Multidimensional Data Security Exchange Modeling and its Optimization Method

Multidimensional Data Security Exchange Modeling and its Optimization Method

Jindong He[1, *], Zhijun Tang[1], Dan Wu[1], Lijin Wu[1], Chenghua Lin[1] and Fucai Luo[2]

[1]State Grid Fujian Electric Power Research Institute, Fuzhou Fujian 350007, China
[2]State Grid Fujian Electric Power Company Limited, Fuzhou Fujian 350003, China

*Corresponding author e-mail: 2468537873@qq.com

Abstract. In order to improve the security of multi-dimensional data exchange and reduce the risk of information leakage, this paper proposes a multi-dimensional data security exchange modeling machine optimization method. Firstly, the data security exchange model is analyzed. Then, the process behavior evaluation model and Boolean operation of the fusion hidden Markov model are used to evaluate the multi-dimensional data security exchange model, so as to obtain better performance evaluation. An optimal process behavior evaluation model selection method based on cost decision tree is proposed, which coordinates the contradiction between the accuracy rate of the evaluation model and the false alarm rate. Finally, a simulation experiment was carried out and the results were compared to verify the effectiveness and reliability of the proposed method.

1. Introduction

With the rapid development of national economy and social informatization, information security has been highly value [1]. Ensuring information security is seen as a basic strategy to safeguard national security, social stability and public interests, safeguard and promote the healthy development of informatization [2].

In recent years, more and more scholars have paid attention to the research on data security exchange. The literature [3] discusses the network gatekeeper technology, and proposes a security domain isolation and data exchange model based on virtual machine monitor (VMM). Then we give up an implement framework of this model based on XEN. Finally, we discuss the security feature and the future appliance effect of the model [3]. The literature [4] puts forward that data exchange, safety and security are required in many measurement and control systems in which programmable logic controllers are used in recent days, and describe the IEC 61131-3 software model at the first, and then introduce the method to translate this model to the information model of OPC UA. Literature [5] puts forward multilevel security management system can effectively manage the security devices, but how to realize data exchange between multilevel security management is an important problem. This paper designs and realizes data exchange between multilevel security systems. Although there are some in-depth studies on data security, there are few studies on the safety exchange of multi-dimensional data. This paper proposes to model the safety exchange of multi-dimensional data and evaluate the optimal process.

2. Data security exchange

2.1. Data security exchange

Initially, data security exchange was used to solve the exchange problem of heterogeneous data, and later to solve the data security exchange problem of information sharing between information systems of different sensitivity levels [6]. Data security exchange is a data security exchange method aimed at establishing a secure exchange chain between information systems [7].

Data exchange is the process of data exchange between different information systems. The goal of data exchange is to realize data sharing in heterogeneous environment so as to make effective use of data resources in information system. It makes full use of the loose structure, heterogeneous data format and cross-platform data, speeds up the flow of data between information systems, and realizes the integration and sharing of data [8].

Data security exchange is used to meet the operational requirements of data exchange and security sharing between information systems with different sensitivity levels, which not only meets the security isolation requirements of information systems with different sensitivity levels, but also meets the requirements of information flow and sharing between systems with different sensitivity levels under heterogeneous environments [9]. When information systems with different sensitivity levels need to share information, there is a threat of information leakage of information systems with high sensitivity.

2.2. Data security exchange model

In order to guarantee the reliability and control of data exchange process, a data security exchange model is established. The model design is shown in figure 1.

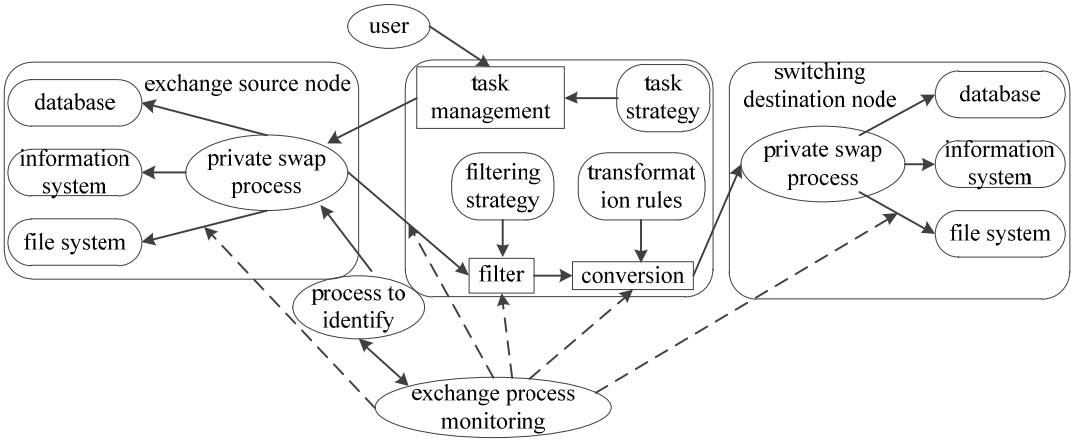

Figure 1. Model design idea

The design idea of the data security exchange model is as follows:

(1) Expand the existing data security enterprise exchange mode ETL to CEFTEL, and provide protection for the implementation of the data exchange process as a whole.

(2) The exchange process is used as the main body to complete the data exchange. Introduces swap tasks, identifies, controls, and supervises dedicated swap processes through the customization of swap tasks.

(3) Through the analysis of some key behaviors in the customized data security exchange, the exchange behaviors with security problems are restrained, and the supervision is implemented throughout the data exchange process.

(4) Introduce the concept of data security exchange network to constrain the data security exchange network connection behavior of the special exchange process.

3. Multi-dimensional process behavior evaluation model

The multi-dimensional process behavior evaluation model can improve the evaluation performance based on the one-dimensional process behavior evaluation model. Based on the process behavior evaluation model based on the hidden Markova model, the multi-dimensional process process behavior evaluation model integrates the evaluation results with Boolean operation to obtain better evaluation performance [10].

3.1. Process behavior evaluation model based on HMM

Hidden Markov Model (HMM) is a dual random process, in which one random process describes the transition of the state, and another describes the state that emits observable statistical probability. As shown in figure 2, the hidden Markov model is used to describe the process behavior.

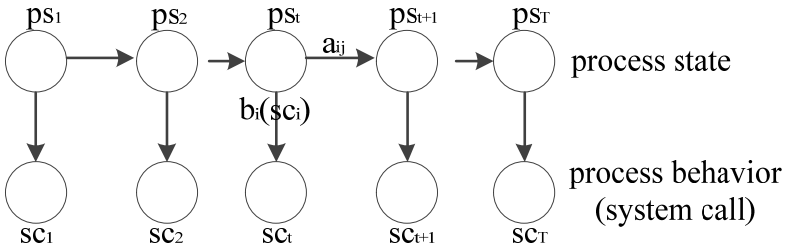

Figure 2. Process behavior evaluation model based on HMM

In figure 2, PST and sc_t represent the state and behavior respectively. The process behavior evaluation model based on the hidden markov model can be expressed as $\lambda=(\pi,A,B)$, represented as:

(1)The process state number is N, process state set $\Theta=\{\theta_1,\theta_2,...,\theta_N\}$;

(2)The number of system calls used by the process during running is M. Each different system call represents a process behavior, and gets the process behavior set $\Delta=\{\delta_1,\delta_2,...,\delta_M\}$;

(3)The initial state probability distribution vector $\pi=\{\pi_i\}$, π_i @ $p(ps_1=\theta_i)_{1\le i\le N}$, where ps_1 is the initial state;

(4)The state transition probability matrix is $A=\{a_{ij}\}$, a_{ij} @ $p(sc_{t+1}=\theta_j \mid sc_t=\theta_i)_{1\le i,j\le N}$, and it represents the probability of switching from state θ_i to state δ_j;

(5)The state output probability matrix is $B=\{b_i(k)\}$, $b_i(\delta_j)$@ $p(sc_t=\delta_j \mid q_t=\theta_i)_{1\le j\le N}$, and it represents the probability of generating behavior δ_j in process state θ_i.

The parameters of the hidden Markov model can be obtained by using the Baum-Welch algorithm. The algorithm can obtain $\lambda=(\pi,A,B)$, and maximum $p(SC \mid \lambda)$, and then construct the hidden Markov model $\lambda=(\pi,A,B)$.

In the evaluation, the forward recursion method is used for SC of system call sequence to obtain its occurrence under hidden Markov model λ The higher the probability of $p(SC \mid \lambda)$.The larger the value of p, the more normal the event sequence is, and conversely, the more likely the sequence is to be abnormal. When $p(SC \mid \lambda)\ge t$, the process behavior is considered normal, while $p(SC \mid \lambda)<t$ the process behavior is considered abnormal, and t is the set threshold.

Figure 3 shows the multi-dimensional process behavior evaluation model. The process behavior data is obtained from the training library. According to the characteristics of the hidden Markov model, multiple different state Numbers are used to train the process behavior data. Thus, multiple different process behavior evaluation models are obtained.

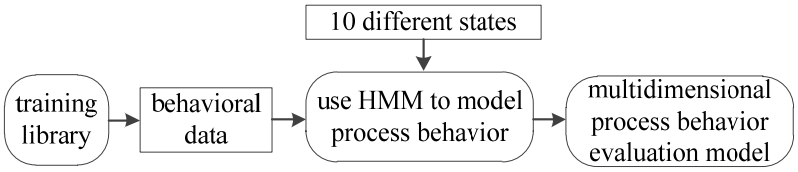

Figure 3. Multidimensional process behavior evaluation model

3.2. Boolean operation based multidimensional process behavior evaluation model fusion algorithm

Lippmann et al. adopted the technology of Receiver Operating Characteristics (ROC) curve to train the process behavior evaluation model through a large amount of data, and changed the evaluation threshold to obtain parameters such as TP, FP, TN and FN under different thresholds, which were used to analyze the relationship between the phase influence of two variables, namely, FPR of false alarm rate and TPR of accuracy. Take the accuracy rate tpr and the false alarm rate fpr as coordinate pairs (FPR, TPR), drawing them on the plane of fpr- tpr, and connect the dots on the plane to form ROC curve. The larger the area under the ROC curve (AUC, Area under Curve) of the process behavior evaluation model, the better the performance of this evaluation model.

$$tpr = \frac{TP}{TP + FN} \tag{1}$$

$$fpr = \frac{FP}{FP + TN} \tag{2}$$

The evaluation model fusion algorithm based on Boolean operation classifies the process behavior evaluation into two aspects: abnormal p (positive) or normal n (negative). The two trained models of process behavior evaluation are referred to as λ_1 and λ_2. For the same behavior trajectory, the judgment results include the following four situations, as shown in table 1 The evaluation results of the two processes were performed with Boolean operation $\lambda_1 \wedge \lambda_2$, $\neg\lambda_1 \wedge \lambda_2$, $\lambda_1 \wedge \neg\lambda_2$, $\neg(\lambda_1 \wedge \lambda_2)$, $\lambda_1 \vee \lambda_2$, $\neg\lambda_1 \vee \lambda_2$, $\lambda_1 \vee \neg\lambda_2$, $\neg(\lambda_1 \vee \lambda_2)$, $\lambda_1 \oplus \lambda_2$, $\neg(\lambda_1 \oplus \lambda_2)$, and then the results were compared with the real abnormal situation, and then the TP, FP, TN, FN, FPR and TPR were obtained after Boolean operation. TP is correct hit, FP is false report, TN is correct neglect, and FN is false report. The hidden Markov model and Boolean algorithm are applied to the evaluation of multi-dimensional data safety exchange.

Table 1. Evaluation results of two behaviors

λ_1	p	p	n	n
λ_2	p	n	p	n

4. Optimal process behavior evaluation method

When using the behavior evaluation model to evaluate process behavior, instead of using the entire ROC curve, it selects a point on it, that is, a process behavior evaluation model, whose accuracy and false alarm rate must meet the needs of practical application. An optimal process behavior evaluation model selection method based on cost decision tree is proposed to coordinate the contradiction between model accuracy and false alarm rate.

Anomaly represents abnormal process behavior, and Normal represents Normal process behavior, and Positive represents abnormal result of behavior evaluation model, and Negative represents Normal result of behavior evaluation model.

The trained process behavior evaluation model was verified on the tested data set to obtain TP, FP, TN, and FN. The accuracy rate of TPR and false alarm rate of FPR were calculated.

The process behavior evaluation model minimizes the loss caused by false and omission, so as to obtain the best behavior evaluation model. Taking process behavior as the root node, a cost decision tree of process behavior evaluation model can be obtained according to the result of behavior evaluation, the response of the result and the actual anomaly of the behavior.

Assume that α is the rate of false positives, β is non-response rates, C_α and C_β respectively false positives and omission of costs, p is the incidence of the abnormal process behaviors, p_1 is the process behavior assessment probability model did not detect the abnormal behavior, when the behavior is normal, the probability of p_2 is for normal process behavior, when the process behavior, the probability of p_3 is for normal process behavior.

Combined with the cost decision tree, according to the total probability formula, the following equation can be obtained:

$$p_1 = P(Negative) = (1-\alpha)(1-p) + \beta p \tag{3}$$

$$1 - p_1 = P(Positive) = \alpha(1-p) + (1-\beta)p \tag{4}$$

If the behavior is normal, the next step can be deduced:

$$\begin{aligned} p_2 &= P(Normal \mid Negative) = (1-\alpha)(1-p) / p_1 \\ &= (1-\alpha)(1-p) / [(1-\alpha)(1-p) + \beta p] \\ 1 - p_2 &= P(Anomaly \mid Negative) = \beta p / [(1-\alpha)(1-p) + \beta p] \\ &\bullet \\ &\bullet \\ &\bullet \end{aligned} \tag{5}$$

When the system chooses not to respond, the cost is:

$$C_{No\,\mathrm{Re}sponse} = Cp_1(1-p_2) + C(1-p_1)(1-p_3) \tag{6}$$

The cost when the system chooses to respond is:

$$C_{\mathrm{Re}sponse} = p_1 p_2 + (1-p_2)p_3 \tag{7}$$

The system obtains the evaluation cost of ROC curve $(1-\beta, \alpha)$ in the way of least cost:

$$C_{evl} = (\alpha, 1-\beta) = Min\{C\beta p, (1-\alpha)(1-p)\} + Min\{C(1-\beta)p, \alpha(1-p)\} \tag{8}$$

By comparing the evaluation cost C_{evl} of each point on the ROC curve, the point with the lowest evaluation cost is regarded as the best advantage on the ROC curve. The objective function of the optimal process behavior evaluation model and its constraint conditions are:

$$\begin{cases} C_{\exp} = Min\{C\beta p, (1-\alpha)(1-p)\} + Min\{C(1-\beta)p, \alpha(1-p)\} \\ f_{optimal}(\alpha, 1\text{-}\beta) = 0, \quad \alpha \in [0,1] \end{cases} \tag{9}$$

5. Experimental simulation

5.1. Experimental comparison

In order to verify the effectiveness of the evaluation model of multi-dimensional process behavior based on Boolean operation, the evaluation model is compared by experimental simulation. The simulation environment is Core i5 6500, 8 GB of memory, virtual operating system Ubuntu12, and kernel version linux2.6. The Forrest data was collected by the University of New Mexico in the United States. The fusion method proposed in this paper is compared with the evaluation model of STIDE and HMM. The normal data of the send mail process in forest data are used. The experimental comparison figure is shown in figure 4.

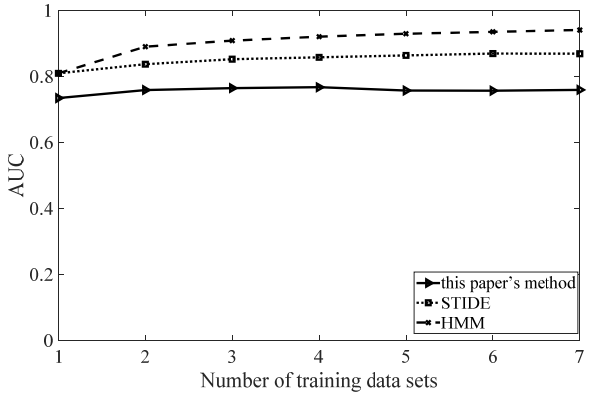

Figure 4. Performance comparison of HMM and STIDE with the method in this paper

The experimental data of "internet-based e-government" is selected as the behavior evaluation model $\lambda_{exchange}$. The leaking cost of the top-secret exchange task is $C_{TopSecret}$, the leaking cost of the secret exchange is $C_{Classified}$, the leaking cost of the secret exchange task is C_{Secret}, and the leaking cost of the public exchange task is C_{Public}. Their evaluation costs are obtained and their corresponding optimal models are selected.

Table 2. The performance evaluation table of $\lambda_{exchange}$

fpr	0	0.01	0.1	0.13	0.2	0.3	0.32	1
tpr	0	0.34	0.67	0.75	0.85	0.90	1	1

Table 3. Exchange the evaluation cost calculation results of tasks at each evaluation point

fpr	0.01	0.1	0.13	0.2	0.3	0.32
tpr	0.34	0.67	0.75	0.85	0.90	1
$C_{TopSecret}$=10	0.7	0.7	0.7	0.59	0.51	0.7
$C_{Classified}$=5	0.7	0.565	0.466	0.365	0.36	0.7
C_{Secret}=3	0.601	0.304	0.316	0.275	0.3	0.7
C_{Public}=1	0.205	0.169	0.116	0.185	0.24	0.3

5.2. Analysis of results

By testing three methods, it can be found that the performance of this algorithm is consistent with HMM when training group 1 data. After adding data from groups 2 to 7, the training performance of the HMM method is inferior to that proposed in this paper. Moreover, the training performance of STIDE method is always inferior to the method mentioned in this paper. It is demonstrated that the

method presented in this paper can select the best-performing behavior evaluation model among multiple process behavior evaluation models, so the performance can be substantially improved.

As can be seen from table 3, the top secret level switching task and the secret level switching task obtain the minimum evaluation cost in (0.3,0.9), the secret level task obtain the minimum evaluation cost in (0.2,0.85), and the open task obtain the minimum evaluation cost in (0.13,0.75). The evaluation results are affected by the underreporting costs based on security requirements and efficiency requirements.

6. Summarizes

In order to solve the problem of single process behavior evaluation model, the multi-dimensional process behavior evaluation model is integrated by Boolean operation, and a multi-dimensional process behavior evaluation model is proposed.

The evaluation model presented in this paper is compared with the traditional STIDE and HMM methods. Aiming at the problem of model selection in process behavior evaluation model, the relationship between accuracy and false alarm rate is discussed, and the evaluation indexes of process behavior evaluation model are studied. Based on the cost decision tree theory, the objective function of selecting the optimal process behavior evaluation model is given, and the optimal process behavior evaluation model is selected on the multi-dimensional process behavior evaluation model. The effectiveness and practicability of the method are illustrated by an application example.

References

[1] Huw Read, Konstantinos Xynos,Andrew Blyth. Presenting DEViSE: Data Exchange for Visualizing Security Events [J]. IEEE Computer Graphics and Applications, 2009, 29 (3): 6-11.

[2] HE Zhangqing, LI Hong, WAN Meilin, et al. Authentication and session key exchange protocol based on Physical Uncolonable Function [J]. Computer engineering and applications, 2018 (18): 17-21.

[3] Dong GuiShan, Liu ZhengJun, Zhao Dong. A security domain isolation and data exchange system based on VMM [C]. 2009 3rd International Conference on Signal Processing and Communication Systems, 28-30 Sept. 2009, Omaha, NE, USA.

[4] Iko Miyazawa, Masashi Murakami, Takashi Matsukuma, et al. OPC UA information model, data exchange, safety and security for IEC 61131–3 [C]. SICE Annual Conference 2011, 13-18 Sept. 2011, Tokyo, Japan.

[5] Luo Yang, Zheng Chang-xing. Data Exchange between Multilevel Security Management System Based Ice [C]. 2010 International Conference on Multimedia Information Networking and Security, China, 4-6 Nov. 2010, Nanjing, Jiangsu.

[6] Zhang zhichang, Chen jingsheng, li bin, tian shiming, dong mingyu, zhu weiyi, qi bing, sun yi. Standardization status and development trend of China's demand response information exchange [J]. Grid technology, 2008, 42 (04): 1183-1190.

[7] Guo renchao, xu yutao. Research on the application of internal and external network data security exchange technology in power grid enterprises [J]. Power big data, 2008, 21 (02): 61-66.

[8] Wu yiqi, he fa zhi, li xiaoxia, CAI xian tao. Service-oriented cloud design security data exchange [J]. Journal of huazhong university of science and technology (natural science edition), 2016, 4 4 (12): 76-80+97.

[9] Yang changchun, jiang dongdong, lu hexin. Data incremental exchange model based on multi-agent collaboration [J]. Computer application and software, 2007, 34 (01): 85-89+105.

[10] Yang xiaodong, gao guo-juan, zhou qixu. E-government data security exchange scheme based on proxy re-signature [J]. Computer engineering, 2007, 43 (02): 183-188.