Acta Technica *62* (2017), No. 5A, 379–390

Information security situation assessment based on conditional evidential network

Sun yang[1], Xiong wei[1], Pei Dong[2]

Abstract. This paper proposes a method of information security situation assessment based on conditional evidential network, which is for the problem of false positives, omissions, inability to produce network security situation maps and the large number of uncertain information in network security assessment. Conditional reliability function is used to describe the conditional dependency of the attack state between nodes, while disjunctive rule of combination is used to implement evidential network reasoning, evaluating the security situation of the whole network. Experimental results show the proposed method can solve the adverse effects of false positives, omissions on the assessment of network security situation with less prior knowledge and network data, and deal with a variety of uncertain information such as probability and fuzzy knowledge.

Key words. conditional evidential network, network security situation, conditional reliability function, evidential network reasoning, Situation assessment.

1. Introduction

Information technology has become the driving force behind the development of modern society, and are changing our way of sharing and communicating. Ubiquitous computing and high-capacity data transmission have turned the Internet into the main medium for information exchange and e-commerce. With the rapid development of information technology and network, the border of national security has surpassed the limitation of geography space and expanded to the information network. Network security has become an important issue concerning national security. At present, the major countries in the world are entering the stage of ubiquitous network application. There is a chance for reform in the field of national Internet governance. At the same time, the scope and content of cyber security threats are constantly expanding and evolving, and the situation and challenges of network

[1]Workshop 1 - Key Laboratory of National Defense Science and Technology for Electronic Information Equipment System, Equipment Academy, Beijing, 100400, China
[2]Workshop 2 - Space Command Department, Equipment Academy, Beijing, 100400, China

security are becoming more and more complicated.

In response to cyber security threats, intrusion detection systems (IDS) came into being. Intrusion Detection System is a kind of network security equipment which monitors network transmission, and sends alarm or takes active reaction measures when finding suspicious transmission. It differs from other cybersecurity devices in that IDS is a proactive security technology. IDS is divided into network-based IDS, host-based IDS and distributed IDS. At present, IDS is developing rapidly, and some researchers have proposed that IDS can completely replace the firewall.

But IDS also has some flaws. Due to the rapid development of modern network technology, network transmission rate greatly accelerated, resulting in an increase in IDS workload, but also means that IDS detection of attack activity is not high reliability. At the same time due to the imperfect pattern recognition technology, IDS's high false alarm rate is also a major problem.

Aiming at the problems such as false positives and inability to form the network security situation in traditional intrusion detection systems, Bass used data mining methods, and evaluated the security situation of computer networks by data fusion of distributed sensors in intrusion detection system [1]. Poolsappasit [2] used the Bayesian network to evaluate the network security situation in real and dynamically time. However, none of these methods considered the issue of uncertainty in the assessment process.

Aiming to the above problems and we consider the uncertainties in the assessment process, and use conditional evidential network to describe the conditional dependence of the attacked states between nodes, and the evidential network reasoning is used to evaluate the network security situation.

2. Conditional evidential network

The reliability function theory is called recognition framework, denoted by Θ. Which includes a finite number of basic propositions,as a subset of Θ, and events in Θ must be mutually exclusive. Assignment $m : 2^{\Theta} \to [0, 1]$ is a basic probability assignment, if and only if formular below is satisfied [3-4]:

$$\begin{cases} \sum_{A\in\Theta} m(A) = 1 \\ bel(A) = \sum_{\emptyset\neq B\subseteq A} m(B), bel(\emptyset) = 0 \\ pl(A) = \sum_{B\cap A\neq\emptyset} m(B), pl(\emptyset) = 0 \\ m(A) > 0 \end{cases} \tag{1}$$

Then is a focal of the assignment. In the formula, the value of the confidence function bel is the reliability value of event, and the value of the likelihood function plis the maximum possible support for event.

Conditional evidential network models uncertainty with the conditional reliability function as the parameter [5-6].

Definition1: Let m be the basic reliability distribution on the identification frame-

work Θ, to$A, B \subseteq \Theta$, conditional basic reliabilityis defined as follows [7]:

$$m(B|A) = \begin{cases} \sum_{X \subseteq \overline{A}} m(B \cup X), B \subseteq A \subseteq \Theta \\ 0, otherwise \end{cases} \tag{2}$$

Definition2: Let bel be the confidence function on the recognition framework Θ, to$A, B \subseteq \Theta$, the conditional confidence function is defined as follows:

$$bel(B|A) = bel(B \cup \overline{A}) - bel(\overline{A}), \forall B \subseteq \Theta \tag{3}$$

Definition2: Let pl be the Likelihood function on the recognition framework Θ, to$A, B \subseteq \Theta$, the conditional likelihood function is defined as follows:

$$pl(B|A) = pl(A \cap B), \forall B \subseteq \Theta \tag{4}$$

DRC theorem: Suppose that the recognition framework of nodes X, Y are Θ_x and Θ_y, abbreviated as X and Y. Under normalized conditions [8-9], ie $bel_X(X|y_i) = 1, \forall y_i \in y$, for $\forall y_i \in \Theta_y, \forall x_i \in \Theta_x$

$$\begin{array}{l} m_X(x|y) = \sum_{(\bigcup_{i:y_i \in y} x_i) = x} \prod_{i:y_i \in y} m_X(x_i|y_i) \\ bel_X(x|y) = \prod_{y_i \in y} bel_X(x|y_i) \\ pl_X(x|y) = 1 - \prod_{y_i \in y} (1 - pl_X(x|y_i)) \end{array} \tag{5}$$

The reasoning of the conditional reliability function is divided into forward reasoning and reverse reasoning by extending Bayes theorem (GBT) and combination disjunctive rule (DRC) [10-11]. This paper uses forward reasoning.

Forward reasoning: If the reliability information of each state or subset of Y are known, denoted as$m_0(y)$ $y \subseteq Y$, then to $\forall x \subseteq X$, there exists[12]

$$m_X(x) = \sum_{y \subseteq Y} m_0(y) m_X(x|y) \tag{6}$$

The above formula is based on the conditional basic credibility, similarly the conditional reliability function and the conditional likelihood function can be expressed as

$$\begin{array}{l} Bel_X(x) = \sum_{y \subseteq Y} m_0(y) Bel_X(x|y) \\ Pl_X(x) = \sum_{y \subseteq Y} m_0(y) Pl_X(x|y) \end{array} \tag{7}$$

3. Network security situation evaluation model based on conditional evidential network

According to the characteristics of network security situation assessment, we establish network security situation assessment model based on conditional evidence network.

3.1. Attack recognition framework

This paper mainly studies the attacked state of network, host and node. The network consists of a series of hosts. Each host includes a series of nodes whose physical meaning can be understood as host vulnerability or vulnerability.

The attack recognition framework indicates the possible state in which the subject can be attacked: attacked or not attacked state, denoted as $\Theta = \{A, \bar{A}\}$, where A indicates attacked, $\bar{A}$indicates that it is not attacked. According to the different subjects, it can be divided into three types: node attack recognition framework, host attack recognition framework and network attack recognition framework.

3.2. Priori attack reliability of atomic attack nodes

The priori attack reliability of atomic attack nodes is assigned reliability by network security alert information in the network or information system, which is assigned by the detection support function. Detection support function represents warning message's support for node attack reliability, which is sent by detectors in the network or information system (such as IDS). So the detection support function is an important factor affecting the node attack reliability of this model.

In a typical attack graph, each alert may be associated with one or more nodes on the attack graph, and each node on the attack graph may also be associated with one or more alerts.

For each evidential node Evidencei, define the evidential reliability distribution function as follows:

$$\begin{cases} m(Evidence_i = A) = p_i \\ m(Evidence_i = \Theta) = 1 - p_i \end{cases} \tag{8}$$

Where pi is the alarm accuracy of the corresponding detector.

The attack reliability of atomic attack node E is calculated according to the Dempster synthesis rule.

Suppose m1 and m2 are two basic reliability distributions on the same recognition frame respectively. The focal elements of m1 and m2 are Xi i=1,2...,l and Yj j=1,2...,n respectively. Focus element $E \neq \emptyset$, and$E = X_i + Y_j$. then

$$[m_1 \oplus m_2](E) = \begin{cases} 0, & E = \emptyset \\ \frac{\sum_{X \bigcap Y = E} m_1(X_i) m_2(Y_j)}{1-k}, & E \neq \emptyset \end{cases} \tag{9}$$

Where $\oplus$represents direct sum $k = \sum_{Xi \bigcap Y_j = \emptyset} m_1(X_i) m_2(Y_j)$, and reflects the degree of evidence conflict. The greater the value of k, indicating that the greater the degree of evidence conflict. The coefficient$(1-k)^{-1}$ is called the normalization factor.

3.3. Conditional reliability function

The conditional reliability function indicates the conditional dependency between the preamble attack and the subsequent attack node attack reliability.

According to the concept of evidence network model, the conditional reliability function can be defined either by node or by edge. In this paper, in order to better characterize the conditional dependencies among multiple nodes, we choose to define the conditional reliability function by node and define it by with AND-OR structure respectively.

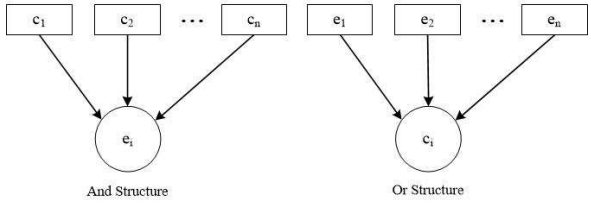

Fig. 1. AND-OR structure

In figure1, the left graph shows AND structure, when all ci are true, then ei is true. In the OR structure, the conditional reliability function is mainly determined by two factors. One is the attack difficulty coefficient $Dif(e_i)$ of the atomic attack node. The more difficult the attack, the less probability that the attack will succeed. Second, reliability transmission has attenuation. Suppose the attenuation coefficient of forward transmission is ρ_a, then the conditional reliability function can be defined as:

$$m(e_i = A|c_1 = A, c_2 = A, \cdots, c_n = A) = \rho_a Dif(e_i) \tag{10}$$

In figure1, the right graph shows OR structure, when either ei is true, then ci is true. When the atomic attack succeeds, the postconditions are subsequently obtained by the attacker, that is, when any ei is true, then ci is true. Therefore, the conditional reliability function of the structure is only determined by the transmission attenuation coefficient, that is,

$$m(c_j = A|e_1 = \bar{A}, \cdots e_i = A, \cdots, e_n = \bar{A}) = \rho_a \tag{11}$$

3.4. Node attack reliability and threat

On the basis of priori attack reliability and conditional reliability function of atomic attack nodes, the attack reliability of the entire network node can be obtained by forward reasoning.

To AND structure, the node attack reliability can be calculated by formula(12)

according to evidential theory:

$$m(e_i = A) = m(e_i = A|c_1 = A, c_2 = A, \cdots, c_n = A)Min(m(c_1), m(c_2), \cdots m(c_n)) \tag{12}$$

Wherein, $Min(m(c_1), m(c_2), ..., m(c_n))$ represents the minimum attack reliability of all conditional nodes

To OR structure, the node attack reliability can be calculated by formula(??) according to evidential theory:

$$m(c_j = A) = m(c_i = A|e_1 = A, e_2 = A, \cdots, e_n = A)Max(m(e_1), m(e_2), \cdots m(e_n)) \tag{13}$$

Wherein, $Max(m(c_1), m(c_2), ..., m(c_n))$ represents the maximum attack reliability of all conditional nodes

Node attack threat refers to the attack on the node and its impact or loss. For node v, its attack threat can be calculated according to formula (14)

$$Threat(v) = m(v)cost(v) \tag{14}$$

Where $m(v)$ is the node attack reliability, and $cost(v)$ is the loss of the node after attack.

3.5. Network security situation

For a host in the network, the attack threat it receives is determined by attack threats from all the condition nodes contained in the host. The attack threat value of each condition node on the host includes the attack reliability and the loss of the node. Therefore, we weight the attack threats of these nodes according to the node weights to obtain the attack threat of the host. The attack threat of the host is equal to the weighted sum of the attack threats of all nodes.

Let the weight of node $c_i(i = 1, ..., n)$in host w be $\theta_i(i = 1, ..., n)$, and $\sum_{i=1}^{n} \theta_i = 1$, and the host attack threat of w can be calculated as follows:

$$Threat_{host}(w) = \sum_{c_i \in Node(w)} \theta_i Threat(c_i) \tag{15}$$

Where $Node(w)$ represents set of all nodes on the host.

Since the host attack threat value of each host device in the network includes the probability of all attacks on the host and the losses caused by these attacks, we weigh the host attack threats of these host devices to obtain the threat of network attacks, That is, the threat of a network attack is equal to the weighted sum of attack threats of all the host devices in the network.

Let the weight of host $h_j(j = 1, ..., m)$ in network net is $\delta_j(j = 1, ..., m)$, and $\sum_{j=1}^{m} \delta_j = 1$.

Then, attack threats of the network can be calculated as formula(16) below:

$$Threat_{net} = \sum_{h_j \in H_{taget}} \delta_j Threat_{host}(h_j) \tag{16}$$

Where H_{target} represents key host set of the network.

Network security situation is defined as the ratio between the overall network threat value and the normalized loss incurred when all the nodes in the network are attacked. Then the network security situation can be calculated according to equation (17):

$$SA_{net} = \frac{Threat_{net}}{\sum_{j=1}^{m} \delta_j \sum_{i=1}^{n} \theta_i cost(c_i)} \tag{17}$$

Wherein, $\sum_{j=1}^{m} \delta_j \sum_{i=1}^{n} \theta_i cost(c_i)$ indicates the normalized loss generated when all the nodes in the network are attacked, θ_i andδ_j are node weight and host weight respectively.

4. Experiments results and analysis

In order to verify the effectiveness of this method, we evaluate the model experimentally. Experimental platform is on Windows8.1 64-bit operating system, Intel Xeon 2.73Ghz processor X2 and 24GHz memory. The network topology is shown in Figure 2. There are two hosts in the network. Host1 and host2 both include a series of vulnerabilities. These vulnerabilities are nodes in the conditional evidential network model. In the figure, the four-pointed star represents vulnerability with lower weight, the hexagonal star represents vulnerability with normal weight, and the sixteen-pointed star represents vulnerability with highest weight. An exclamation mark indicates alert from IDS, which is the evidential node in the conditional evidential network.

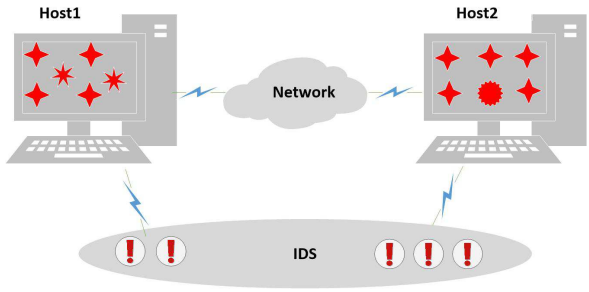

Fig. 2. Network topology

The attack graph condition evidential network is shown in Figure 3. Host 1 includes six nodes C1, C2, C3, C4, E1 and E2, and the host 2 includes six nodes C5, C6, C7, C8, C9 and E3. Asset weight of host 1 is 0.6, and asset weight of host 2 is 0.4. The initial attack reliability of each node and the loss after being attacked are:

m(v=A)=0??m(v=Θ)=1??cost(v)=100, where v is attack node in the network.

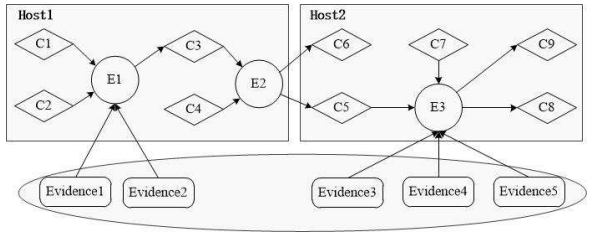

Fig. 3. Attack graph of condition evidence network

Define the node attack difficulty coefficient and weight as shown in table 1.

Table 1. Node attack difficulty coefficient and weight

Node	Dif (difficulty coefficient)	weight
C1	0.6	0.1
C2	0.6	0.1
C3	0.7	0.1
C4	0.7	0.1
C5	0.8	0.1
C6	0.8	0.1
C7	0.8	0.1
C8	0.9	0.1
C9	0.9	0.1
E1	0.65	0.3
E2	0.75	0.3
E3	0.85	0.5

When the detector detects a safety alert, five evidence nodes are generated. These evidence nodes reliability distribution of reliabilityare shown in table 3.

Table 3. Evidential nodes reliability

Evidence1	0.9	0.1
Evidence2	0.7	0.3
Evidence3	0.5	0.5
Evidence4	0.6	0.4
Evidence5	0.8	0.2

The attack reliability of atomic attack nodes E1 and E3 can be synthesized as follows according to the reliability synthesis rules:

The reliability of node E1 is synthesized by the reliability of Evidence1and Evidence2, and the reliability of node E3 is synthesized by the reliability of Evidence3, Evidence4and Evidence5

where evidence confliction coefficient k=0.9*0.3+0.7*0.1=0.34.

Similarly, according to the above formula

According to the direction of evidence transmission, the attack reliability of nodes E1, C3, E2, C5, C6, E3, C8 and C9 may be updated when the security alarm is generated. Then, the attack reliability and attack threats of each node is calculated according to the formula in section 2.4.

m(E1=A)=0.65;
m(C3=A)=m(C3=A|E1=A) m(E1=A)=0.9*0.65=0.585;
m(E2=A)=0;
m(C5=A)=0;
m(C6=A)=0;
m(E3=A)=0.86;
m(C8=A)=m(C8=A|E3=A) m(E3=A)=0.9*0.86=0.774;
m(C9=A)=m(C9=A|E3=A) m(E3=A)=0.9*0.86=0.774;
Then, the threat value of host 1 is calculated as follows;
$Threat_{net}(h_1)$=100 (0.3*m(E1=A) + 0.1*m(C3=A) + 0.3*m(E2=A)) =25.35
The threat value of host 2 is calculated as follows:
$Threat_{net}(h_2)$=100 (0.1*m(C5=A) +0.1*m(C6=A) +0.5*m(E3=A) +0.1*m(C8=A)+ 0.1*m(C9=A))= 86*0.5+77.4*0.1+77.4*0.1=58.48

Network threat is calculated as follows:

$$Threat_{net} = 0.6 \times Threat_{net}(h_1) + 0.4 \times Threat_{net}(h_2) = 38.602$$

Network security situation is calculated as follows:

$$SA_{net} = 38.602/100 = 38.602\%$$

The calculation results in this experiment show that when the IDS equipment issued the above five evidential nodes(Evidence 1-5) security alerts, then 38.602% of the network's assets are under attack threat.

Compared with the existing evaluation methods such as Bayesian Networks, Data Mining and HMM, the evidence network model used in this paper has the advantage of requiring less prior knowledge and data, and higher computational efficiency in larger network. Bayesian networks require large amounts of data for accurate probability determination and ambiguity estimation; data mining requires large amounts of data for pattern recognition; and Hidden Markov Models (HMMs) also require more data to determine implicit parameters of Markov processes. In addition, since the conditional evidence network can fuse the noisy information, the situation assessment results can still be well obtained in the case of IDS false positives. The conditional evidence network can reverse the inference, restore the attack scenario, and support processing IDS false negative problem. Therefore, compared with the

traditional evidence theory, the conditional evidence network model in this paper has the advantage of being able to solve the adverse effects of IDS false positives and false negatives on the assessment of network security situation.

5. Conclusion

In order to solve the problem of large number of uncertain information in the assessment of network security situation, this paper models and reasons the evaluation problems by conditional evidential network. Experiments results show that the proposed method can efficiently process network security situation evaluation problem with many kinds of fuzzy and uncertain information. Compared with the traditional Bayesian network method, this method does not need accurate probability judgment and fuzzy estimation and is less demand for prior knowledge and data. Advantages of algorithm in efficiency are quantified by experiments.

There are still some shortcomings in this method, mainly because the node attack difficulty coefficient and the evidence attenuation parameter in the conditional reliability parameter table still depend on the expert's experience, which is only suitable for the limited data in the initial stage of the problem. By accumulating data, we can deeply study evidence network model method based on data driven in future work.

References

[1] B. H. Narjes and B. Y. Boutheina: *Learning structure in evidential networks from evidential databases.* S. Destercke and T. Denoeux (2015), 301–311.

[2] L. Wafa and B. Y. Boutheina: *Propagation of belief functions in singly-connected hybrid directed evidential networks.* Beierle and A. Dekhtyar *98* (2015), 234–248.

[3] T. Bass, A. Arbor: *Multisensor data fusion for next generation distributed intrusion detection systems.* In: Proceeding of iris national symposium on sensor and data fusion (1999), 24–27.

[4] N. Poolsappasit, R. Dewri, I. Ray: *Dynamic security risk management using bayesian attack graphs.* IEEE Transactions on dependable and secure computing *9* (2012), No. 1, 61–74.

[5] B. B. Yaghlane, K. Mellouli: *Inference in directed evidential networks based on the transferable belief model.* International journal of approximate reasoning *48* (2008), No. 2, 399–418.

[6] F. Aguirre, M. Sallak, F. Vanderhaegen: *An evidential network approach to support uncertain multiviewpoint abductive reasoning.* Information Sciences *253* (2013) 110–125.

[7] H. Lee, J. S. Choi, R. Elmasri: *A static evidential network for context reasoning in home-based care.* IEEE Transactions on systems man and cybernetics-part a: Systems and Humans *40* (2010), No. 6, 1232–1243.

[8] E. Pollard, M. Rombaut, B. Pannetier: *evidential networks: an application to convoy detection.* Springer-Verlag (2010), 31–39.

[9] P. Smets: *The disjunctive rule of combination and the generalized bayesian theorem.* International journal of approximate reasoning *9* (1993), No. 1, 633–664.

[10] L. Wafa, B. Y. Boutheina: *Reasoning in singly-connected directed evidential networks with conditional beliefs.* SETN 2014 *2014*, 221–236.

[11] L. Wafa, B. H. Narjes, and B. Y. Boutheina: *Approximate inference in directed evidential networks with conditional belief functions using the monte carlo algorithm.* A. Laurent et al. (Eds.): IPMU 2014, Part III, CCIS 444 (2014), 486–497.

[12] B. H. Narjes and B. Y. Boutheina: *Learning parameters in directed evidential networks with conditional belief functions.* F. Cuzzolin (Ed.): BELIEF 2014 (2014), No. 4, 294–303.

Received November 16, 2017