\}, D)$

connects LAN by Internet.

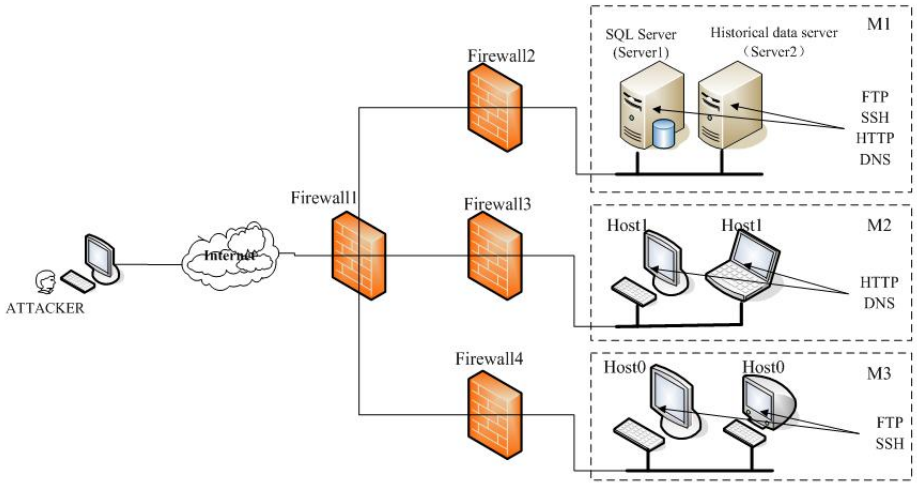

FIGURE 4. The topology graph of LAN

In LAN, the SQL sever Server1 and historical data server Server2 are set in region M1 and it is separated from other networks by Firewall 2. Server1 takes charge backing up data for Server 2 and supplies service named FTP SSH HTTP DNS and so on for each host in region M1 and M2, and improves service named SSH for Server2; In region M2, Host1 supplies service named HTTP and DNS; In region M3, Host0 supplies service named FTP and SSH for the target host stored key data.

5.2. **Data and Analysis of Experiment.** Using Scanner to scan the experimental network, we can get Vulnerability and vulnerability information of each host as the following table 2 shown. The utilized quantitative value of difficult degree in each vulnerability can

be inquired by attribute values of AccessComplexity in NVD database, and the query result is in table 2.

Table 2. The vulnerability information of each host

Security Domain	Host	Vulnerability	Complexity(P_2)
M1	Server1	CVE-2014-8595	0.61
		CVE-2011-4667	0.35
	Server2	CVE-2011-4955	0.61
		CVE-2012-6682	0.71
M2	Host1	CVE-2005-4900	0.35
		CVE-2014-7242	0.61
M2	Host0	CVE-2016-9722	0.71
		CVE-2011-2683	0.35

Utilizing proposed 3 algorithms in previous paper, we calculate node confidence of Bayesian attack graph in figure 4. First, we utilize the conditional calculation method of node confidence(no considering the attack cost and relationship of nodes) to calculate the node confidence in figure 4, and the calculation results is in table 3.

Table 3. The calculation results of conditional calculation method of node confidence

P_1 × P_2	$\pi(v_1)$	$\pi(v_2)$	$\pi(v_3)$	$\pi(v_4)$	$\pi(v_5)$	$\pi(v_6)$	$\pi(v_7)$	$\pi(v_8)$	$\pi(v_9)$	$\pi(v_{10})$	$\pi(v_{11})$	$\pi(v_{12})$	$\pi(v_{13})$	$\pi(v_{14})$
1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
0.2	1.0	1.0	1.0	1.0	0.2	0.2	0.2	0.2	0.008	0.078	0.04	0.04	0.17	0.10
0.6	1.0	1.0	1.0	1.0	0.6	0.6	0.6	0.6	0.22	0.59	0.36	0.36	0.68	0.87

In table 3, the node which is needed to define attack path is v_5,andv_6, v_9 and v_{10}, v_{11},v_{12} and v_{13}. Owing to the traditional calculation method does not consider attack cost and effect of correlation among nodes, so we assume that the conditional probability of resource node obtained by attacker is a fixed value.(The fixed value is set as 0.2,0.6,1.0 respectively).

Then we consider the effect of attack cost and correlation among nodes on node confidence, and utilize the suggested method to make an experiment get the calculation results of node confidence shown in table 4(i expresses the experimental number).

Table 4. The calculation results of the paper calculation method of node confidence

i	$\pi(v_1)$	$\pi(v_2)$	$\pi(v_3)$	$\pi(v_4)$	$\pi(v_5)$	$\pi(v_6)$	$\pi(v_7)$	$\pi(v_8)$	$\pi(v_9)$	$\pi(v_{10})$	$\pi(v_{11})$	$\pi(v_{12})$	$\pi(v_{13})$	$\pi(v_{14})$
1	1.0	1.0	1.0	0.13	0.31	0.36	0.06	0.04	0.02	0.34	0.04	0.02	0.22	0.14
2	1.0	1.0	1.0	0.13	0.31	0.36	0.06	0.04	0.01	0.32	0.04	0.04	0.2	0.12
3	1.0	1.0	1.0	0.13	0.31	0.18	0.6	0.6	0.04	0.28	0.04	0.04	0.17	0.09

When i=1, we assume that the consumptive attack cost in every attack path is 1 and $P_2(q_{3,6}) > P_2(q_{3,5})$, the result is that $\pi(v_6) > \pi(v_6)$ which definite the attack path which definite the attack path.

When i=2, we assume that keeping the other node condition unchanged while the successful conditional probability of attack path$q_{9,13}, q_{10,13}$ only changed.Because node v_{13} and v_{14} is related tov_9 and v_{10},so the confidence of node v_{13} and v_{14} varies with it. Comparing with i=1, this situation solves the effect which node correlation has on the calculation result of node confidence.

When i=3, we assume that the condition related to node v_5 does not change and increase the value of $Cost(q_{36})$, $Cost(q_{14})$ and $Cost(q_{24})$.The obtained value of $\pi(v_{10})$ based on this assumption is less than the value based on former two assumption.

Therefore, the proposed calculation method solves the miscalculation question of node confidence efficiently and definite the attack path.

5.3. **The Prediction of Attack Path.** For predicting the attack path, we introduce the cost-ratio parameter $C(X)$ to analyze the effect of attack cost on attack path selected by attacker.

According to the setting of parameter$C(X)$, we predict the probability of attack path implemented by attacker q_{35} or q_{36}, thus it generates confidence trend map of node v_{10} dynamically. Figure 5 shows the confidence trend map of node v_{10} changing with $C(X)$. X-axis represents cost-ratio C(X), Y-axis represents confidence, red line represents confidence $\pi(v_{10})$ of node v_{10},black line which is parallel to X-axis is threshold that represents what attacker want to occupy the node v_{10} by implementing attack path q_{36}, and the confidence of node must reach the minimum value. In figure 5, with the increasing of parameter

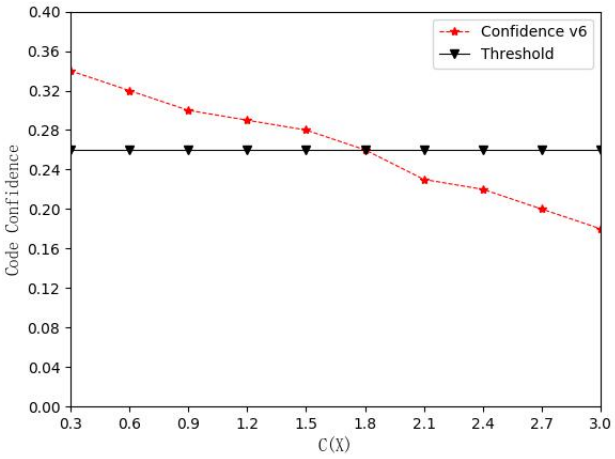

FIGURE 5. The prediction of attack path

$C(X)$,$\pi(v_{10})$ shows the tendency of decreasing gradually. The value of threshold is 0.26 which represents that attacker want to occupy the node 10 successfully by implementing attack path q36, and the minimum value of $\pi(v_{10})$ must reach 0.26.Thus, combining with parameter such as the 0.26 set by manager and warning value of cost-ratio, when attacker selects one path whose cost reaches warning value, the system can make an early warning and take the corresponding measure.

6. **Conclusion.** In order to solve the miscalculation question of node confidence caused by node correlation and conditional probability and improve the efficiency of prediction about propagation path of network security threat, the paper proposes a calculation method of node confidence.The contents proposed above not only makes up the deficiency about Homer paper and considers the effect of conditional probability on node confidence, but also proposes the calculation formula of conditional probability which is not proposed in literature[12-14].The further research work includes researching on the existing problem of propagation path of network security threat deeply and verifying the efficiency about proposed method in more complex network.

Acknowledgment. This project is supported by Research Fund for the Doctoral Program of Higher Education of China (No. 20124116120004), and supported by Educational Commission of Henan Province of China (No. 13A510325).

REFERENCES

[1] Ramos A, Lazar M, Filho R H, et al. Model-Based Quantitative Network Security Metrics: A Survey[J] *IEEE Communications Surveys & Tutorials*, PP(99):1-1, 2017.

[2] Idika N, Bhargava B. Extending Attack Graph-Based Security Metrics and Aggregating Their Application[J] *IEEE Transactions on Dependable & Secure Computing*, vol.70, 2016.

[3] Bopche G S, Mehtre B M. Attack Graph Generation, Visualization and Analysis: Issues and Challenges[C]// *International Symposium on Security in Computing and Communication. Springer, Berlin, Heidelberg*, 2014:379-390.

[4] Homer J, Zhang S, Ou X, et al. Aggregating vulnerability metrics in enterprise networks using attack graphs[J]. *Journal of Computer Security*, 21(4):561-597, 2013.

[5] Zhang S, Ou X, Homer J. Effective Network Vulnerability Assessment through Model Abstraction[J]. 6739:17-34, 2011.

[6] Homer J, Ou X M, Schmidt D. A sound and practical approach to quantifying security risk in enterprise networks[R]. *Technical report, Kansas State University*, 2009.

[7] Younis A, Malaiya Y K, Ray I. Assessing vulnerability exploitability risk using software properties[M]. *Kluwer Academic Publishers*, 2016.

[8] Younis A A, Malaiya Y K, Ray I. Using Attack Surface Entry Points and Reachability Analysis to Assess the Risk of Software Vulnerability Exploitability[C]// *IEEE, International Symposium on High-Assurance Systems Engineering*, IEEE, 2014:1-8.

[9] Tsu-Yang Wu, Chien-Ming Chen, King-Hang Wang,Chao Meng, Eric Ke Wang. A Provably Secure Certificateless Public Key Encryption with Keyword Search. *Journal of the Chinese Institute of Engineers*, DOI:10.1080/02533839.2018.1537807.

[10] Zhang Y X, Zhao D M, Liu J X, et al. Approach to Forecasting Multi- Stage Attack Based on Fuzzy Hidden Markov Model[J]. *Electronics Optics & Control*, 2015.

[11] Alhomidi M, Reed M. Attack Graph-Based Risk Assessment and Optimisation Approach[J]. *International Journal of Network Security & Its Applications*, 6(3):31-43, 2014.

[12] Fadlallah A, Sbeity H, Malli M, et al. Application of Attack Graphs in Intrusion Detection Systems: An Implementation[J]. 2016, 8(1):2016-2017.

[13] Li H, Wang Y, Cao Y. Searching Forward Complete Attack Graph Generation Algorithm Based on Hypergraph Partitioning [J]. *Procedia Computer Science*, 107:27-38, 2017.

[14] Chien-Ming Chen, Bin Xiang, King-Hang Wang, Kuo-Hui Yeh, Tsu-Yang Wu.A Robust Mutual Authentication with a Key Agreement Scheme for Session Initiation Protocol. *Applied Sciences*, 2018, 8, 1789; doi:10.3390/app8101789.