Method on network information system security assessment based on rough set

Wang Qiangmin¹, Lin Mengquan², Li Jianhua²

¹ School of Information Security Engineering, Shanghai Jiaotong University, Shanghai 200030, P. R. China
qmwang@sjtu.edu.cn

² Department of Electronic Engineering, Shanghai Jiaotong University, Shanghai 200030, P. R. China
linmq@cdgdc.edu.cn, lijianhua888@sjtu.edu.cn

Abstract

Based on analyzing the concept of network information system security assessment, this paper proposed an information system security basic attribute model for information system security assessment based on optimal security tree criteria. Then a novel heuristic rough set attributes reduction algorithm was proposed to reduce the attributes and build weight set. The validity and practical value of it were illustrated by an example. The main idea of the security assessment is improve the assessment efficiency and reduce the cost.

1. Introduction

With the rapid development of information technology, Internet has become an important platform for global information transmission. It has broken the geography boundary of information exchange, and changed the traditional mode of information exchange. With some Internet properties of opening, universality and complexity, the security and reliability of information transmission and storage are facing with more and more severe threats. It has been one of the important parts of society security, and closely related with daily work, life and study activity. It even threatens the military and economy security of a country. Therefore, it is necessary to know about the security state of an information system, and to manage and control the security problems. The security assessment is an important means to get a better understanding to the information system security state. Its results are important bases to build up a network security solution.

There are two security assessments involved in information security research. One is risk assessment, the other is security assessment. The risk assessment of an information system aims to identify threats and existing vulnerabilities, which are the opportunities of the intruders’ side. Information system security assessment extends the work of risk assessment. Its objective is the evaluation of the capabilities of defense system. Both assessments will involves vulnerability testing, but the follow-up tasks are different. After security assessment, security policy and implementation procedures are set up. On the other hand, after security assessment, the weak points of a defense system are boosted to maintain the required standards of a security program. There are already many studies on IS risk assessments[1-4]. Recently, IS security assessment have also aroused some researchers’ interests[5-7]. Shuchuan Chao[6] developed an information security assessment model to evaluate the security system. They organized these components as a two-layer structure – security controls and sub-security controls. They used web-survey to get the different importance levels of security controls and sub-security controls. Hallberg, J. [7] proposed a framework for information system assessment.

2. The basic model for Information system security assessment

The information system security assessment model based on rough set that we studied is based on the tree model of information system. It is illustrated in figure 1.

In this model, the security level of the overall system is affected by the single security factor of the system and the weight of every security. Although there are several advantages in the attributes tree model of information system the attribute [5], there are too many security factors included in an information system. According to the security criteria cc and BS 7799, there are more than 350
factors related with an information system. When we assess the security level of an information system, we should assess every factor of the end leaves in the attribute tree model and calculate the weighted factors to get the overall result. The accumulated calculating error may cause assess deviation of overall security level, it even get the error result. Furthermore, constructing the weight of the security factors is also a difficult problem. At present, most of the methods to get the weight of security factors are using AHP and Delphi. These methods need a mass of samples or expert consultation.

3. The basic problem of security factors optimization

There are some correlations among the security factors of the information system assessment model based on tree structure. These factors with high correlations can be removed through using correlation theory. This can avoid calculating these factors repeatedly. However correlation theory could realize only with assured data. When rough set theory is used, we could accumulate some information system assessment datum, then use reduction principle to simplify the model and get the more compact and easier realized security assessment model.

Rough set theory, proposed by Pawlak in 1982[8], is a tool for dealing with uncertainty and vagueness aspects of knowledge model. The main idea of rough sets corresponds to the lower and upper approximations based on equivalence relations. It could be used to identify the correlation among the factors and remove the redundant data. It has been widely used in the area of AI, information system analyzing, data mining, pattern recognition, fault diagnostics, etc. When we optimize the model using rough set theory, it does not need to provide any other transcendental knowledge except the data provided to solve problems. It can discover the connotative knowledge only if we classify the measured data itself and find the underlying rules among the data. Under the condition of keeping the most important data information, it can reduce the data set and get the minimum expression of the knowledge.

Researchers always wish to get the minimum factors from reduction because that the fewer the factors are, the lower the cost is when we use the reductive model to assess an information system. However, when we use rough set theory to deal with the database of the information system or decision table and reduce the data, it may obtain several reduction results.

In fact, the process of looking for the minimum reduction of information system and decision table is iterative process. It has been proven that computing the optimal (minimal) reduction of decision table is a NP-hard problem by Wong S.K.M. and Ziarko[9]. Therefore before applying rough set theory to assessment model, we should optimize the security factors of an information system.

4. Constructing security assessment system model

According to the above analyzing, it could be found that the fewer factors provided for rough set reduction, the easier the reduction algorithm is realized. In this paper, we proposed a method, which is according to assessment experience and knowledge of information security experts, to analyze and classify the original security factors of an information system tree structure and find the proper nodes on the tree structure model as the basic factors of security assessment model. For example, In figure 1, there are several security factors such as “access control policy”, “user registration”, “user password management” and etc. All these factors can be included in the upper layer factor “access control”.

When assessing an information system security, we only need collect the security factor “access control” and look its including contents as the assessment meaning. This will help the assessor to easily obtain the overall performance of “access control”. It will also reduce dramatically the assessment factors that need to be collected and processed, reduce assessment cost and improve the assessment efficiency. We named the information system security assessment model, which is constructed based on the basic factors, as the optimal node security assessment model (ONSAM). The ONSAM model has several advantages.

1. It can reduce the assessing factors and benefit reduction algorithm.
2. It also can reduce the accumulate error due to the fewer assessing factors.
5. Hybrid heuristic rough set attributes reduction assessment model

Using rough set theory, we could analyze the accumulated assessing data from information system and reduce the assessment factors. This can optimize the assessment model. In this paper, based on the ONSAM model, we discuss how to describe the information system security model based on rough set and reduction process.

5.1 Description of information system security model based on rough set

A security assessment original information system can be described by a 4-tuple in rough set theory.

\[ S = (U, A, V, f) \]  

where \( U \) represents a finite set of objects, in this case, it is same type information system objects, i.e. a record in decision table, \( U = \{x_1, x_2, \cdots, x_N\} \), \( x_j \) is basic object \( (i=1,2,\cdots,N) \) : \( A \) refers to a finite set of attributes, \( A = \{a_1, a_2, \cdots, a_M\} \), \( a_j \) is a attribute of the objects \( (j=1,2,\cdots,M) \); \( V \) is the union of all the domains of the attributes in \( A \), \( A = C \cup D \), \( C \cup D = \Phi \). The attribute set \( A \) often consists of two subsets, one refers to condition attributes \( C \). These values are obtained from the collected information in assessing process; \( D \) is decision attributes (i.e. the assessing results set \( R \)), which represent security level of an information system. The information system with condition attributes and decision attributes is entitled decision table. \( V \) is the union of all the domains of the attributes in \( A \). \( V = \{v_1, v_2, \cdots, v_M\} \), \( v_j \) is the domain of the attribute \( a_j \) \( (j=1,2,\cdots,M) \) : \( f \) is information function, \( f: U \times A \rightarrow V \), \( f(x_i, a_j) \in v_j \) is the domain of the attribute \( a_j \) of the basic object \( x_j \) \( (i=1,2,\cdots,N; j=1,2,\cdots,M) \).

The security assessment information system represented by formula (1) may exist dependency among the attributes, repetition among the decision rules and some redundant attributes. Using rough set theory to the security assessment information system, we can reduce the attributes and decision rules and get the equivalent reduced security assessment information:

\[ S^0 = (U^0, A^0, V^0, f^0) \]  

where \( U^0 = \{x^0_1, x^0_2, \cdots, x^0_n\} \subseteq U = \{x_1, x_2, \cdots, x_N\} \), \( n \leq N \), \( A^0 = \{a^0_1, a^0_2, \cdots, a^0_m\} \subseteq A = \{a_1, a_2, \cdots, a_M\} \), \( m \leq M \), \( C^0 \subseteq C \), and \( D^0 \subseteq D \).

Constrcuting the security assessment information system original decision table is the base of rough set reduction. Collecting the assessment data from \( n \) same kinds of network information systems, we construct the security assessment information system original decision table. It is shown in table 1. Then we can use the attributes reduction of rough set to evaluate the significance of the security factors and remove these security factors that do not affect the assessment results.

| Table 1 security assessment information system original decision table |
|------------------------|------------------------|------------------------|
| \( S = (U, A, V, f) \) | \( A = C, D > C, D \) |
| \( a_1 \)                | \( \cdots \)            | \( a_M \)                |
| \( x_1 \)                | \( v_{11} \)            | \( \cdots \)            | \( v_{1M} \)            |
| \( x_2 \)                | \( v_{21} \)            | \( \cdots \)            | \( v_{2M} \)            |
| \( \cdots \)            | \( \cdots \)            | \( \cdots \)            | \( \cdots \)            |
| \( x_N \)                | \( v_{N1} \)            | \( \cdots \)            | \( v_{NM} \)            |

5.2 Attribute importance degree heuristic reduction

One fundamental aspect of rough set theory involves a search for particular subsets of condition attributes. Such subsets are called attribute reductions. Many types of attribute reductions have been proposed, each of the reductions aimed at some basic requirements. Skowron A. introduced the notion of discernibility matrix which became a major tool for searching for reductions in information systems\(^{[10]}\). This method could get all the reduction results because the reduction of decision table is a NP-hard problem. Therefore, it needs resort to heuristic algorithm to solve the problem when the information system has large amount of data.

5.3 Hybrid importance degree heuristic attribute reduction

In this paper, we use the information system factors as a constraint condition and apply it as a heuristic rule to reduction process. The heuristic attribute reduction uses core as a base of reduction. According to the importance degree of the attributes, these attributes are added one by one from high to low until the set become a reduction set that has the same equivalence relation with the original rough set. Then an optimal reduction set could be gained.
We entitle the heuristic attribute reduction as hybrid importance degree heuristic attribute reduction (HIDHAR).

Different kinds of information systems have different requirements on different security factors. For example, financial network system may emphasize information confidentiality while public service network system may emphasize the information availability. We can get the importance degree of the attributes in different kind of network information systems from experts and users’ experiences. We synthesis the importance degree of attribute utilizing the experiences and the reduction attribute importance degree. For example, we calculate the weighted average of these importance degrees as the heuristic rule. This make that the results of reduction is more practical and optimal.

### 5.3.1 The definition of attribute importance degree

Given $S$ is a decision table, $C$ and $D$ are condition attributes and decision attributes respectively. $R \subseteq C$, for any attribute $a \in C - R$, the importance degree of $a$ is defined as:

$$SGF(a, R, D) = k(R \cup \{a\}, D) - k(R, D) \quad (6)$$

where $k(R, D) = card(POS_R(D)) / card(POS_C(D))$ the importance degree of attribute $a$ is also defined as:

$$SGF(a, R, D) = \gamma_{R \cup \{a\}} - \gamma_R \quad (7)$$

$$SGF(a, R, D)$$ could be calculated using the method of reference [6].

The experience attribute importance degree set can be gotten through expert consultation. Then the experience attribute importance degree set knowledge database can be constructed. It will be provided to heuristic reduction algorithm.

$$W_i = \alpha \cdot SGF(a_i, R, D) + (1 - \alpha) \cdot W(a_i), i = 1, 2, \cdots, m \quad (8)$$

where, $W_i$ refers to the synthetic attribute importance degree of attribute $a_i$, $W(a_i) \in [0, 1]$ is the experience attribute importance degree of attribute $a_i$. $\alpha \in [0, 1]$ is a weighted factor which can be changed to let the algorithm to be used in different kinds of information system.

### 5.3.2 The description of HIDHAR

The steps of the HIDHAR attribute reduction can be described as following.

1) Calculate the discernibility equivalence classes $IND(C)$ and $IND(d)$ of the condition attributes set $C$ and the decision attributes $D$. For every attribute subset $B$, the discernibility equivalence classes is given as:

$$IND_{D^B}(B) = \{(x_i, x_j) \in U \times U \mid \forall a \in B, a(x_i) = a(x_j), i \neq j; i, j = 1, 2, \cdots, m\}$$

Calculate the lower approximation set $C(d)$, and check whether the information decision system is consistency. If $POS_{C(d)}(d) = C(d) = U$, this represents the decision table is consistency. General we suppose that the decision system is consistency.

3) Calculate the discernibility equivalence classes $IND(C/\{c_i\})$ of the condition attributes, where $c_i (i = 1, 2, \cdots, m)$, then we can get $POS_{C/\{c_i\}}(d)$.

4) Calculate approximation quality of classification

$$r_{c_{/\{c_i\}}}(d) = \frac{card(POS_{C/\{c_i\}}(d))}{card(U)}$$

5) Calculate the attribute importance degree of the condition attribute $c_i (i = 1, 2, \cdots, m$).

$$r_{c_i}(d) = \frac{card(POS_{C}(d))}{card(U)}$$

$$I(c_i) = r_{c_i}(d) - r_{c_{/\{c_i\}}}(d)$$

$$W_i = \alpha \cdot I(c_i) + (1 - \alpha) \cdot W(a_i), i = 1, 2, \cdots, m$$

6) Calculate the synthetic attribute importance degree

The attributes are simplified according to the calculated results of attributes importance degree. For one attribute $c_i (i = 1, 2, \cdots, m)$, if $W_i \leq \beta$, $\beta \geq 0$, then attribute $c_i (i = 1, 2, \cdots, m)$ is not necessary for decision attribute $D$, otherwise attribute $c_i (i = 1, 2, \cdots, m)$ is necessary for decision attribute $D$, where, the value of $\beta$ is adjusted by experience and practical condition.

7) For these unnecessary condition attributes $(c_i, c_j)$, it is required to check whether the attributes $c_i$ and $c_j$ ($i \neq j$) can be removed at the same time. If the attributes $c_i$ and $c_j$ can not removed at the same time, go to step 3.

### 5.4 An practical application of SIDHAR method

In order to testify the validity of SIDHAR method, we chose the networks of high education institutions as samples. We get 29 samples data from 21 universities by expert consultation. There are 32 security factors that are taken as condition attributes in the consultation table. These factors include
“access control”, “information encryption”, “security audit” and etc. We divided the decision attribute into three levels i.e. “safe”, “satisfying requirement” and “unsafe”.

The last calculated reduction results include 18 security factors, i.e. Password Administration, Security Audit, Management Organization and Policy, Number of Identify Password, Database security, Access Control, Device security, Audit Log, Information Device Security, Environment security, Division Confidentiality Level of Information, Information Encryption, Mark Level of Information Confidentiality, IC Card Physical Protection, Division of Security Domain, Measure of Access Control, Check Policy and Protect from Electromagnetic Leak. The weights of these factors are 0.1856, 0.1568, 0.15415, 0.1057, 0.0852, 0.0625, 0.0625, 0.0341, 0.0341, 0.02275, 0.01705, 0.1705, 0.01705, 0.01135, 0.01135, 0.01135, 0.0057 and 0.0057 respectively. The reduction results indicate that the networks of high education institutions emphasize the security factor on Password Administration, Security Audit, Management Organization and Policy, Number of Identify Password, Database security, Access Control and Device security. This coincides with the practical condition. At the same time, we notice that more than 40 percent of the security factors are reduced by the attribution reduction. This will make the optimal information system security assessment model much simpler than the original model and reduce a large part assess work and cost.

6. Conclusion

Information system security assessment is a complicated and challenging issue. A new idea of combining experts’ experiences and rough set is proposed to structure information system security assessment model. On the base of the model, a heuristic attributes reduction algorithm is put forward based on rough set theory and experts’ experiences. According to the results of rough set attributes reduction in networks of high education institutions, using the reduction security model to assess this kind of network is very efficient. Practical example proves the method is valid. The method is also applied to other kind of information system, such as financial system and government system with different requirement of system security.

7. References


[6] Chao Shu-chuan, An approach to information system security assessment, Dissertation of St Cleveland at University, 2005


