Network Situation Assessment of Host Node Based on Improved D-S Evidence Theory

Liangyuan HE¹, Tao WAN², Chuanlin ZHANG¹, Fei XIA¹, Sheng WANG¹, Yajie WANG¹
¹College of Automation Engineering, Shanghai University of Electric Power, Shanghai 200090, China
²Hangzhou DPtech Technologies Co., Ltd., Hangzhou, Zhejiang 310051, China
clzhang@shiep.edu.cn

Abstract. In recent years, network security has become a hot spot of social concern. In order to avoid dependence on network training, improved D-S evidence theory is applied to the situation assessment of host. The fuzzy function is adopted to generate the BPA of the corresponding feature, and the weighting matrix is added to handle conflict evidences in the fusion process. Finally, the situation level of the host node is obtained. The effectiveness and feasibility of the method is proved by analyses of examples.

1. Introduction
Network security situational awareness involves data mining, intrusion detection, data fusion and other related technologies. Many outstanding research results have appeared in recent years. In 2016, a hierarchical security threat situation quantitative evaluation model for bottom-up overall assessment strategies was proposed, which combines cybersecurity with hierarchical theory [1]. Liu et al. assumed that the weakness in the network was unchanging. Combining the analytic hierarchy process (AHP) and the attack graph, a security evaluation method for the power communication network was proposed. However, the weakness of the actual network keeps changing [2]. In addition, considering single feature during assessment process is not suitable. When it comes to multiple features contained in the model, BP neural network algorithm was adopted by literature [3] to establish the relationship between network situation level and perceptual parameters, and quantitatively assessed the situational awareness. However, BP neural network has strong dependence on training samples with defined labels. A network security situation assessment model based on information fusion for network security model was proposed in literature [4-5], fusing source information generated by multiple detection devices. Owing to the ability of expressing uncertainty and less calculation, Dempster-Shafer evidence theory (DST) is consistent with the uncertainty of network security assessment issues.

In summary, a security assessment of host node based on the improved DST is proposed in this paper. Firstly, the construction of the network situational awareness model is based on the improved Basic Probability Allocation (BPA) method and the improved evidence combination rules. The network security situation of the host node at that moment is finally obtained.
2. Preliminaries

2.1. D-S evidence theory
DST was first proposed by Dempster in 1967 and improved by Shafer in 1976 [6]. First, Θ is defined as the frame of discernment, which is a finite non-empty set of N elements that are mutually exclusive and exhaustive. The power set consisting of \(2^N\) elements of Θ is denoted as \(P(Θ)\). The BPA function is defined as a numerical mapping between 0 and 1 of the power set \(P(Θ)\). The following conditions must be met:

\[
m(∅) = 0, \sum_{A∈Θ} m(A) = 1
\]

For example, if \(A\) is a subset of the frame of discernment, and satisfies \(m(A)>0\), then \(A\) is called one of the focal elements of \(Θ\). The value \(m(A)\) of \(A\) indicates the degree of support of the focus element \(A\).

Suppose that the frame of discernment consists of two mutually independent evidence bodies \(m_1\) and \(m_2\), and the corresponding Dempster combination rules are:

\[
m_{1⊕2}(B) = \frac{1}{K} \sum_{B∩C=A} m_1(B)m_2(C)
\]

\[
K = \sum_{B∩C=∅} m_1(B)m_2(C)
\]

2.2. Triangular fuzzy number
Due to the small sample set for BPA construction and easy calculation, the extended fuzzy number [7] is adopted defined as \(A=(a, b, c, w)\), where \(a\), \(b\), and \(c\) are real number. As shown in the Fig.1, if \(w=1\), \(A\) is called a regular triangular fuzzy number. Besides, there are trapezoidal and Gaussian fuzzy numbers. According to the data distribution characteristics of the research object [8], this paper chooses the triangular fuzzy number to calculate the BPA.

![Regular triangular fuzzy number](Fig.1)

2.3. The situation assessment model of Host node
So far, research scholars have proposed several models of network security situational awareness. Representative of them is the hierarchical model of literature [9]. On this basis, combined with the perceptual elements and evaluation target of network security, a network security situational awareness model based on host nodes is constructed in this paper, as shown in Fig.2.

![The network security situational awareness model of host nodes](Fig.2)
number of intrusion attacks and the number of vulnerabilities reflect the potential threat of the host. The result of assessment can be obtained by fusing the evidences generated in the feature layer.

3. Method of this paper

3.1. The construction of BPA

In this paper, the frame of discernment for the assessment of the host node situation is \( \Theta = \{H, M, L, \emptyset\} \). \( m(\emptyset) \) indicates that the current host node posture does not belong to the mass function of the level H, level M and level L. Usually, it does not exist, denotes as \( m(\emptyset) = 0 \). For example, \( m(H) \) represents the good statue of the current host node situation; \( m(M) \) represents the moderate statue of current host node situation; \( m(L) \) represents the poor statue of current host node situation. The situational elements that affect the host node situation are recorded as \( x_1, x_2, x_3, x_4, x_5, \) and \( x_6 \), respectively. In addition, the value of the mass function of each of the focal elements in the frame of discernment satisfies the basic condition in section 2.

The relationship between node elements and node states is shown in Table 1 by querying the snort security manual [10].

| Symptom parameters | m(H)       | m(M)       | m(L)       |
|--------------------|------------|------------|------------|
| X1 CPU utilization | \( <=60\% \) | \( 60\%-90\% \) | \( >=90\% \) |
| X2 Memory usage    | \( <=70\% \) | \( 70\%-90\% \) | \( >=90\% \) |
| X3 Hard disk usage | \( <=50\% \) | \( 50\%-70\% \) | \( >=70\% \) |
| X4 IDS/the number  | 0-200      | 200-300    | >=300      |
| X5 Flow in unit time | 0-3000   | 3000-3500  | >=3500     |
| X6 Number of vulnerabilities | 0-4       | 4-10      | >=10       |

As can be seen from Table 1, if obtained host running asset accounts for a larger proportion, it means that the current running state of the host is worse, and the possibility that the current situation is at the L level is higher. Similarly, as the number of intrusions, the amount of traffic and the number of vulnerabilities increase, the statue of the host itself is more likely to be the L level. Based on this, BPA can be established by the following steps [8]:

Step 1 Analyse the distribution of features and obtain the feature distribution stability points of different levels (level H, level M, level L) under the corresponding elements. Fuzzy number membership is defined as \( A_i = (a_i, b_i, c_i, \omega) \), where \( i = 1, 2, 3; \omega=1; \)

Step 2 Draw the corresponding triangular fuzzy number membership image according to the obtained \( A_i \);

Step 3 According to the definition given in section 2.2, the corresponding membership function is substituted into the \( A_i \) parameter to obtain the corresponding \( \mu_i(x) \) by formula (4).

Step 4 If the sample intersects the fuzzy function, the intersecting vertical coordinates are marked as the BPA of the proposition.

\[
\mu(x) = \begin{cases} 
0, & x < a \\
\frac{x-a}{b-a}, & a < x < b \\
1, & x = b \\
\frac{c-x}{c-b}, & b < x < c \\
0, & x = c
\end{cases}
\]

If a sample intersects with multiple fuzzy functions, the fuzzy values of multiple intersecting points are the BPA of multiple propositions. When the sum of the generated BPAs \( \sum_{i=1}^{n} m_i < 1 \) (\( n \) is the number of recognition frame elements), other BPAs are generated as:
where $d_k$ is the distance between the center point of the other fuzzy triangle number and $\max (m_i)$.

When the sum of the generated BPAs $\sum_{i=1}^{n-2} m_i < 1$, $m_n = 1 - \sum_{i=1}^{n-2} m_i$ is generated.

When the sum of the generated BPAs $\sum_{i=1}^{n-2} m_i > 1$, normalization is supposed to be performed.

### 3.2. Evidence correction

When the classical DST deals with the fusion problem, once the evidence source is highly conflicted, Zadeh paradox [11] will appear. In order to solve the problem that the classical fusion rules are not in line with the fact, it is necessary to make appropriate corrections to the evidence sources with high conflicts. So far, most scholars have adopted the method of seeking weighted evidence or average evidence [12].

Suppose that there are two evidences $m_1, m_2$, and correspondingly the focal elements of $\Theta$ are $A_i$ and $B_j$. Then the similarity coefficient between the two evidences can be calculated as:

$$d_{12} = \frac{\sum_{i,j} m_1(A_i)m_2(B_j)}{\sqrt{(\sum m_1^2(A_i))(\sum m_2^2(B_j))}}$$

If the value of $d_{12}$ is equal to 1, the evidences are identical. If the result is 0, the conclusion is reversed. The corresponding similarity matrix can be easily obtained by the similarity coefficient, which is denoted as Sim. Then the weight of evidence can be calculated by formula (8-9).

$$\text{Sim} = \begin{bmatrix}
1 & d_{12} & \cdots & d_{1n} \\
\vdots & \vdots & \ddots & \vdots \\
d_{n1} & d_{n2} & \cdots & 1
\end{bmatrix}
$$

$$\text{Sup}(m_k) = \sum_{k=1}^{n} d_{kt}, k = 1, 2, \ldots, n;$$

$$w(m_k) = \frac{\text{sup}(m_k)}{\sum_{k=1}^{n} \text{sup}(m_k)}, k = 1, 2, \ldots, n$$

Finally, the weighted evidence can be obtained according to formula (9).

$$m_{avg}(A_i) = \sum_{i=1}^{n} w(m_k)m_k(A_i), i = 1, 2, \ldots, n$$

### 3.3. Combination rule

For the combination of conflicting evidence sources, Literature [13] replaced the conflicting evidence with the generated weighted evidence, and finally adopted the classical Dempster combination rule. However, it is difficult to determine which evidence is highly conflicting in practical applications. Once the replacement is wrong, it is likely that the opposite situation will be reached. Therefore, it is necessary to retain conflict evidence and assign conflict information to the identification framework [14]. The specific steps can be referred to paper [15].

### 4. Simulations

In order to verify the reliability of the selected method, the following two examples are simulated. The data format used in the simulation is a set of data of six characteristic parameters per unit time, and a total of ten sets of data. In order to obtain the fuzzy number of each feature element, the distribution law of each feature data is analysed. First, this section demonstrates the composition of membership with an example, and then simulates ten sets of sample data separately. Finally, through comparison with BP neural network method, it is verified whether the situation evaluation of the method at the host node is effective.
Example 1
When the target samples are $A_1 = (0, 29, 60, 100); A_2 = (54, 75.4, 90, 1); A_3 = (84, 92.7, 100, 1)$, the obtained membership degree is shown in Fig. 3.

Table 2. The BPA distribution of six symptoms from different membership functions.

|     | X1   | X2   | X3   | X4   | X5   | X6   |
|-----|------|------|------|------|------|------|
| m(H)| 0.5484 | 0.1995 | 0.3756 | 0.3614 | 0.5201 | 0.4239 |
| m(M)| 0.2613 | 0.3731 | 0.4041 | 0.3506 | 0.2666 | 0.3030 |
| m(L)| 0.1903 | 0.4274 | 0.2204 | 0.2879 | 0.2133 | 0.2731 |

If CPU utilization is 15% $m(H) = 0.5180$. According to Step 4, $m(M) = 0.3971; m(L) = 0.0920$ can be calculated. If CPU utilization is 75%, $m(H) = 0.2761; m(M) = 0.7176; m(L) = 0.0062$.

Similarly, according to the above method, the corresponding fuzzy number is established according to the sample data distribution characteristics of each feature element.

Example 2
For instance, use the fuzzy number of Example 1 to calculate the corresponding feature BPA allocation and save it in Table 2. The combination in Table 3 was performed using the improved D-S evidence theory of Section 2 and Section 3. Moreover, the assessment results obtained via BP neural network are also recorded in Table 3.

Table 3. The combination results of Table 2.

| Method               | m(H)   | m(M)   | m(L)   |
|----------------------|--------|--------|--------|
| BP neural network    | 0.4981 | 0.4540 | 0.0479 |
| This paper           | 0.5429 | 0.2915 | 0.1659 |

As can be clearly seen in Table 2, except for the BPA of feature 2 and feature 3, other features have the highest support probability of level H. According to common sense, the situation of the host node at this moment is level H. In this paper, when $\max(m(A_i)) - \min(m(A_j)) > \varepsilon, \varepsilon = 0.2, i = 1, 2, ..., n$, the final decision result is $A_i$. Therefore, according to the above rules, combined with the calculation results in Table 3, the final state of the Example 2 host node is level H. In contrast, the support probability for level H in this method is significantly higher than the results of BP neural network. Therefore, method proposed in this paper to evaluate the situation level of the host node is valid.

In addition, by analysing the results of ten sets of data fusion, the final decision of this method is consistent with the expected decision. Three groups assessed by BP neural network were judged by mistake. Besides, when a certain feature data is at the extreme of the level interval, the final result of the BP neural network has a local optimal phenomenon of the final result.

5. Conclusion
In this paper, the improved D-S evidence theory is applied to the situation level assessment of the host node. The examples show that the method is effective in the network situation assessment of the host node.
node. The method takes all the situational features of host node into account. It also has certain ability to handle conflict information. In the fusion process, reasonable conflict information can be distributed to conflict evidence, thus avoiding the one-vote veto phenomenon of the fusion result. However, there are still some problems with this method. In the process of calculating the BPA, once the collected data happens to be at the end of the interval, it is likely to obtain a support probability for focal element is 0. Obviously, it is unreasonable. In addition, owing to the fact that the situation assessment of a certain network is not limited to the host node, it is our future work to promote the method to other device nodes and make it clear that how to fuse different layers.

Acknowledgments
Thanks are due to the project colleagues for assistance with the simulations and valuable discussion.

References
[1] Kokkonen, Tero. Architecture for the cyber security situational awareness system[J]. Internet of things, smart spaces, and next generation networks and systems. Springer, Cham, 2016: 294-302.
[2] Liu, Nian, et al. Security assessment for communication networks of power control systems using attack graph and MCDM[J]. IEEE Transactions on Power Delivery, 2010, 25(3): 1492-1500.
[3] Xu, Zhou, et al. Study on security risk assessment of power system based on BP neural network[J]. Journal of Computational and Theoretical Nanoscience, 2016, 13(8): 5277-5280.
[4] Zhang, Wei, et al. A novel trust management scheme based on Dempster–Shafer evidence theory for malicious nodes detection in wireless sensor networks[J]. The Journal of Supercomputing, 2018, 74(4): 1779-1801.
[5] Shafer, G., A Mathematical Theory of Evidence. Princeton University Press, 1976, Princeton.
[6] Deng Yong, HAN Deqiang. The Method of Generating Probability Assignment in General Evidence Theory[J]. Journal of Xi’an Jiaotong University, 2011, 45(02): 34-38 (in Chinese).
[7] Wang Lei, Jiang Weijian, et al. Application of improved D-S evidence theory in human body fall detection in substation[J]. Journal of Electronic Measurement and Instrument, 2017, 31(07): 1090-1098 (in Chinese).
[8] Yang Jianbo, et al. A general multi-level evaluation process for hybrid MADM with uncertainty[J]. IEEE Transactions on Systems, Man, and Cybernetics, 1994, 24(10): 1458-1473.
[9] Miao Ke. Research on network security situational awareness technology based on index extraction [D]. Beijing University of Posts and Telecommunications, 2015:3-33 (in Chinese).
[10] Zadeh, L.A. Review of Shafer’s mathematical theory of evidence[J]. AI Magazine, 1984, 5: 81-83.
[11] Lin Yun, Li Yuxiao, et al. Multisensor Fault Diagnosis Modeling Based on the Evidence Theory[J]. IEEE Transactions on Reliability, 2018, 67(2): 513-521.
[12] Hao Zhiwei, Wu Yong, et al. Analysis of BP network and Improved Evidence Synthesis Rules for Air Target Recognition[J]. Electronics Optics & Control, 2014, 21(12): 36-39+66 (in Chinese).
[13] RONALD R Y. On the dempster-shafer framework and new combination rules[J]. Information sciences, 1987, 41(2): 93-137.
[14] Xia Fei, Ma Xi, et al. Application of Improved D-S Evidence Theory in Fault Diagnosis of Lithium Battery for Electric Vehicles[J]. CAAI Transactions on Intelligent Systems, 2017, 12(04): 526-537 (in Chinese).