A Security Risk Analysis Method for Information System Based on Information Entropy

Sha Fu*, Zhongli Liu, Hangjun Zhou, Wenbin Liu and Bo Li

Department of Information Management, Hunan University of Finance and Economics, Changsha, 410205, People's Republic of China

Abstract: According to the problem of the uncertain information is difficult to quantify in information systems security risk analysis process, proposed an information system security risk analysis method based on information entropy. This method use information entropy to measure the risk of information systems, introduce the information entropy theory, and get the value of the risk of various risk factors with the combination of qualitative analysis and quantitative calculation, to evaluate the risk factors of concern in the system and take appropriate control measures. The paper constructs an information system security risk analysis model and through the case analysis verified the proposed method can be effectively applied to information system security risk analysis.

Keywords: Entropy-weight coefficient, fault tree analysis, information entropy, information system security, risk analysis, weight vector.

1. INTRODUCTION

Now human being has stridden into the global information era. The information has extended to all fields in the country and society. When people are enjoying huge convenience brought by information, they also bear unprecedented information influence and control. With increasing dependency of the country economy and social development on information and information system, the main information systems, which are required by critical infrastructures such as power, telecommunication, traffic and bank and directly decide the national interest and people's livelihood, are suffering from the traditional physical damage and emerging virtual attacks, so the facing security risks and threats are becoming more severe. To ensure normal and secure operation of the information system, it is required to find severe defects which may lead to crash-down. One of effective means to solve these problems is effective security risk evaluation for the information system.

Based on survey on the research references at home and abroad, the information system security analysis is performed mainly by using OCTAVE method, Markov method, Grey Theory, Analytic Hierarchy Process (AHP), Comprehensive Fuzzy Assessment method, BP neural network and Probabilistic Risk Analysis (PRA) method. Now some achievements are achieved [1-6]. Ashish Gehani [7] proposed a risk management and decision model based on the host. Daniel Bilar [8] proposed an information security risk decision method based on the vulnerabilities in the software running on the network hosts, which can be utilized and lead to different dangerous effects and risks. If overall risk of different software is over the threshold set by the system administrator, a security decision should be made to reduce overall risk. Peng Junhao [9] proposed security risk measurement model based on utility, which introduces the risk function and relative loss effect for risk measurement, but this method is not suitable for analysis on hierarchical model. Zhao Dongmei [10] proposed the fuzzy risk evaluation model of the information system, which gets the final comprehensive assessment *via* weighted sum. Although it is a simple handling method, higher subjective preference exists in comprehensiveness and accuracy. The information system risk analysis methods are plentiful, their analysis emphasis and applications are different, and the risk analysis features higher subjective randomness and higher error, so it is difficult to perform objective and accurate assessment.

Based on weaknesses in above research, this paper constructs a hierarchical model for information system security risk measurement and analysis by using the fault analysis method to solve the defect of other methods which are not suitable for analysis on hierarchical risk coefficients and proposes an information system security risk analysis method based on the information entropy, which is purposeful, simple and practicable and can further improve accuracy of quantitative assessment of the information system risks.

2. INFORMATION ENTROPY THEORY

The concept “entropy” was first proposed by Clausius in 1865. Later, C.E. Shannon, a founder of information theory, introduced the concept “entropy” into the information area in 1948 and utilized the “information entropy” to measure disorder degree of information [11]. Shannon applied the entropy to measure uncertainty or information quantity of a random event, which lays theoretical foundation for modern information theory and extends the entropy quantification to research on system uncertainty and disorder quantification [12]. The concept "entropy" can reflect uncertainty degree of the system under the micro status from the micro view.

Assume that the system is under n different statuses, namely $\{S_1,S_2,\ldots,S_n\}$,P_i indicates probability of the system under the status S_i, wherein $i = 1,2,\ldots,n,0\leq P_i\leq 1$ and $\sum_{i=1}^{n} P_i = 1$, the entropy of the system X is:

$$H = -\sum_{i=1}^{n} P_i \ln P_i \tag{1}$$

The entropy has an extreme. When the system statuses have the equal probability, namely $P_i = 1/n$, the entropy has the maximum $H_{max} = \ln m$. H_{max} is used to normalize the equation (1) to get the relative importance entropy of the security risk coefficient A_i:

$$e_i = -\frac{1}{\ln m}\sum_{j=1}^{m} p_{ij} \ln p_{ij} \tag{2}$$

when p_{ij} ($j = 1,2,\ldots,m$) has equal values, the entropy e_i has the maximum 1, namely e_i meets the condition $0\leq e_i\leq 1$. When the entropy reaches the maximum, (1-e_i) is used to measure the weight of the security risk coefficient A_i. After normalization, the weight ϕ_i of the risk coefficient A_i is:

$$\phi_i = \frac{1}{n-E}\left(1-e_i\right) \tag{3}$$

wherein $E = \sum_{i=1}^{n} e_i$, ϕ_i meets the condition $0\leq \phi_i \leq 1$, $\sum_{i=1}^{n} \phi_i = 1$.

3. RISK ANALYSIS METHOD BASED ON INFORMATION ENTROPY

3.1. Related Concepts of Fault Tree Analysis Method

The fault tree analysis (FTA) is a tool to analyze the system reliability and security and is an evolutionary analysis method from the effect to the cause. This method regards the easily observed faults as the top event of the fault tree and regards the final cause of the fault leading to this top event as the bottom event. The middle events reflect causal relationship relation between top events and bottom events [13]. All possible causes leading to top event and their mutual relations are analyzed in details and the fault tree is constructed by the causal relation between the fault phenomena and causes. The weak steps and critical parts of the system, measures to take and security requirements are identified based on the analysis results.

3.2. Construct Fault Tree

This paper constructs a fault tree according to "practicable rules for information security management" -ISO/IEC 27002 standard by using the fault tree analysis method. The specific process is described as follows:

1). Identify top event of the fault tree, namely assess general conditions of the information system security risks.

2). The assessment structure is divided into the management elements, control target and control measure according to ISO/IEC 27002 standard. Three levels are cause-effect relation. Security elements of different levels are the causes of higher level and are effects of lower level. The fault tree analysis method is used to analyze logic relation among security elements at different levels based on this feature for complete and effective extraction of security coefficients in the standard.

3). Identify the bottom event of the fault tree, namely select risk coefficients at the bottom level of the fault tree by referring to the control measures in ISO/IEC 27002 standard, the events at the bottom level are not crossed to meet the requirements of the fault tree analysis method.

Construct the fault tree for information system security risk according to the above qualitative analysis, shown as the Fig. (**1**), and compute the comprehensive security risk value of the information system *via* the information entropy risk analysis algorithm. The Fig. (**1**) indicates level 1, level 2 and level 3 by using α, β and γ.

3.3. Weight and Conformity Instruction of Information Entropy

Based on the fault tree analysis, the entropy weights of the risk coefficients at different levels are regulated according to the structural features of the ISO/IEC 27002 standard. L1 of the fault tree includes X management elements, which reflect several security coefficients in overall risks of the information system. The weight is expressed as α_i. L2 of the fault tree indicates that each management element should be reflected *via* several control targets. The weight of j^{th} control target of i^{th} management element is expressed as $\beta_{i,j}$. L3 of the fault tree indicates that each control target is reflected *via* implementation of several control measures. The weight of k^{th} control measure in j^{th} control target in i^{th} management element (namely conformity) is expressed as $\gamma_{i,j,k}$. The conformity is quantified as 5 levels, shown as the Table **1**.

3.4. Description of Information Entropy-Based Risk Analysis Method and Evaluation Steps

The information entropy-based risk analysis method constructs a fault tree according to the architecture in the ISO/IEC 27002 standard, gets original data according to assessment practices, computes risk value of the risky coefficients by using the information entropy analysis algorithm, and combines the qualitative analysis with the quantitative computing, so it can overcome insufficient analysis on the influence degree of control measures on risks in the standard and ensure creditability and repeatability of the risk analysis conclusions. The security risk is completely computed for the constructed security risk fault tree of the information system. The specific steps are described as follows:

(1) Identify quantified value of the conformity. Conformity can reflect conformity between security conditions and security control measures of the information system. The quantified values can take a value between two lev-

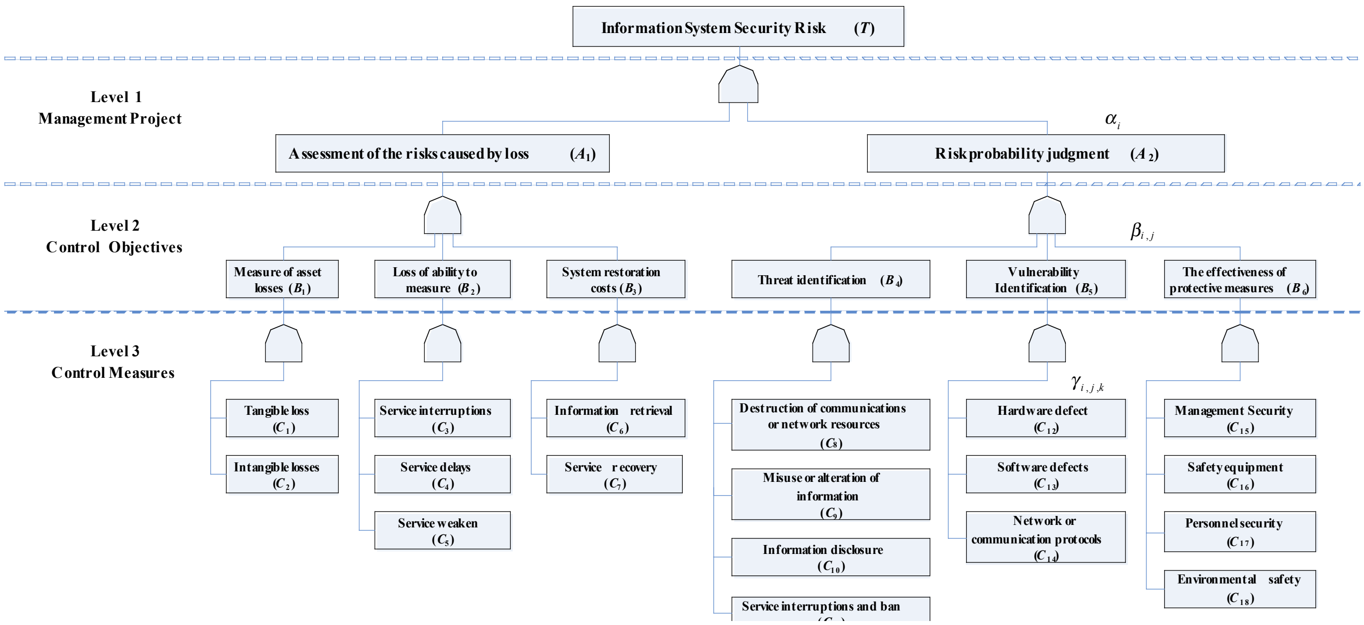
Information System Security Risk (*T*)
Level 1
Management Project
α_i
Assessment of the risks caused by loss (A_1)
Risk probability judgment (A_2)
Level 2
Control Objectives
$\beta_{i,j}$
Measure of asset losses (B_1)
Loss of ability to measure (B_2)
System restoration costs (B_3)
Threat identification (B_4)
Vulnerability Identification (B_5)
The effectiveness of protective measures (B_6)
Level 3
Control Measures
$\gamma_{i,j,k}$
Tangible loss (C_1)
Intangible losses (C_2)
Service interruptions (C_3)
Service delays (C_4)
Service weaken (C_5)
Information retrieval (C_6)
Service recovery (C_7)
Destruction of communications or network resources (C_8)
Misuse or alteration of information (C_9)
Information disclosure (C_{10})
Service interruptions and ban
Hardware defect (C_{12})
Software defects (C_{13})
Network or communication protocols (C_{14})
Management Security (C_{15})
Safety equipment (C_{16})
Personnel security (C_{17})
Environmental safety (C_{18})

Fig. (1). Fault tree of information system security risk.

Table 1. Definition of conformity.

Conformity	Quantified value	Conformity measure
Not compliant	0	No corresponding security control measures
Not very compliant	0.25	Certain security control measures are available, but they can not reach the security baseline requirement or only play partial role
Rough complaint	0.5	Better security control measures are taken, but they can not reach a better effect
Better compliant	0.75	The security control measures can roughly reach the security baseline requirement or have better effect
Fully compliant	1	The security control measures fully comply with the security baseline requirement and are completely executed.

els according to the actual conditions. *E.g.* The current security conditions are between basic conformity and better conformity. To reflect the subjective conditions, the conformity can be 0.6 or 0.65. assume the corresponding conformity of kth control measure in jth control target in ith management element of the existing fault tree, it is substituted in the following equation for calculation:

(2) Calculate the entropy weight coefficient $\beta_{i,j}$ and risk weight vector $\phi_{i,j}$ of control targets of L2 according to the equation (4) and (5).

$$\beta_{i,j} = -\frac{1}{\ln m}\sum_{k=1}^{m}\gamma_{i,j,k}\ln\gamma_{i,j,k} \tag{4}$$

$$\phi_{i,j} = \frac{1}{n-E}\left(1-\beta_{i,j}\right) \tag{5}$$

wherein $E=\sum_{j=1}^{n}\beta_{i,j}$,$\phi_{i,j}$ meets :$0\leq\phi_{i,j}\leq 1$, $\sum_{j=1}^{n}\phi_{i,j}=1$.

(3) Compute the entropy weight coefficient α_i and risk weight vector ϕ_i of management elements of L1 according to the equation (6) and (7):

$$\alpha_i = -\frac{1}{\ln m}\sum_{j=1}^{m}\beta_{i,j}\ln\beta_{i,j} \tag{6}$$

$$\phi_i = \frac{1}{n-E}\left(1-\alpha_i\right) \tag{7}$$

wherein $E=\sum_{i=1}^{n}\alpha_i$,ϕ_i meets the condition $0\leq\phi_i\leq 1$, $\sum_{i=1}^{n}\phi_i = 1$。

(4) Compute the entropy weight coefficient *e* and overall risk weight vector ϕ of general security risk of the information system.

$$e = -\frac{1}{\ln m}\sum_{i=1}^{m}\alpha_i\ln\alpha_i$$

$\phi = 1 - e$ and $0 \le e, \phi \le 1$

(5) Decide the action priority, make corresponding measures and reasonable and feasible risk handling scheme, and perform internal auditing and management review according to the overall risk weight vector of the information system based on the risk level description and quantitative representation in the Table **2** in order to monitor the risks.

4. APPLICATION CASE

Multiple experts and consultants are invited to assess security risk fault tree of the information system based on their professional knowledge and rich experiences according to related description of the security risk control points of the information system.

Step 1: Prepare: Based on the definition of weight vector and conformity, the conformity $\gamma_{\mathrm{i,j,k}}$ of 18 control measures (6 groups) of L3 of the fault tree obtained from the expert's comments is {(0.65,0.50), (0.75,0.60,0.60), (0.55,0.80), (0.45, 0.80,0.65,0.75), (0.50,0.75,0.75), (0.65,0.70,0.85,0.70)}.

Step 2: Compute the entropy weight coefficient $\beta_{\mathrm{i,j}}$ and risk weight vector $\phi_{i,j}$ of 6 control targets of L2 according to equation (4) and (5) as follows:

$$\beta_{1,1} = -\frac{1}{\ln m}\sum_{k=1}^{m}\gamma_{1,1,k}\ln\gamma_{1,1,k}$$
$$= -\frac{1}{0.6931}\left[0.65\times\left(-0.4308\right)+0.50\times\left(-0.6931\right)\right]$$
$$= 0.9040$$

Similarly, we can get:

$\beta_{1,2}$=0.7544, $\beta_{1,3}$=0.7319

$\beta_{2,1}$=0.7456, $\beta_{2,2}$=0.7083, $\beta_{2,3}$=0.6618

$$E = \sum_{j=1}^{n}\beta_{1,j}$$
$$= \left(0.9040 + 0.7544 + 0.7319\right) = 2.3903$$

$$\phi_{1,1} = \frac{1}{n-E}\left(1-\beta_{1,1}\right)$$
$$= \frac{1}{3-2.3903}\left(1-0.9040\right) = 0.1575$$

Similarly, we can get:

$\phi_{1,2} = 0.4028$, $\phi_{1,3} = 0.4397$

$$E = \sum_{j=1}^{n}\beta_{2,j}$$
$$= \left(0.7456 + 0.7083 + 0.6618\right) = 2.1157$$

$$\phi_{2,1} = \frac{1}{n-E}\left(1-\beta_{2,1}\right)$$
$$= \frac{1}{3-2.1157}\left(1-0.7456\right) = 0.2877$$

Similarly, we can get:

$\phi_{2,2} = 0.3299$, $\phi_{2,3} = 0.3824$

Step 3: compute the entropy weight coefficient α_i and risk weight coefficient ϕ_i of 2 management elements of L1 according to the equation (6) and (7) as follows:

$$\alpha_1 = -\frac{1}{\ln m}\sum_{j=1}^{m}\beta_{i,j}\ln\beta_{i,j}$$
$$= -\frac{1}{1.0986}\begin{bmatrix}0.9040\times\left(-0.1009\right)+0.7544\times\left(-0.2818\right)\\ +0.7319\times\left(-0.3121\right)\end{bmatrix}$$
$$= 0.4845$$

Similarly, we can get:

$\alpha_2 = 0.6703$

$$E = \sum_{i=1}^{n}\alpha_i$$
$$= 0.4845 + 0.6703 = 1.1548$$

$$\phi_1 = \frac{1}{n-E}\left(1-\alpha_1\right)$$
$$= \frac{1}{2-1.1548}\left(1-0.4845\right) = 0.6099$$

Similarly, we can get:

$\phi_2 = 0.3901$

Step 4: Compute the entropy weight coefficient e and overall risk weight vector ϕ of the overall security risk of the information system as follows:

$$e = -\frac{1}{\ln m}\sum_{i=1}^{m}\alpha_i\ln\alpha_i$$
$$= -\frac{1}{0.6931}\left[0.4845\times\left(-0.7246\right)+0.6703\times\left(-0.4000\right)\right]$$
$$= 0.8934$$
$$\phi = 1 - e = 0.1066$$

Table 2. Risk level description and quantitative expression.

Risk level	*Quantitative expression of risk*	*Risk level*	*Measrue to take*
1	$0 < \phi \le 0.2$	*low* risk	accept risk
2	$0.2 < \phi \le 0.4$	*plain* risk	notice and prevention
3	$0.4 < \phi \le 0.6$	*medium* risk	enhance prevention and control
4	$0.6 < \phi \le 0.8$	*high* risk	inform related department of purposeful control
5	$0.8 < \phi \le 1.0$	*very* high risk	take control measure to reduce risk or change system

Table 3. Computing results of risk weight vector.

Level					*Overall risk weight vector* $\phi = 0.1066$						
1			$\phi_1 = 0.6099$						$\phi_2 = 0.3901$		
2	$\phi_{1,1} = 0.1575$		$\phi_{1,2} = 0.4028$		$\phi_{1,3} = 0.4397$		$\phi_{2,1} = 0.2877$		$\phi_{2,2} = 0.3299$		$\phi_{2,3} = 0.382$
3	$\gamma_{1,1,1}$ 0.65	$\gamma_{1,1,2}$ 0.50		$\gamma_{1,2,1}$ 0.75	$\gamma_{1,2,2}$ 0.60	$\gamma_{1,2,3}$ 0.60		$\gamma_{1,3,1}$ 0.55	$\gamma_{1,3,2}$ 0.80		
	$\gamma_{2,1,1}$ 0.45	$\gamma_{2,1,2}$ 0.80	$\gamma_{2,1,3}$ 0.65	$\gamma_{2,1,4}$ 0.75	$\gamma_{2,2,1}$ 0.50	$\gamma_{2,2,2}$ 0.75	$\gamma_{2,2,3}$ 0.75	$\gamma_{2,3,1}$ 0.65	$\gamma_{2,3,2}$ 0.70	$\gamma_{2,3,3}$ 0.85	$\gamma_{2,3,4}$ 0.70

The overall risk of the information system and risk weight vector of risk coefficients of different levels are assessed by analyzing above data. The arranged results are shown as the Table **3**.

It is determined that the overall risk weight vector of the information system $\phi = 0.1066$ from the Table **2**, so it is low-risk, the security risk level is low, and the system is relatively secure and reliable. Considering the weight vector of management elements of L1, if the risk value of "risk occurrence probability assessment" is 0.3901, it is a plain risk level. If the risk value of "evaluation of risk-caused loss" is 0.6099, it is high risk level. The latter is the main coefficient leading to overall risk of the information system and is also the key in risk management, so effective measures should be taken to reduce severity of the effect. Considering the weight vector of the control targets of L2, the weight vectors of the risk coefficients of 6 control target are balanced. The risk values are mainly distributed at the plain risk level. The instance analysis indicates that the risk analysis results of the information system based on the information entropy can better match the manual evaluation results of the security experts, so it indicates that this risk analysis method is scientific and reasonable.

CONCLUSION

This paper constructs the fault tree based on qualitative analysis according to the requirements and architecture in ISO/IEC 27002 standard, proposes a security risk analysis method of the information system based on the information entropy, and constructs the security risk analysis model of the system. This method solves the problem of difficult quantification of the uncertain information in security risk analysis of the information system, can accurately and quantitatively evaluate risk conditions of the information system, and provide a scientific and feasible new idea for comprehensive assessment of security risk of the information system. in addition, we will further study how to gradually correct and perfect this analysis method according to the change of the actual conditions by continuously applying this analysis method in practices in future.

CONFLICT OF INTEREST

The authors confirm that this article content has no conflict of interest.

ACKNOWLEDGEMENTS

This study was supported by the Scientific Research Fund of Hunan Provincial Education Department (grant code: 14C0184) and also supported by the Teaching Reform Project of Hunan Provincial Colleges and Universities (grant code: 20130511). Supported by the construct program of the key discipline in Hunan province.

REFERENCES

[1] D. Mats, "Generalized evaluation in decision analysis", *European Journal of Operational Research*, vol. 162, no. 7, pp. 442-449, 2005.

[2] S. Dzazali, and A. Zolait, "Assessment of information security maturity: an exploration study of Malaysian public service organizations", *Journal of Systems and Information Technology*, vol. 14, no. 1, pp. 23-57, 2012.

[3] M. Ahmed, and L. Sharif, "Information security: securing a network device with passwords to protect information", *TRIM*, vol. 6, no. 1, pp. 62-76, 2011.

[4] D. Qin, L. Zhang, and J. Li, "Risk assessment approach for information security based on FAHP", *Computer Engineering*, vol. 35, no. 15, pp. 156-158, August 2009.

[5] W. Zhang, B. Du, and Y. Yang, "Application of fuzzy analytic hierarchy process to TV and radio information assurance evaluation indicator systems", *Acta Electronica Sinica*, vol. 36, no. 10, pp. 2061-2064, October 2008.

[6] Y. Fu, X. Wu, and C. Yan, "The method of information security risk assessment using bayesian networks", *Journal of Wuhan University(Natural Science Edition)*, vol. 52, no. 5, pp. 631-634, October 2006.

[7] A. Gehani, "*Support for Automated Passive Host-based Intrusion Response*", Ph. D. thesis, Duke University, 2003.

[8] D. Bilar, "*Quantitative Risk Analysis of Computer Networks*", Ph. D. thesis, Dartmouth College, 2003.

[9] J. Peng, G. Xu, Y. Yang, and Y. Tang, "Measure model of security risk based on utility", *Journal of Beijing University of Posts and Telecommunications*, vol. 29, no. 2, pp. 59-61, 69, April 2006.

[10] D. Zhao, J. Ma, and Y. Wang, "Model of fuzzy risk assessment of the information system", *Journal on Communications*, vol. 28, no. 4, pp. 51-56, 64, April 2007.

[11] J. Liang, and Y. Qian, "Information granule and entropy theory in information system", *Science in China (Series E:Information Sciences)*, vol. 38, no. 12, pp. 2048-2065, December 2008.

[12] W. Ding, J. Wang, Z. Guan, and H. Zhu, "Algorithm of attribute reduction based on extension entropy of variable precision thresholding in incomplete information system", *Journal of Chinese Computer Systems*, vol. 31, no. 12, pp. 2372-2376, December 2010.

[13] L. Zheng, L. Song, R. Guo, and J. Zhang, "Application of FAT in information security risk assessment", *Computer Science*, vol. 38, no. 10A, pp. 106-108, 118, October 2011.

Received: September 22, 2014 Revised: November 30, 2014 Accepted: December 02, 2014