Journal of Universal Computer Science, vol. 25, no. 3 (2019), 203-221
submitted: 28/3/18, accepted: 30/12/18, appeared: 28/3/19

Ontology and Weighted D-S Evidence Theory-Based Vulnerability Data Fusion Method

Xiaoling Tao
(Guangxi Colleges and Universities Key Laboratory of Cloud Computing and Complex Systems, Guilin University of Electronic Technology
State Key Lab of Integrated Service Networks, Xidian University, China
txl@guet.edu.cn)

Liyan Liu
(Guangxi Colleges and Universities Key Laboratory of Cloud Computing and Complex Systems, Guilin University of Electronic Technology, China
yiuyi.1990@qq.com)

Feng Zhao
(Guangxi Colleges and Universities Key Laboratory of Cloud Computing and Complex Systems, Guilin University of Electronic Technology, China
zhaofeng@guet.edu.cn)

Yan Huang
(Department of Computer Science, Georgia State University, USA
yhuang30@student.gsu.edu)

Yi Liang
(Department of Computer Science, Georgia State University, USA
yliang5@student.gsu.edu)

Saide Zhu
(Department of Computer Science, Georgia State University, USA
szhu5@student.gsu.edu)

Abstract: With the rapid development of high-speed and large-scale complex network, network vulnerability data presents the characteristics of massive, multi-source and heterogeneous, which makes data fusion become more complex. Although existing data fusion methods can fuse multi-source data, they do not consider that the multi-source data may affect the accuracy of fusion result. To solve this problem, we propose an ontology and weighted D-S evidence theory-based vulnerability data fusion method. In our method, we utilize ontology to describe the network vulnerability semantically and construct the network vulnerability ontology hierarchically. Then we use weighted D-S evidence theory to perform the operation of probability distribution and fusion processing. Besides, we simulate our method on MapReduce parallel computing platform. The experiment results show that our method is more effective and accurate compared with existing fusion approaches using single detection tool and traditional D-S evidence theory.

Key Words: data fusion, D-S evidence theory, network vulnerability, ontology
Category: G.1.0, I.6.0, I.6.4, J.2

1 Introduction

Network vulnerability data refers to the vulnerability information in the implementation of hardware devices, the security configuration strategy of the software, and the design process of protocols [Bishop and Bailey, 1999]. Network vulnerability data is a typical multi-source security data, which comes from different instrumentation tools and they are different in number and format. Researchers can evaluate and predict the network security situation vulnerability data analysis. Therefore, the vulnerability data is a main source for network security situational awareness.

While the vulnerability data can be used to predict the network security situation [Yu et al., 2018], with the rapid development of Internet technology, the vulnerability data presents the characteristics of massive, multi-source and heterogeneous [Alhazmi and Malaiya, 2006], which may make the predication more difficult. On the one hand, there is no uniform description for multi-source vulnerability data because they are scattered in different systems. Therefore, it is difficult to efficiently manage vulnerability data, which will increase the difficulty for judging the security incidents. On the other hand, the multi-source vulnerability data is collected from different instrumentation tools, whose ability to detect network threats is disparate. The traditional data fusion methods cannot analyze the multi-source vulnerability data simultaneously, thus failing to identify network threats and false negative alerts. Therefore, how to effectively manage and comprehensively analyze the vulnerability data is a serious challenge in network vulnerability data fusion.

Driven by various data fusion models, experts and researchers also apply different data fusion algorithms to the data fusion process of network security situational awareness to achieve the fusion of some specific situation indicators or situation data.Many scholars use clustering algorithm to correlate and fuse network security alarm events [Ning et al., 2002], [Julisch and Dacier, 2002]. They aggregated similar high alarm events by clustering algorithm and generated high-level events, which is very effective in similar and repeated alarm. However, they cannot effectively use the correlation timing relationship and causal relationship among alarms, and they also cannot identify complex attack scenarios. Almgren at al. [Almgren et al., 2008] used the Bayesian network to integrate the alarm of different subjects. The model considered the alarm quality of different sensors, and solved the conflict produced by multi-source alarm data. The drawback is that the prior probability model of every sensor is difficult to obtain.

In recent years, ontology is used to solve the network security situation awareness problems. Sadighian et al. [Sadighian et al., 2013] proposed an ontology-based alert correlation framework. They utilized ontology to describe and store the data of alarms, vulnerabilities and attacks. Then they used the logic rules of ontology to relate and filtrate the uncorrelated alarms flexibly. However, their method is not very practical. Besides, D-S evidence theory is always used in network security situation awareness filed. In 2013, Liu et al. [Liu et al., 2013] used D-S evidence theory to propose a multi-source information fusion method for network situation assessment, which pays attention to flow traffic capturing and analysis. However, it lacks the ability of analyzing and assessing other security attributes, such as vulnerabilities and threats. Zhong et al [Zhong and Zhao, 2012] proposed a D-S evidence theory-based vulnerability data fusion method in 2012, nevertheless, their method does not consider the credibility of different detection tools.

Despite there are plenty of works on network vulnerability data fusion by utilizing different data fusion algorithms, most of them focus on the data integrated approach and data analytical model. So, there are lack of suitable solutions for data semantic isomerism, and few work pay attentions to application of ontology description in network security situation awareness. At the same time, most of the existing work research the dispose of the traffic data and alarm data, only few attention are paid to the network vulnerability data fusion methods. As far as we know, the existing schemes do not consider the differences between the network situation data that are from multi-probe tools. Therefore, we propose an ontology and weighted D-S evidence theory-based vulnerability data fusion method.

Our contributions. In this paper, we propose a novel ontology and weighted D-S evidence theory-based vulnerability data fusion method. Our method gives different confidence values to multi-source data, which will make the fusion results more correct and comprehensive. The main contributions of this paper are as follows:

- We propose an ontology and weighted D-S evidence theory-based vulnerability data fusion method. Our method can describe the multi-source data uniformly through constructing a vulnerability data ontology, which can solve the problem of inconformity of the multi-source data.

- We introduce a weight-based D-S evidence theory method to fuse multi-source vulnerability data. We can solve the problem that the multi-source data affect the fusion results by giving different confidence values to multi-source data, which can make the fusion results more precise and efficient.

1.1 Related work

Data fusion technology dates back to 1970s, it was mainly engaged in the military at the beginning. With the rapid development of data fusion technology, it is gradually extended to civilian. Nowadays, data fusion has been widely used in plenty of fields, such as urban mapping [Gamba, 2013], forest-related studies [Delalieux et al., 2014], oil slick detection and characterization [Fingas and Brown, 2014], disaster management [Dell'Acqua and Gamba, 2012], remote sensing [Mura et al., 2015] and so on.

The network security situation is considered that it has the ability to integrate multi-source information, and it has become a hot topic in recent years. In traditional network security situational awareness methods, D-S evidence theory is often used to fuse information which comes from different security devices. In 2005, Yu et al. [Yu et al., 2005] put forward an alarm information fusion method, which is based on the weighted D-S evidence theory. Their method can improve the reliability of security events, and reduce the false alarm by giving different sensor configurations corresponding confidence and weights. In 2011, based on the importance of network topology, host and service, Zhang et al. [Zhang et al., 2011] adopted the D-S evidence theory to integrate the multi-source data submitted by multi-sensor.

In the literature [Wang et al., 2012], the dynamic adjustment strategy of Agent weight integrated into D-S classification optimization was introduced. Chen et al. [Chen and Feng, 2014] used the evidence distance to obtain the weight of different evidences to deal with the conflict of evidence in the D-S evidence. Huang [Huang, 2015] improved traditional D-S evidence theory, and proposed a time window D-S evidence theory in 2015. The novel time window D-S evidence theory is more accurate than the traditional D-S, and can achieve good result in network attack detection.

Recently, ontology as a tool of knowledge expression [Haug et al., 2013] has become a research hotspot. Researchers have paid a lot of attentions to the ontology, such as model consistency [Oellrich et al., 2015], logical consistency [Azevedo et al., 2015] and relational consistency [Ebrahimipour and Yacout, 2015]. In recent years, ontology has been applied to the field of network situation awareness. Ontology provides a unified conceptual interface and a rich semantic description for heterogeneous data, and it is independent of the data mode. Bhandari and Gujral [Bhandari and Gujral, 2014] presented an ontological approach to perceive the current network security status, which may be used to infer impact of various events happening in the network. Si et al. [Si and Zhang et al., 2014] established an ontology-based network security situation elements fusion model. In their scheme, the concepts and objects of situation elements were described uniformly by web ontology language (OWL) and Semantic Query-enhanced web rule language (SQWRL) inference rules. They achieved good integration effect by improving the complementary and reducing the redundancy.

1.2 Organization

The rest of the paper is organized as follows: we define some preliminaries in Section 2, and we give our overall system framework in Section 3. An vulnerability ontology construction method is described in detail in Section 4, and an vulnerability data fusion method based on weighted D-S evidence theory is presented in Section 5. Then the experiment results are presented and analyzed in Section 6. Finally, we give a brief conclusion of the paper in the last section.

2 Preliminaries

To have a better understanding of the method proposed, we will introduce some preliminaries in this section.

Dempster-Shafer evidence theory. Dempster-Shafer (D-S) evidence theory is a very useful method for dealing with incomplete, uncertain, and unclear data or information. D-S evidence theory is originally proposed by Dempster [Dempster, 1967] in the mid-1970s, then Shafer extended it to a complete set of mathematical reasoning theory [Shafer(76)]. It provides a good method for the expression and synthesis of uncertain information and introduces confidence interval and the belief function to remove the dependence on priori information. Besides, it uses interval estimation to describe uncertain information and D-S evidence theory to fuse multi-source data, which can increase the credibility. We will give some basic concepts of D-S evidence theory in the following.

- **The frame of discernment** $\varTheta$. We assume that there is a problem need to be judged, and all the possible results are putted in a set $\varTheta$. If all elements of the set $\varTheta$ are pairwise mutually exclusive, we call the incompatibility event set $\varTheta$ the frame of discernment.

  $2^{\varTheta}$ is all propositions corresponds to the set of all subset of $\varTheta$. $\forall A \subseteq \varTheta$, A is an element of $2^{\varTheta}$, A is called the event on the frame of discernment $\varTheta$. Evidence theory is computed on the elements of $2^{\varTheta}$.

- **Basic probability assignment function**. Let $\varTheta$ be a frame of discernment, the basic probability assignment function m is a mapping function $2^{\varTheta} \rightarrow [0, 1]$, if m satisfies:

$$m(\varPhi) = 0 \tag{1}$$

$$\sum m(A) = 1 \tag{2}$$

  where, m is the basic probability assignment function on the frame of discernment $\varTheta$, $\forall A \subseteq \varTheta$, and $m(A)$ is the basic probability assignment of the event A.

– **Belief function**. Given a body of evidence with BPAF m, we can compute the total belief provided by the body of evidence for a proposition. This can be done by belief function Bel:

$$Bel(A) = \sum_{B \subseteq A} m(B), (\forall A \subseteq \Theta) \tag{3}$$

here, $Bel(A)$ is the total belief committed to A, that is, the mass of A itself plus the mass attached toall subsets of A.

– **Plausibility function**. Since the Belief function does not reflect the plausibility of A, we propose the Plausibility function to describe the uncertainty of A. $\forall A \subseteq \Theta$, we define

$$Pl(A) = 1 - Bel(\overline{A}) \tag{4}$$

Dempster's rule of combination represents the conjunctive operation of the evidence. Given several belief functions on the same frame of discernment based on the different evidences, if they do not entirely conflict, we can calculate a belief function using Dempster's rule of combination.

– **Dempster's rule of combination**. Assume that m_1 and m_2 are two independent basic probability assignment functions on the basic probability assignment Θ. The two pieces of evidence can be fused to produce a joint basic probability assignment function by the following equation,

$$m_1 \oplus m_2(A) = \frac{1}{1-K} \sum_{B \cap C = A} m_1(B) m_2(C) \tag{5}$$

when

$$k = \sum_{B \cap C = \Phi} m_1(B) m_2(C), k \neq 1 \tag{6}$$

Dempster's rule of combination satisfies the associative and commutative property. Combination rules of limited number of basic probability assignment functions are as follows:

$$(m_1 \oplus m_2 \oplus \ldots \oplus m_n)(A) = \frac{1}{1-K} \sum_{\bigcap_{i=1}^{n} A_i = A} m_1(A_1) m_2(A_2) \ldots m_n(A_n) \tag{7}$$

when

$$K = \sum_{\bigcap_{i=1}^{n} A_i = \emptyset} m_1(A_1)m_2(A_2)\ldots m_n(A_n), K \neq 1 \quad (8)$$

3 Architecture description

3.1 System model

To deal with the characteristics of the network vulnerability data, such as massive, multi-source and heterogeneous, we design a novel network vulnerability situation awareness model for cloud [Yu et al., 2017, Li et al., 2019]. Our model contains four layers: network vulnerability data acquisition layer, network vulnerability data ontology construction layer, network vulnerability situation data fusion layer and application service layer, as shown in Figure 1.

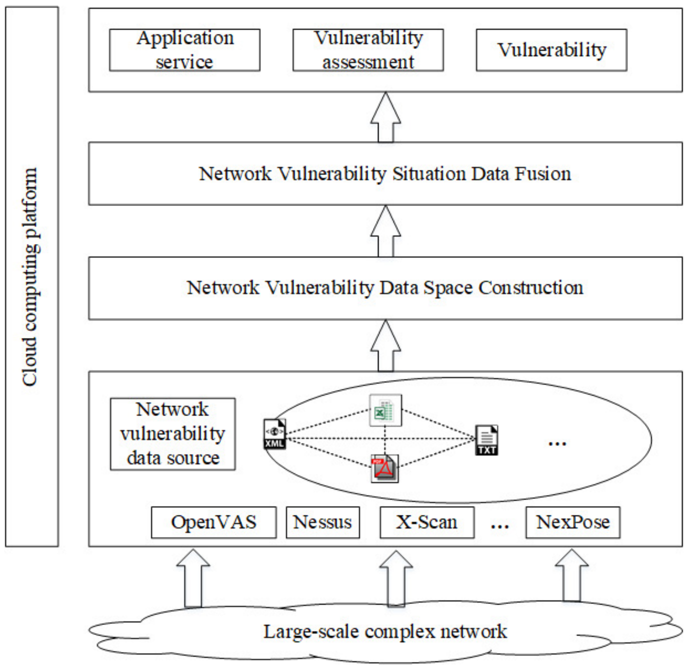

Figure 1: Cloud Vulnerability Assessment Model of Network Vulnerability

– **Network vulnerability data acquisition layer.** The network vulnerability data acquisition layer is the basic layer of the system model. The vulnerability data are different in terms of syntax, semantics and format, which are collected by a variety of scanning tools. And the detection tools are distributed over large-scale and complex networks.

- **Network vulnerability data ontology building layer.** As one of the core layers of the system model, this layer aims to classify and describe multi-source network vulnerability data by using ontology, and establishes the hierarchical relationship of resource description. Then the vulnerability data will be describe uniformly.
- **Network vulnerability situation data fusion layer.** This layer is another core layers of the model, the main task of this layer is to utilize the proposed vulnerability fusion method to integrate the multi-source vulnerability data, then we can obtain more comprehensive and accurate trend indicators of vulnerability.
- **Application service layer.** The application service layer is based on the network vulnerability data fusion, and mainly fulfills the task of application of the network situation-aware security service, such as assessing the network security situation, predicting the security layer of the network, and so on.

3.2 Design goals

In this paper, we design an ontology and weighted D-S evidence theory-based vulnerability data fusion method, which can describe and identify multi-source and heterogeneous network vulnerability data uniformly. The main design goals of our method are as follows:

1) We aim to utilize ontology to classify and describe the network vulnerability data uniformly, and then eliminate the inconsistency between the multi-source vulnerability data. Besides, we plan to establish the hierarchical relationship of resource description.

2) There are many differences among the vulnerability evidences which come from different scanning tools. We aim to increase correctness of the vulnerability data fusion results by introducing the weighted D-S evidence theory. And then, we will obtain a better result when there is conflicting evidence. So we can make the fusion results better reflect the real situation.

4 Vulnerability ontology construction method

4.1 Vulnerability ontology structure

We utilize the ontology to classify vulnerability data and establish the hierarchical relationship. By constructing the vulnerability ontology, on the one hand, we can make a clear definition of the concept of vulnerability; on the other hand, we can define the relationship among concepts clearly, and eliminate the inconsistency of heterogeneous data effectively.

In the process of constructing the ontology, VulProperty will record the associated attributes of the vulnerability, including the severity of the vulnerability, the name of the vulnerability, the generic vulnerability scoring system-CVSS, the summary of the vulnerability, and public vulnerability number-CVE. And VulVector will record the carrier information of the vulnerability, including software/operating systems, protocols, hosts and ports. Besides, Method will record the way of the vulnerability detected, Result will record the detection result of the vulnerability, Solution will record the solutions for the different vulnerabilities and P will record the probability of the vulnerability appears, Tool will record the tools used for scanning. The vulnerability ontology structure is shown in Figure 2.

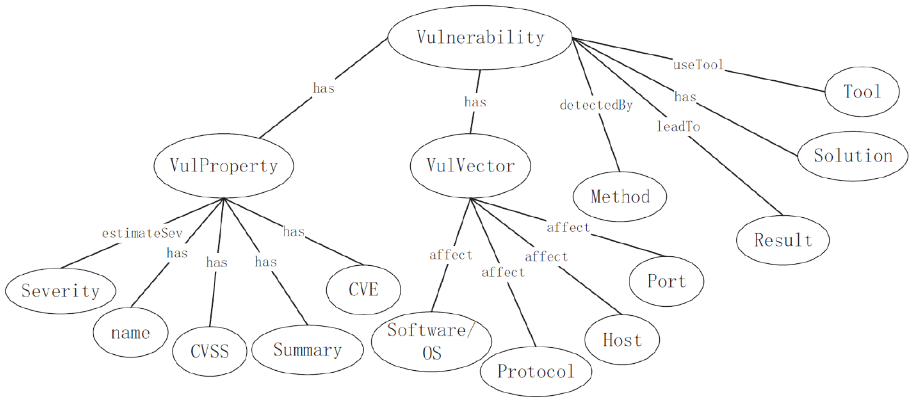

Figure 2: Vulnerability ontology structure

4.2 Vulnerability ontology description

Web Ontology Language (WOL) is an ontology description language, which is W3C recommended language standard for describing ontology, and we use it to describe vulnerability ontology. Compared with the XML, WOL is better for describing the relationship among concepts because it contains rich primitives. In our method, we describe the vulnerability ontology by OWL DL, which is a subset of OWL. As its strong expression ability and reasoning ability, OWL DL is widely used to describe semantic Web ontology.

OWL contains three relationships: inclusion, equivalence, and disjoint, denoted as subClassof, equivalentWith and disjointWith, respectively. We can show the links among the elements in the vulnerability ontology by hierarchically expressing the relationships between classes. For example, the relationship between the vulnerability detection tool and the vulnerability can be expressed as follows:

<owl:Class rdf:ID="Tool">
<rdfs:subClassOf rdf:resource="#Vulnerability"/ >
< /owl:Class>

It defines Tool as a subclass of Vulnerability, where the namespace"owl" represents OWL and"rdf" stands for RDFS.

Owl:ObjectProperty and owl:DatatypeProperty are two key primitives of OWL DL, in which object attributes (ObjectProperty) are used to illustrate the relationships between the two classes. For example, the object property detectBy defines the relationship between the Vulnerability class and the Method class as follows:

<owl:ObjectProperty rdf:ID="detectedBy">
<rdfs:subClassOf rdf:resource="#Vulnerability"/ >
<rdfs:range rdf:resource="#Method/ >
< /owl:ObjectProperty>

5 Weighted D-S evidence theory-based vulnerability data fusion method

Data consistency is the premise of data fusion. So we use the method of building vulnerability ontology to describe the multi-source vulnerability data uniformly which could eliminate the inconsistency between heterogeneous data and lays the foundation for the fusion of multi-source vulnerability data.

In the network security situation awareness system, the reliability of each detection tool is different, which can be determined by related background knowledge and experience. Next we will describe the details of the weighted D-S combination rules.

1) We assume that the number of the detection tools is r, and the relative weights of the evidence $E_1, \ldots, E_r$ are $w_1, \ldots, w_r$, which is determined on the basis of the detection ability of the detection tools. If $w_f = \max\{w_1, \ldots, w_i\}$, we call the E_f key evidence, and other evidences as non-critical evidences. Then $\beta_i = w_i/w_f, (i = 1, 2, \ldots, r)$ are the weights of each evidence against key evidence, and the fusion probability assignment satisfies:

$$m'_f(A_i) = m_f(A_i) \tag{9}$$

$$m'_h(A_i) = \beta_i m_h(A_i), A_i \neq \Theta, h \neq f \tag{10}$$

$$m'_h(\Theta) = \beta_i m_h(\Theta) + 1 - \beta_i, s \neq f \tag{11}$$

where $m_f(A_i)$ is the basic probability assignment of the maximum detection tool f, and $m'_f(A_i)$ is the weighting fusion probability assignment of f; $m_h(A_i)$ is the basic probability assignment of the non-critical evidence detection tool h, and $m'_h(A_i)$ is the weighting fusion probability assignment of h; $m_h(\Theta)$ is the unknown basic probability distribution of h, and $m'_h(\Theta)$ is the unknown weighting fusion probability distribution of h.

2) After computing the fusion probability, we can compute the fusion probability for different detection tools by using D-S evidence combination rules.

Then we can fuse the multi-source vulnerability data, specifically, we define the identification framework $\Theta = \{exit, unexit\}$, $exit$ indicates that the vulnerability exists, and $unexit$ indicates that the vulnerability does not exist. There is a vulnerability if the following rules are satisfied.

$$\begin{cases} m(exit) > \varepsilon_1 \\ m(\Theta) < \varepsilon_2 \\ m(exit) > m(unexit) \end{cases} \tag{12}$$

where ε_1 and ε_2 are pre-set thresholds, and the fusion architecture is shown in Figure 3.

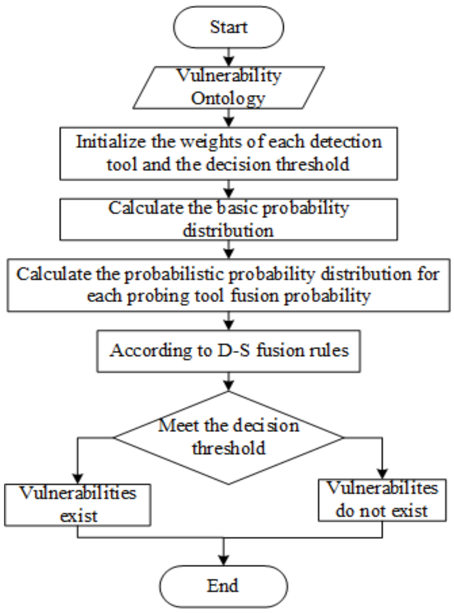

Figure 3: Vulnerability data fusion architecture based on weighted D-S evidence theory

The weighted D-S evidence theory-based vulnerability data fusion architecture is described as follows:

1) Input the constructed vulnerability ontology.
2) Initialize the each probe tool weight $E_1, \ldots, E_r$ and the pre-set thresholds $\varepsilon_1, \varepsilon_2$.
3) Calculate the basic probability assignment $m(exit), m(unexit)$ and $m(\Theta)$ according to each CVE number, different probe tools and result severity.
4) Calculate the fusion probability assignment according to the formulas (9), (10), (11).
5) Determine whether a vulnerability exists according to the formula (12). if pre-set thresholds are satisfied, then vulnerability exists, otherwise the vulnerability does not exist.

6 The experiment design and the result analysis

6.1 Experimental data

To verify the validity of the proposed ontology and weighted D-S evidence theory-based vulnerability data fusion method, we scan hosts with the help of Nessus and OpenVAS to obtain vulnerability data which includes Scan Information and Host Information. The composition of the data is shown in Table 1.

Table 1: The composition of data

Categories	Attributes
Host Information	IP
	MAC Address
	OS
Result Details	CVSS
	CVE
	Severity
	Solution
	Detection Method
	Port
	Protocol
	Vulnerability Name
	Summary

The scan reports of Nessus and OpenVAS give the vulnerability Severity, which contains 5 levels: Critical, High, Medium, Low and Info. Through semantic analysis of the Nessus scan reports, we can get the BPA allocation, as shown in Table 2.

Table 2: Nessus BPA allocation

Severity	$m(exit)$	$m(unexit)$	$m(\Theta)$
Critical	0.9	0.1	0
High	0.8	0.2	0
Medium	0.6	0.4	0
Low	0.4	0.6	0
Info	0.1	0.9	0

As the scan reports of OpenVAS give the probability of existence of each vulnerability is P, the probability that the vulnerability does not exist is 0, and the unknown probability is $1 - P$. Because P is around 0.7, So when an entry is present only in Nessus, the probability of constructing an OpenVAS corresponding entry with a basic probability distribution is 0.7, and the unknown probability is 0.3. The probability of constructing a flaw that exists only in OpenVAS is 0, the unknown probability is 0.3, and the probability of existence is 0.7.

6.2 Experimental Study

The specific configuration of the host and software used in the experiment are shown in Table 3.

Table 3: Configuration of host and software

Categories	Configuration
CPU	Intel 2.53GHZ
Memory	6GB
Operating System	Win7
Software	Eclipse 4.6.1

6.3 Experimental results and analysis

According to the overall framework of the proposed method, we first construct the vulnerability ontology by using the ontology editing tool Protégé 3.4.4, and the vulnerability ontology is shown in Figure 4.

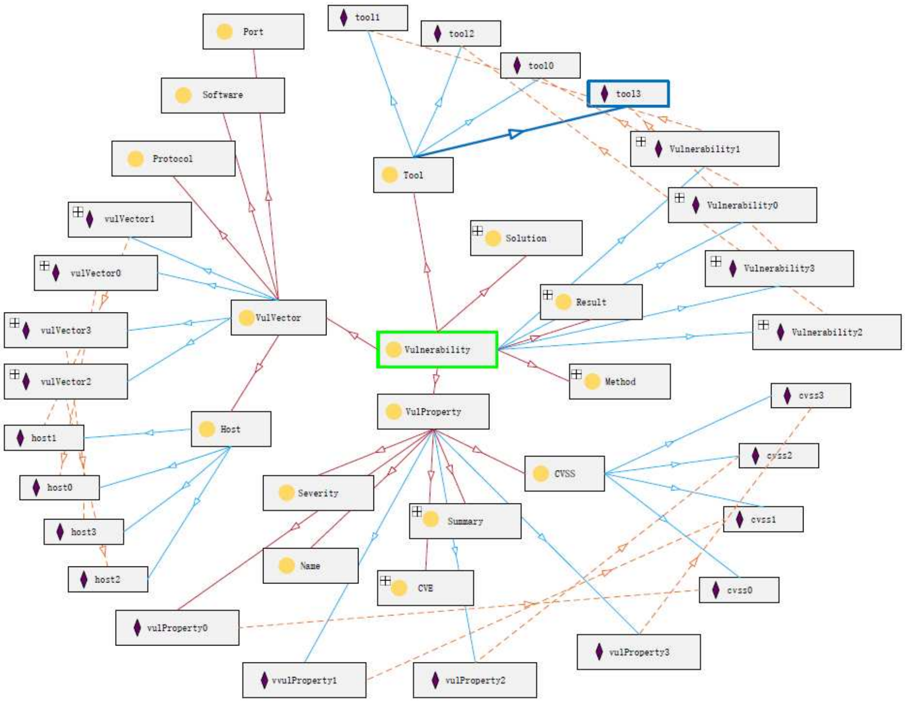

Figure 4: Vulnerability ontology

Secondly, considering the different characteristics of each scan tool, we ser the relative weight of Nessus to 0.7, and that of OpenVAS to 0.3. When calculating the entries only given in Nessus, the fusion results are shown in Table 4.

From Table 4 we can get the reliability function of the basic propositions for different severity recognition frameworks.

When the severity is Critical:

$Bel(exit) = 0.863; Bel(unexit) = 0.136;$

When the severity is High:

$Bel(exit) = 0.737; Bel(unexit) = 0.263;$

When the severity is Medium:

$Bel(exit) = 0.512; Bel(unexit) = 0.487;$

When the severity is Low:

$Bel(exit) = 0.318; Bel(unexit) = 0.681;$

When the severity is Info:

$Bel(exit) = 0.072; Bel(unexit) = 0.927;$

Table 4: Weighted D-S evidence theory fusion results

Severity	Detection tool	Evidence weight	$m(exit)$	$m(unexit)$	$m(\Theta)$
	Nessus	1	0.9	0.1	0
Critical	OpenVAS	3/7	0	0.3	0.7
	Fusion result	–	0.863	0.136	0.001
	Nessus	1	0.8	0.2	0
High	OpenVAS	3/7	0	0.3	0.7
	Fusion result	–	0.737	0.263	0
	Nessus	1	0.6	0.4	0
Medium	OpenVAS	3/7	0	0.3	0.7
	Fusion result	–	0.512	0.487	0.001
	Nessus	1	0.4	0.6	0
Low	OpenVAS	3/7	0	0.3	0.7
	Fusion result	–	0.318	0.681	0.001
	Nessus	1	0.1	0.9	0
Info	OpenVAS	3/7	0	0.3	0.7
	Fusion result	–	0.072	0.927	0.001

By setting $\varepsilon_1 = 0.5, \varepsilon_2 = 0.1$, according to the judgment rule, when the vulnerability entries are only given in Nessus, the severity is Critical, High and Medium, respectively. Only when the severity is Critical and High, the vulnerability exists. If we use D-S evidence theory (That is, when the evidence provided by each tool is equal), the fusion results are shown in Table 5.

From Table 5, we can get the fusion results as follows, and when the severity is Critical:

$Bel(exit) = 0.729; Bel(unexit) = 0.270;$

When the severity is High:

$Bel(exit) = 0.545; Bel(unexit) = 0.454;$

When the severity is Medium:

$Bel(exit) = 0.310; Bel(unexit) = 0.689;$

When the severity is Low:

$Bel(exit) = 0.166; Bel(unexit) = 0.833;$

When the severity is Info:

$Bel(exit) = 0.072; Bel(unexit) = 0.927;$

Similarly, by setting $\varepsilon_1 = 0.5, \varepsilon_2 = 0.1$, only when the severity is Critical and High, the vulnerability exists. Comparing with the weighted D-S evidence theory, the unweighted D-S evidence theory only identifies some vulnerability. This is because after adopting the unweighted fusion method, each of the probe tool evidence weight is equal, that is, low weight and high weight probe tool play the same role. However, the weighted D-S evidence theory-based fusion method

Table 5: D-S evidence theory fusion results

Severity	Detection tool	$m(exit)$	$m(unexit)$	$m(\Theta)$
	Nessus	0.9	0.1	0
Critical	OpenVAS	0	0.7	0.3
	Fusion result	0.729	0.270	0.001
	Nessus	0.8	0.2	0
High	OpenVAS	0	0.7	0.3
	Fusion result	0.545	0.454	0.001
	Nessus	0.6	0.4	0
Medium	OpenVAS	0	0.7	0.3
	Fusion result	0.310	0.689	0.001
	Nessus	0.4	0.6	0
Low	OpenVAS	0	0.7	0.3
	Fusion result	0.166	0.833	0.001
	Nessus	0.1	0.9	0
Info	OpenVAS	0	0.7	0.3
	Fusion result	0.072	0.927	0.001

takes probe tools evidence effectiveness into account, that is, high weight plays more important role whereas low weight probe tools. Therefore, the weighted D-S evidence theory-based fusion method is more accurate.

In order to verify the effectiveness of the proposed method, we test five hosts of experiment environment. Among them three machines with IP 192.168.15.1, 192.168.15.2 and 192.168.15.4, respectively, are installed with open source vulnerability testing system, therefore, there exist lots of vulnerability. The results of using weighted D-S evidence theory and single probe tool are shown in Figure 5. The results of using weighted D-S evidence theory and D-S evidence theory are shown in Figure 6.

Figure 5 shows that weighted D-S evidence theory fusion method can synthesize the results of different detection tools, which is more comprehensive than those of the OpenVAS and Nessus methods.

As we can see in Figure 6, the number of the vulnerability obtained by the weighted D-S evidence theory fusion method is larger than that obtained by the traditional D-S evidence theory fusion method. This is because when vulnerability entries only exist in the Nessus, the proposed weighted D-S evidence theory fusion method can recognize vulnerability when the severities are Critical, High and Medium, whereas, the D-S evidence theory fusion method can only identify vulnerability when the severities are Critical and High.

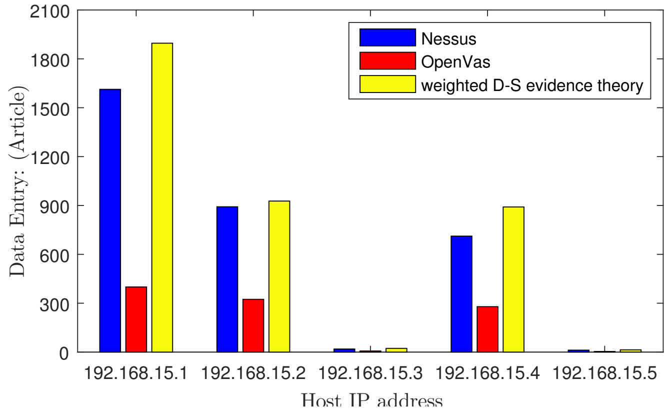

Figure 5: Fusion results of single probe tool and weighted D-S evidence theory

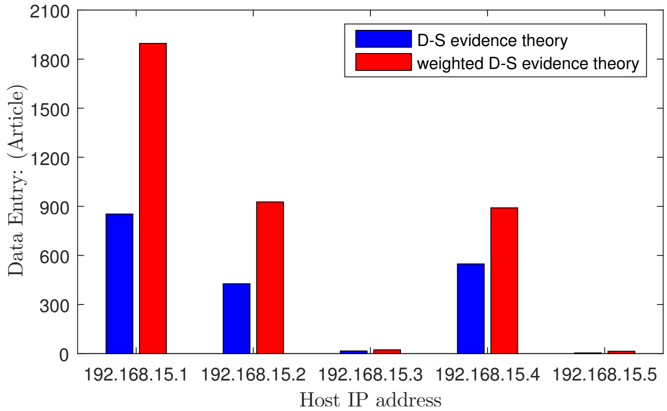

Figure 6: Fusion results of D-S evidence theory and weighted D-S evidence theory

7 Conclusion

In this paper, we propose a novel network vulnerability data fusion method, which is based on ontology and weighted D-S evidence theory. Our method uses ontology semantic description to achieve consistency of multi-source network vulnerability data. Besides, we adapt the D-S evidence theory to increase the accuracy of data fusion result. Then we verify our method through experiments, and the experiment results show that our method can improve the comprehensiveness and accuracy of the data fusion results.

Acknowledgement

This work was supported by the National Natural Science Foundation of China (No. 61363006) and the Open Projects of State Key Laboratory of Integrated Service Networks (ISN) of Xidian University (No.ISN19-13) and the National Natural Science Foundation of Guangxi (No. 2016GXNSFAA380098) and the Science and Technology Program of Guangxi (No. AB17195045).

References

[Alhazmi and Malaiya, 2006] YAlhazmi, O., Malaiya, Y.: " Prediction capabilities of vulnerability discovery models"; Reliability and Maintainability Symposium, RAMS '06, Newport Beach, USA, (February 2006).

[Almgren et al., 2008] Almgren, M., Lindqvist, U.,Jonsson, E.: "A multi-sensor model to improve automated attack detection"; Proceedings of the 11th international symposium on Recent Advances in Intrusion Detection RAID '08, Cambridge, USA (September 2008).

[Azevedo et al., 2015] Azevedo, C., Iacob, M., Almeida, J., Sinderen, M., Pires, L., Guizzrdi,G.: "Modeling resources and capabilities in enterprise architecture: A well-founded Ontology-based Proposal for ArchiMate"; Information systems, 54, c (December 2015) 235-262.

[Bhandari and Gujral, 2014] Bhandari, P., Gujral, M.: "Ontology based approach for perception of network security state"; Engineering and Computational Sciences (RAECS), 2014 Recent Advances in. IEEE '14, Chandigarh, India (March 2014).

[Bishop and Bailey, 1999] Bishop, M., Bailey, D.: "A Critical Analysis of Vulnerability Taxonomies"; A Critical Analysis of Vulnerability Taxonomies, (1999).

[Chen and Feng, 2014] Chen, B., Feng, J.: "Multisensor information fusion of pulsed GTAW based on improved DS evidence theory"; International Journal of Advanced Manufacturing Technology, 71, 1-4 (March 2014) 91-99.

[Delalieux et al., 2014] Delalieux, S., Zarco-Tejada, P.,Tits, L., Bello, M., Intrigliolo, D., Somers, B.: "Unmixing-based fusion of hyperspatial and hyperspectral airborne imagery for early detection of vegetation stress"; IEEE Journal of Selected Topics in Applied earth Observations and Remote Sensing, 7, 6 (July 2014) 2571-2582.

[Dell'Acqua and Gamba, 2012] Dell'Acqua, F., Gamba, P.: "Remote sensing and earthquake damage assessment: Experiences, limits, and perspectives"; Proceedings of the IEEE, 100, 10 (July 2012) 2876-2890.

[Dempster, 1967] Dempster, A.: "Upper and lower probabilities induced by a multivalued mapping"; The Annals of Mathematical Statistics, 38, 2 (Apr 1967) 325-339.

[Ebrahimipour and Yacout, 2015] Ebrahimipour, V., Yacout, Y.: "Ontology-based schema to support maintenance knowledge representation with a case study of a pneumatic valve"; IEEE Transactions on Systems, Man, and Cybernetics: Systems, 45, 4 (January 2015) 702-712.

[Fingas and Brown, 2014] Fingas, M., Brown, C.: "Review of oil spill remote sensing"; Marine Pollution Bulletin, 83, 1 (July 2014) 9-23.

[Gamba, 2013] Gamba, P.: "Human settlements: A global challenge for EO data processing and interpretation"; Proceedings of the IEEE, 101, 3 (April 2013) 570-581.

[Haug et al., 2013] Haug, P., Ferraro, J., Holmen, J., Wu,X., Mynam, K.,Ebert, M., Dean, N., Jones, J.: "An ontology-driven, diagnostic modeling system"; Journal of the American Medical Informatics Association, 20, e1 (March 2013) e102-e110.

[Huang, 2015] Huang, Z.: "A Method of Evaluating Network Attack Probability based on TWDS"; International Smposium on Computers and Informatics, (2015).

[Julisch and Dacier, 2002] Julisch, K., Dacier, M.: "Mining intrusion detection alarms for actionable knowledge"; Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '02, Edmonton, Canada (July 2002).

[Li et al., 2019] Li, Y., Yu, Y., Susilo, W., Min, G., Ni, J., Choo, R.,. "Fuzzy Identity-Based Data Integrity Auditing for Reliable Cloud Storage Systems". IEEE Trans. on Dependable and Secure Computing, 16, 1 (2019), 72-83.

[Liu et al., 2013] Liu, X., Wang, H., Lyu, H., An, S.: "Quantitative awareness of network security situation based on fusion"; Journal of Jilin University Engineering and Technology Edition, 43, 6 (November 2013) 1650-1657.

[Mura et al., 2015] Mura, M., Prasad, S.,Pacifici,F., Gamba, P., Benediktsson, J.: "Challenges and opportunities of multimodality and data fusion in remote sensing"; Proceedings of the IEEE, 103, 9 (August 2015) 1585-1601.

[Ning et al., 2002] Ning, P., Cui, Y., S.Reeves, D.: "Constructing attack scenarios through correlation of intrusion alerts"; Proceedings of the 9th ACM Conference on Computer and Communications Cecurity. CCS '02, Washington, USA (November 2002).

[Oellrich et al., 2015] Oellrich, A., Walls, R., Cannon, E., Cannon, S., Cooper, L., Gardiner, J., et al.: "An ontology approach to comparative phenomics in plants"; Plant methods, 11, 1 (January 2015) 10.

[Sadighian et al., 2013] Sadighian, A., Fernandez, J.,Lemay, A., Zargar, S.: "ONTIDS: A Highly Flexible Context-Aware and Ontology-Based Alert Correlation Framework"; Intl symposium on Foundations and Practice of Security '13, Springer New York (October 2013).

[Shafer(76)] Shafer, G.: "A mathematical theory of evidence", Princeton University Press.

[Si and Zhang et al., 2014] Si, C., Zhang, H.,Wang Y., Liu, J.: "Network security situation elements fusion method based on ontology"; Proceeding ISCID '14 Proceedings of the 2014 Seventh International Symposium on Computational Intelligence and Design - Volume 02, ISCID '14, December, USA (December 2014).

[Wang et al., 2012] Wang, J., Zhang, Q., Zhi, H.: "Fault diagnosis and optimization for agent based on the ds evidence theory"; International Conference in Swarm Intelligence '12, Springer, Berlin, (2012).

[Yu et al., 2018] Yu, Y., Li, Y., Du, X., Chen, R., Yang, G., "Content Protection in Named Data Networking: Challenges and Potential Solutions". IEEE Communications Magazine 56, 11 (2018), 82-87.

[Yu et al., 2005] Yu, D., Frincke, D.: " Alert confidence fusion in intrusion detection systems with extended Dempster-Shafer theory"; Proceedings of the 43rd annual Southeast Regional Conference-Volume 2. ACM-SE '05, New York, USA, (March 2005).

[Yu et al., 2017] Yu, Y., Au, M. H., Ateniese, G., Huang, X., Susilo, W., Dai, Y., Min, G., "Identity-Based Remote Data Integrity Checking with Perfect Data Privacy Preserving for Cloud Storage". IEEE Trans. Information Forensics and Security 12, 4 (2017), 767-778.

[Zhong and Zhao, 2012] Zhong, Q., Zhao, Z.: "Vulnerability Data Fusion Method Based on the DS theory of Evidence"; Instrumentation, Measurement, Computer, Communication and Control (IMCCC), 2012 second iInternational Conference on. IEEE '12, Harbin, China (2012).

[Zhang et al., 2011] Zhang, Y., Huang, S.,Guo,S., Zhu, J.: "Multi-sensor data fusion for cyber security situation awareness"; Procedia environmental sciences, 10, 1 (December 2011) 1029-1034.