Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (10): 1740-1745, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.1740.1745

Risk Assessment of Information Security Based on Grey Incidence and D-s Theory of Evidence

[1]Ling Liu, [1]Tonggang Bao, [2]Jiahang Yuan and [2]Cunbin Li
[1]Jinzhong Electric Power Supply Company, Shanxi Electric Power Corporation, Jinzhong, China
[2]School of Economic and Management, North China Electric Power University, Beijing, China

Abstract: As new threats continue to emerge, the information system cannot be safe forever. To ensure information security, a security risk assessment is needed. Compared to traditional methods, such as AHP, fuzzy logic, and grey analysis, an approach based on grey incidence and D-S theory of evidence is put forward to evaluate information system security in this paper. Firstly, the uncertainty in index parameter values is analyzed, according to the actual condition and history statistical data, the vacant index parameter values may meet three kinds of distributions: uniform distribution, exponential distribution, and normal distribution. The corresponding prior estimates are given to fill the vacant values up. Then, the concept of interval conversion operator is defined, using grey incidence to determine the uncertain degrees of different indices, and the mass functions are obtained by the uncertain degrees. Finally, mass functions are fused in accordance with the rule of combination and sequence the information system security risk according to the belief function value. An example application has proved the feasibility and effectiveness of this method. The results indicate this method can obviously reduce the overall uncertainty and provide a new thought to information security risk assessment approaches.

Key words: Information security, risk assessment, grey incidence, D-S theory of evidence

INTRODUCTION

As an important part of national information construction, information system, its security involves the fundamental interests of the country and the users. As new threats continue to emerge, the information system can not be safe forever. So ensuring information security is a relative and dynamic process. Making a security risk assessment is an effective way to protect information.

Information security risk assessment is an estimate of the possibility of invasion and destruction and the consequences of the confidentiality, integrity and availability of the information during its generation, transmission and storage according to the relevant information security standard. It assesses the negative influence on frequency of threat, severity extent of survivabulity and influence on asset being used by risk of information system. It is the process of identifying risk which is the foundation of information security management construction. The potential risk could be found in the system through the risk assessment, and then we could take measures to keep risk under control.

There are many methods of information security risk assessment. Considering the fuzziness and uncertainty of the risk parameters, Chen and Chen (2003) assess information system security risk by using fuzzy logic method. Zhao *et al.* (2005) combine AHP with fuzzy logic, using fuzzy logic to compensate the deficiency of AHP. Discount ratio is used to improve Demspter's rule of combination of evidence, then the improved rule is used to combine the risk factors of network, by Gao and Zhu (2008). The uncertainty of risk factors is reduced. In Fu, (2010) 's study, the risk factors of information systems are classified into three aspects of influence on asset, frequency of threat and severity extent of survivabulity, which are analyzed based on the fuzzy set theory and entropy theory to integrate the respective risk assessment results of three factors to obtain the final risk grade. Owing to the characteristic of uncertainty information in the assessment process, the reasoning algorithm on Bayesian Networks is presented by Fu *et al.* (2006) and the conditional probability matrix of the reasoning rule is given by him based on the expert knowledge. After the assessment process is analyzed, the model of information security risk assessment is constructed. Zhao and Xue (2009) propose Multi Attribute Group Decision-making (MAGDM) security assessment method based on Variable Precision Rough Set (VPRS). This assessment method, combines VPRS with the Analytic Hierarchy Process (AHP), uncovers the inherent information hidden

Corresponding Author: Ling Liu, Jinzhong Electric Power Supply Company, Shanxi Electric Power Corporation, Jinzhong, China

in data via the quality of classification, and makes a synthetic security assessment of the information system. Demotier *et al*. (2006) show that theory of evidence is a valuable framework of risk analysis under uncertain environment and shortages of information. By Gao and Luo (2009), a grey relational decision making algorithm is put forward to evaluate information system security for solving uncertainty.

Although a lot of efforts had been made on information security risk assessment, these methods have different degrees of deficiency. First, using different methods to assess the same object may obtain different conclusions. Second, some methods have a high subjective content. Last, some algorithm as theoretical results has its limitation, in other words the theory is divorced from practice. The contributions of this paper are: (1) By combining grey incidence with D-S theory of evidence, a new method is presented for information security risk assessment taking into account the information uncertainty in index parameter values. (2) The paper studies information fusion based on D-S theory of evidence.

UNCERTAINTY ANALYSIS

Uncertainty in index parameter values can be divided into two kinds, one is grey index parameter values, an uncertainty value within the certain interval or a number set. The other is vacant index parameter values meant that there is absolutely no information about it. The complexity and dynamic of information system is a major cause of missing data, we use $(*)_{ij}$ to indicate the missing data.

For the grey index parameter values, we can ask experts to estimate. Supposing s_{ij} is the assessed value under the index j of the information system i:

$$s_{ij} \in [s_{ij}^{l}, s_{ij}^{u}]$$

For the vacant index parameter values, we could transform the vacant index parameter values into grey ones. The value range, expectation value μ and variance σ^2 could be known according to the actual condition and history statistical data. We could discuss the vacant index parameter values under three circumstances:

- Let $p_{ij}(t)$ be a probability distribution density for $s_{ij}(\phi)$ taken random from $[s_j^{l*}, s_j^{u*}]$ provided that $s_{ij}(\phi) \in [s_j^{l*}, s_j^{u*}]$. According to the distribution estimation theory of maximum entropy, provided that:

$$p_{ij}(t) = (s_j^{l*} - s_j^{u*})^{-1}$$

the information entropy reaches a maximum. Thus $s_{ij}(\phi)$ may meet uniform distribution, it is reasonable that we use interval grey number $s_{ij}(\phi) \in [s_j^{l*}, s_j^{u*}]$ to fill the vacancy. Meanwhile:

$$\forall 0 < \varepsilon \le \frac{1}{2}(s_j^{l*} - s_j^{u*})$$

the degree of satisfaction is:

$$F_{ij}(\varepsilon) = 2\varepsilon(s_j^{l*} - s_j^{u*})^{-1}$$

provided that:

$$s_{ij}(\phi) \in [\frac{s_j^{l*} + s_j^{u*}}{2} - \varepsilon, \frac{s_j^{l*} + s_j^{u*}}{2} + \varepsilon]$$

- If $s_{ij}(\phi) \in [\alpha, +\infty)$ and the expectation value is μ, According to the distribution estimation theory of maximum entropy, $s_{ij}(\phi)$ may meet exponential distribution. $\forall \beta \in (\alpha, +\infty)$, the degree of satisfaction is:

$$F_{ij}(\beta) = \int_{\alpha}^{\beta} (\mu - \alpha)^{-1} e^{-\frac{t-\alpha}{\mu-\alpha}} dt$$

provided that $s_{ij}(\phi) \in [\alpha, \beta]$

- If $s_{ij}(\phi) \in (-\infty, +\infty)$, the expectation value is μ and variance is σ^2. According to the distribution estimation theory of maximum entropy, $s_{ij}(\phi)$ may meet normal distribution. $\forall 0 < \varepsilon \le \mu$, the degree of satisfaction is:

$$F_{ij}(\varepsilon) = \int_{\mu-\varepsilon}^{\mu+\varepsilon} (\sqrt{2\pi}\sigma)^{-1} e^{-\frac{(1-\mu)^2}{2\sigma^2}} dt$$

provided that $s_{ij}(\phi) \in [\mu - \varepsilon, \mu + \varepsilon]$. τ is the confidence level, written as:

$$\tau = \frac{mn - q}{mn}$$

where, q is the number of vacant data.

Based on the above analysis, we can fill up the vacant values and obtain the initial matrix $s = (s_{ij})_{m\times n}$.

Definition 1: let G_α be a conversion operator that transform interval numbers into real numbers, written as:

$$G_\alpha(s_{ij}) = \frac{(s_{ij}^{l} + s_{ij}^{u})}{2} + \alpha\frac{s_{ij}^{u} - s_{ij}^{l}}{2}$$

where α is risk factor, $S(\alpha) \in [-1, 1]$.

MASS FUNCTION DETERMINED BY GREY INCIDENCE

When we make the assessment, we should fuse the information in accordance with D-S theory of evidence.

Fusion is based on the mass function, and the uncertain degree of indices is the crux of the mass function.

The uncertain degree of the index j is:

$$DOI(I_j) = \frac{1}{m} \left| \sum_{i=1}^{m} (r_{ij})^q \right|^{\frac{1}{q}} \quad (1)$$

In the formula, q = 2, because we use Euclidean distance to improve resolution. r_{ij} is the integrated grey correlation coefficient. The expression is:

$$r_{ij} = \frac{1}{(1 + r_{ij}^{+} / r_{ij}^{-})^2} \quad (2)$$

In order to avoid distortion, we calculate two correlation coefficients. The optimal correlation coefficient r_{ij}^{+} and the worst correlation coefficient r_{ij}^{-} are calculated by the two formulas:

$$r_{ij}^{+} = \frac{\min_i \min_j |x_{ij} - X^{+}| + \xi \max_i \max_j |x_{ij} - X^{+}|}{|x_{ij} - X^{+}| + \xi \max_i \max_j |x_{ij} - X^{+}|} \quad (3)$$

$$r_{ij}^{-} = \frac{\min_i \min_j |x_{ij} - X^{-}| + \xi \max_i \max_j |x_{ij} - X^{-}|}{|x_{ij} - X^{-}| + \xi \max_i \max_j |x_{ij} - X^{-}|} \quad (4)$$

In the formulas:

$$X^{+} = \max_{\substack{1 \le i \le m \\ 1 \le j \le n}} x_{ij} = \{x_1^{+}, x_2^{+}, ..., x_n^{+}\}$$

$$X^{-} = \min_{\substack{1 \le i \le m \\ 1 \le j \le n}} x_{ij} = \{x_1^{-}, x_2^{-}, ..., x_n^{-}\}$$

X^{+} and X^{-} are said to be the ideal optimal sequence and the ideal worst sequence of the weighted normalized matrix. $\xi \in [0,1]$, distinguishing coefficient, usually, $\xi = 0.5$. $X = (x_{ij})_{m \times n} = (w_j \cdot y_{ij})_{m \times n}$ is the weighted normalized matrix where ω is the index weight, $Y = (y_{ij})_{m \times n}$ is the normalized matrix.

Accordingly, the mass function can be obtained by the Eq. 5:

$$m_j(i) = [1 - DOI(I_j)] \times y_{ij} \quad (5)$$

$m_j(i)$ stands for the mass function under the index j of the information system A_i. Because of the limitation and uncertainty of cognitive ability, we know:

$$\sum_{i=1}^{m} m_j(i) < 1$$

Drawing on the experience of Li and Liu (2011), the Identification framework Θ which is endowed with this part of the mass function will partake in fusion calculation so that the uncertainty in the assessment could be reduced. Therefore, we could obtain the overall uncertainty mass function:

$$m_j(i+1) = 1 - \sum_{i=1}^{m} m_j(i) \quad (6)$$

D-S THEORY OF EVIDENCE

The Dempster-Shafer theory is a mathematical theory of evidence. It is also an uncertainty fusion algorithm, it allows one to combine the consistency information from the imprecise judgment and description and arrive at a degree of belief. The belief is represented by a belief function.

Definition 2: Identification framework Θ. Identification framework is the mutually exclusive and nonempty set. It represent all possible hypotheses that determine the claim of a system under consideration.

Definition 3: Basic Probability Assignment (BPA). The function m is $m: 2^{\Theta} \rightarrow [0,1]$, when it has two properties. First, the mass of the empty set is zero: $m(\phi) = 0$. Second, the masses of the remaining members of the set add up to a total of 1:

$$\sum_{X \subseteq \Theta} m(X) = 1$$

The mass m(A) of A expresses the proportion of all relevant and available evidence that supports the claim. The value of m(A) pertains only to the set A and makes no additional claims about any subsets of A, each of which have, by definition, their own mass.

From the mass assignments, the upper and lower bounds of a probability interval can be defined. This interval contains the precise probability of a set of interest (in the classical sense), and is bounded by two non-additive continuous measures called belief function and plausibility function:

$$bel(A) \le P(A) \le pl(A) \quad (7)$$

The belief function bel(A) for a set A is defined as the sum of all the masses of subsets of the set of interest:

$$Bel(A) = \sum_{B \subseteq A} m(B) \quad (8)$$

The plausibility function Pl(A) is the sum of all the masses of the sets B that intersect the set of interest A:

$$pl(A) = \sum_{B \cap A \neq \phi} m(B) \quad (9)$$

The two measures are related to each other as follows: $pl(A) = 1 - bel(\overline{A})$. We know different sources express their beliefs over the framework in terms of belief constraints, then Dempster's rule of combination is the appropriate fusion operator. This rule derives common shared belief between multiple sources and ignores all the conflicting belief through a normalization factor. The combination is calculated as the following formula:

$$m(A) = m_i(B) \oplus m_j(C) = \begin{cases} 0 & B \cap C = \phi \\ \dfrac{\sum_{B \cap C = A, \forall B, C \subseteq \Theta} m_i(B) \times m_j(C)}{1 - \sum_{B \cap C = \phi, \forall B, C \subseteq \Theta} m_i(C) \times m_j(B)} & B \cap C \neq \phi \end{cases} \quad i, j = 1, 2, \ldots, m \quad (10)$$

The combination formula has two important characteristics, commutativity and associativity:

$$m_1 \oplus m_2 = m_2 \oplus m_1$$

$$m_1 \oplus (m_2 \oplus m_3) = (m_1 \oplus m_2) \oplus m_3$$

Also, we could combine many mass functions by the Eq. 11:

$$\oplus \ldots \oplus m_n)(A) = \frac{\sum_{A_1 \cap A_2 \cap \ldots \cap A_n = A} m_1(A_1) m_2(A_2) \ldots m_n(A_n)}{1 - \sum_{A_1 \cap A_2 \cap \ldots \cap A_n = \phi} m_1(A_1) m_2(A_2) \ldots m_n(A_n)} \quad (11)$$

where, m_1, m_2, ..., m_n are the mass functions belong to the identification framework Θ, $A \subseteq \Theta$.

THE STEPS OF ASSESSMENT

- **Step 1:** Investigate the index parameter values and construct the initial matrix $S = (s_{ij})_{m \times n}$, transform the interval values into real numbers
- **Step 2:** Normalize initial the risk matrix $G = (g_{ij})_{m \times n}$ by the equation:

$$\left(\frac{g_{ij}}{\sum_{i=1}^{m} g_{ij}}\right)_{m \times n}$$

and then obtain the weighted normalized matrix $X = (x_{ij})_{m \times n}$

- **Step 3:** Calculate the integrated grey correlation coefficient r_{ij} and then determine the uncertain degree of the index $DOI(I_j)$
- **Step 4:** The mass function $m_j(i)$ and the overall uncertainty mass function $m_j(i+1)$ can be obtained by the Eq. 5 and 6
- **Step 5:** According to the rule of combination, we could combine many mass functions by the Eq. 10
- Step 6: Sequence the results

THE EXAMPLE OF APPLICATION

$A = \{A_1, A_2, \ldots, A_m\}$ represents information system. According to the definition of information security risk assessment, we select three indicators Virus, system flaw and Percentage of unavailable data that reflect frequency of threat, severity extent of survivability and influence on asset. Investigate the data of five information systems in the current month as the following Table 1.

In the information system A_2, the value of system flaw is unknown. But we know that its minimum value is 4 and its expectation value is 6. First, we should fill up the vacant. According to the uncertainty analysis, the vacant index parameter values most likely meet exponential distribution. That means we could use $s_{42}(\phi) \in [4, \beta]$ to fill up the vacant. The degree of satisfaction is 0.7, the satisfaction function is:

$$F_{42}(\beta) = \int_4^{\beta} (6-4)^{-1} e^{-\frac{t-4}{6-4}} dt = 1 - e^{-\frac{\beta-4}{2}}$$

After calculation, $\beta = 6.4$. So we can use $s_{42}(\phi) \in [4, 7]$ to fill up the vacant, the degree of satisfaction is 0.777>0.7. Then, we get complete data as the following Table 2.

We call α with an interval value of zero, then the interval index parameter values are transformed into real

Table 1: The initial data

	Virus	System flaw	Percentage of unavailable data
A_1	[6,10]	[2,5]	[35%, 45%]
A_2	[7,13]	[3,8]	[30%, 60%]
A_3	[4,9]	[5,10]	[20%, 50%]
A_4	[3,5]	$(*)_{ij}$	[10%, 30%]
A_5	[5,7]	[4,9]	[20%, 40%]

Table 2: Complete data

	Virus	System flaw	Percentage of unavailable data
A_1	[6,10]	[2,5]	[35%,45%]
A_2	[7,13]	[3,8]	[30%,60%]
A_3	[4,9]	[5,10]	[20%,50%]
A_4	[3,5]	[4,7]	[10%,30%]
A_5	[5,7]	[4,9]	[20%,40%]

numbers. So we establish the risk matrix $G = (g_{ij})_{m\times n}$.

$$G = \begin{pmatrix} 8 & 3.5 & 0.4 \\ 10.5 & 5.5 & 0.45 \\ 6.5 & 7.5 & 0.35 \\ 4 & 5.5 & 0.2 \\ 6 & 6.5 & 0.3 \end{pmatrix}$$

Gao and Luo(2009) afford us the index weight $\omega = (0.3, 0.2, 0.5)$. The weighted normalized matrix is:

$$X = \begin{pmatrix} 0.069 & 0.025 & 0.118 \\ 0.09 & 0.039 & 0.132 \\ 0.056 & 0.053 & 0.103 \\ 0.034 & 0.039 & 0.059 \\ 0.051 & 0.046 & 0.088 \end{pmatrix}$$

The ideal optimal sequence is $X^+ = (0.09, 0.053, 0.132)$, the ideal worst sequence is $X^- = (0.034, 0.025, 0.059)$. The optimal correlation coefficient matrix r^+ is:

$$r^+ = \begin{pmatrix} 0.63 & 0.562 & 0.718 \\ 1 & 0.717 & 1 \\ 0.516 & 1 & 0.557 \\ 0.396 & 0.717 & 0.333 \\ 0.486 & 0.832 & 0.455 \end{pmatrix}$$

the worst correlation coefficient matrix r^- is:

$$r^- = \begin{pmatrix} 0.514 & 1 & 0.384 \\ 0.395 & 0.729 & 0.332 \\ 0.627 & 0.569 & 0.454 \\ 1 & 0.729 & 1 \\ 0.677 & 0.639 & 0.555 \end{pmatrix}$$

The integrated grey correlation coefficient matrix is:

$$R = \begin{pmatrix} 0.202 & 0.406 & 0.121 \\ 0.08 & 0.254 & 0.063 \\ 0.301 & 0.133 & 0.202 \\ 0.511 & 0.254 & 0.562 \\ 0.339 & 0.189 & 0.302 \end{pmatrix}$$

The uncertain degree of three indices are $DOI(I_1) = 0.143$, DOI(I2) = 0.118, DOI(I3) = 0.137.

The overall uncertainty mass function matrix is:

$$M = \begin{pmatrix} 0.196 & 0.108 & 0.203 \\ 0.257 & 0.170 & 0.229 \\ 0.159 & 0.232 & 0.178 \\ 0.098 & 0.170 & 0.102 \\ 0.147 & 0.201 & 0.152 \\ 0.143 & 0.118 & 0.137 \end{pmatrix}$$

The overall uncertainty mass functions are $m_1(6) = 0.143$, m2(6) = 0.118, m3(6) = 0.137.

In this case, the identification framework Θ is information system. That is $\Theta = \{A_1, A_2, A_3, A_4, A_5\}$, $2^{\Theta} = \{\{A_1\},\{A_2\},\{A_3\},\{A_4\},\{A_5\},\{A_1, A_2, A_3, A_4, A_5\}\}$. Based on the rule of combination, calculate the confidence function of each subset:

$$bel(A_1) = (m_1 \oplus m_2 \oplus m_3)(A_1) = 0.1839$$

$$bel(A_2) = (m_1 \oplus m_2 \oplus m_3)(A_2) = 0.4324$$

$$bel(A_3) = (m_1 \oplus m_2 \oplus m_3)(A_3) = 0.1885$$

$$bel(A_4) = (m_1 \oplus m_2 \oplus m_3)(A_4) = 0.0286$$

$$bel(A_5) = (m_1 \oplus m_2 \oplus m_3)(A_5) = 0.1113$$

$$bel(A_1, A_2, A_3, A_4, A_5) = (m_1 \oplus m_2 \oplus m_3)(A_1, A_2, A_3, A_4, A_5) = 0.055$$

According to maximum principle of confidence function, sequence the risk value: $A_2 > A_3 > A_1 > A_5 > A_4$. The overall uncertainty is on the decrease in the integration process, it reduce from 13.3% which is the initial mean to 5.5%. This shows that the combination of methods could effectively deal with a multi-source uncertain information, reduce the uncertainty and improve the reliability.

CONCLUSION

Nowadays, information systems are widely used in various industries, ensuring information security become the core work of information risk management. By combining grey incidence with D-S theory of evidence, this paper proposed a method of information security risk assessment to deal with the uncertain information. The risk assessment is illustrated with an example, it shows that the method is effective, feasible and can take advantage of the information of each index. Meanwhile, it provides a comprehensive scientific and practical new idea for information system security and has a strong

guiding significance. However, no matter which method you choose, it can only reduce the impact of the uncertainty on results to some extent, but the uncertainty can not be eliminated completely.

ACKNOWLEDGMENT

The authors would like to acknowledge the supports by National Natural Science Foundation of China (No. 71271084; 71071054).

REFERENCES

Chen, S.J. and S.M. Chen, 2003. Fuzzy risk analysis based on similarity measures of generalized fuzzy numbers. IEEE Trans. Fuzzy Syst., 11: 45-55.

Demotier, S., W. Schon and T. Denoeux, 2006. Risk assessment based on weak information using belief functions: A case study in water treatment. IEEE Trans. Syst. Man Cybernetics Part C: Appl. Rev., 36: 382-396.

Fu, Y., 2010. An approach for information systems security risk assessment on fuzzy set and entropy-weight. Acta Electronica Sinica, 38: 1489-1494.

Fu, Y., X.P. Wu and C. Yan, 2006. The method of information security risk assessment using bayesian networks. Wuhan Univ. (Nat. Sci. Ed.,), 52: 631-634.

Gao, H.S. and J. Zhu, 2008. Security risk assessment model of network based on D-S evidence theory. Comput. Eng. Appl., 44: 157-159.

Gao, Y. and J. Luo, 2009. Information security risk assessment based on grey relational decision-making algorithm. J. Southeast Univ. (Nat. Sci. Ed.,), 39: 225-229.

Li, P. and S.F. Liu, 2011. Interval-valued intuitionistic fuzzy numbers decision-making method based on grey incidence analysis and D-S theory of evidence. Acata Automatica Sinica, 37: 993-998.

Zhao, D.M., J.H. Wang, J. Wu and J.F. Ma, 2005. Using fuzzy logic and entropy theory to risk assessment of the information security. Proceedings of the 4th International Conference on Machine Learning and Cybernetics, Volume 4, August 18-21, 2005, Guangzhou, China, pp: 2448-2453.

Zhao, L. and Z. Xue, 2009. Mult-i attribute group decision-making information system security assessment based on VPRS. J. ShangHai JiaoTong Univ., 43: 1161-1166.