Influence of integrity control tools parameters on the metrological data unauthorized modification threat probability

A.I Frid
Department of Computer
Science and Robotics
Ufa State Aviation Technical University
Ufa, Russia
e-mail: arkfrid@mail.ru

T. I. Fazliakhmetov
Department of Computer
Science and Robotics
Ufa State Aviation Technical University
Ufa, Russia
e-mail: sidh@mail.ru

Abstract[1]

Assessing the influence of integrity control tools on probability of metrological data unauthorized modification threat is discussed in the paper. Familiarity with time required to overcome the integrity control tool permits to estimate a probability of integrity control tool overcoming. It is shown how artificial neural network parameters can influence on this time.. It is calculated a risk before and after this method introduction and its effectiveness.

1. Introduction

Currently more attention is paid to the protection of metrological data (MD) from unauthorized modification (UAM) and their complete removal because decisions on management of production processes are taken based on these data. The consequences of these decisions can be disastrous.

The purpose of research presented in the paper is to increase the security level of MD. One of the metrological information protection problems is considered - to ensure its integrity. For achieving this goal it is necessary to solve the following tasks:

1. Offer an integrity control method of the MD.
2. Evaluating the effectiveness of the proposed method.

The conception and the basic algorithm of solving the first problem are described in [1, 2, 3]. The basic idea is to use the semantic meaning of information - the context of information flows [4]. In context of information security of MD this means that MD is functionally related with each other. For example, the flow rate in a pipeline depends on pressure on the input and output of the pipeline, viscosity, temperature and other parameters. Solution of the data protection in this case is to calculate the flow rate value by the other flow parameters and compare calculated value with the real value. If the calculated and real values are equal or close it is decided that there is not UAM. If these values are not equal it is decided that there is an UAM.The decision of the second problem is proposed in [2], where a model for evaluating the effectiveness of the MD integrity controls in production systems is offered. However, this model does not define how to calculate the total probability of the MD UAM threat and how integrity controls affects to this probability. Therefore, the task of assessing the influence of any integrity control mechanisms on probability of MD UAM threat is set in this paper. Also the task of evaluating the effectiveness of the proposed method in accordance with this influence is set.

2. Calculation of integrity control mechanisms overcoming probability

Model for circulating of complex threats assessment to information on the subject of informatization is proposed in [5]. By this method the set of threats is considered as the set of unauthorized access channels, information leakage channels and channels of destructive effects on the informational environment (UAALE). Special model is constructed for each type of channel in the view of graphs structuring UAALE channels. This model is taking into account the threats interaction with the information security means. By this model the threat implement probability is equal to multiplication of the probabilities of success in barriers overcoming by the attacker (intruder). We will apply a similar method for the construction the model of MD UAM threat. The general view of this model is shown in Figure 1.

In Fig. 1 IST - information security tool, P_i - probability of overcoming the appropriate information protecting tool. This model shows the attacker is needed to overcome n information protecting tools to get full access to the MD.

[1] **Proceedings of the 14th international workshop on computer science and information technologies CSIT'2012, Ufa – Hamburg – Norwegian Fjords, 2012**

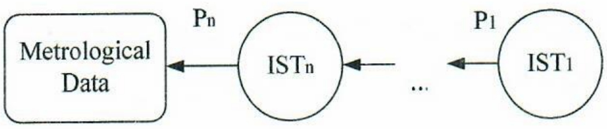

Fig. 1. Model of MD UAM threat

The probability of MD UAM in this case is calculated as:

$$P_{total} = \prod_{i=1}^{n} P_i \,. \qquad (1)$$

Introduce the integrity control tool to protection system and combine all the protection tools besides the integrity control tool to the single node. Then the model takes the format shown in Figure 2.

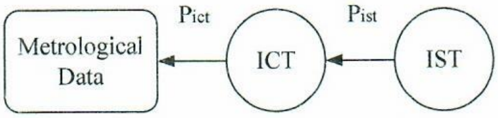

Fig. 2. Model of MD UAM threat with integrity control mechanism

In Fig. 2 ICT - integrity control tool, P_{ict} - the probability of the integrity control tool overcoming, P_{ist} - the probability of all information protection tools overcoming (besides the integrity control tool). Then the MD UAM probability in this case is calculated as:

$$P_{total} = P_{ict} \cdot P_{ist} \,. \qquad (2)$$

Thus, by (2) we can conclude that the task of assessing the influence of some integrity control tool on probability of MD UAM threat reduces to determination of this integrity control tool overcoming probability.

To solve this problem following formula is proposed to calculate the probability:

$$P = f(t_{ict}) \,, \qquad (3)$$

where t_{ist} - the time that is needed to overcome the information protection tool that implements integrity control. For example, if the integrity control mechanism is implemented using a hash function, it may be the time required to recalculate hashes of modified MD. For the integrity control method previously proposed in [1, 2, 3] based on the artificial neural network (ANN), it is the time required to retrain ANN by new modified data. Now the challenge is to define a specific function f used to calculate MD UAM threat probability. Function f should to satisfy to the following constraints:

1. If t_{ict} increases then P decreases.
2. $0<P<1$.
3. If $t_{ict} \to -\infty$ then $P \to 1$.

Proceeding to the listed constraints the following function is offered to use as the f:

$$P = e^{-\lambda t_{ict}} \,, \qquad (4)$$

where λ – coefficient characterizing the attacker qualification. In addition to above described constraints while using the formula (4) it is important to note that the t_{ist} range is from 0 to 8 hours because the attacker has not more than 8 hours. This limit is determined by the duration of work day at the company. Thus, a plot of the MD UAM threat probability P on the time required to overcome the integrity control tool t_{ict} is shown in Figure 3.

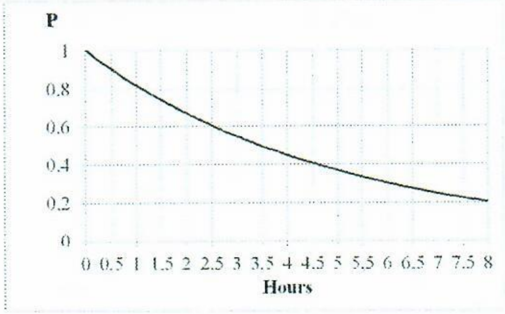

Fig. 3. Plot *P* on t_{ict}

It is a problem to choose coefficient λ to use it the formula (4). This coefficient is proposed to select from the following considerations. Consider there are 3 attacker's skill levels: low, medium and high. Assume that for attacker with medium skill level when t_{ict} is equal to 4 hours the probability of the integrity control overcoming is equal to 0,5 . This assumption can be expressed by the equation:

$$0{,}5 = e^{-4\lambda} \,. \qquad (5)$$

When solve this equation for λ we find that λ = 0,17. When round λ to one decimal place we obtain λ = 0,2. Thus, for attacker with medium skill level we assign λ = 0,2.

Next we need to define λ to attackers with high and low skill level. Figure 4 shows plots for different λ.

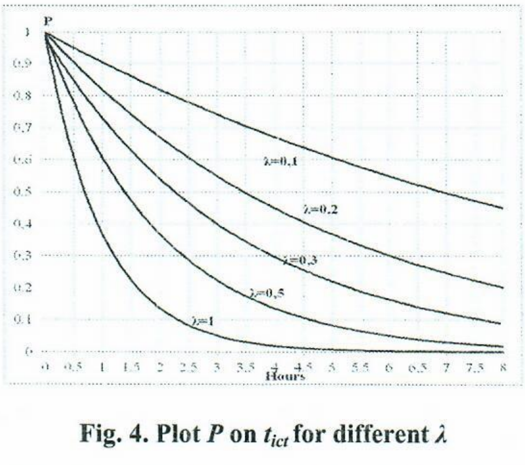

Fig. 4. Plot *P* on t_{ict} for different λ

It is true that the attacker with higher qualification has got greater probability to overcome integrity control at the same t_{ict}. Therefore the higher plot matches to the higher attacker's qualification. Fig. 4 shows that the higher plot matches to the smaller λ. Consequently the higher qualification matches to the lower coefficient λ. In this paper λ = 0,1 is proposed for an attacker with high skills, and λ = 0,5 is proposed for an attacker with low skills. Selected coefficients λ for attackers with different skills are listed in Table 1.

Table 1

The λ for attackers with different skills level

Attacker skills level	λ
Low	0,5
Medium	0,2
High	0,1

3. Evaluation of proposed integrity control method effectiveness

In order to evaluate the effectiveness of proposed solution it is necessary to assess the risk before and after this solution introduction. Assume that the initial protection system uses hash function as the integrity control tool. Then the MD UAM threat model can be represented as follows:

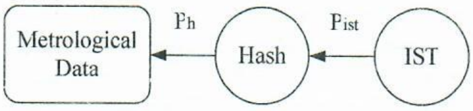

Fig. 5. Model of MD UAM threat with hash function

In Fig. 5 P_h - the probability of hash overcoming. In [2] the MD UAM risk is proposed to estimate using the following equation:

$$R = I \cdot K \cdot F \cdot P \quad (6)$$

Symbols in the formula (6) indicate following:

- R – risk level;
- I – information cost;
- K – destructiveness coefficient;
- P – threat probability;
- F – threat frequency.

Denote the total probability of the MD UAM threat before the proposed method introduction as P_{tb}. By figure 5 and equation (2) P_{tb} is calculated as follows:

$$P_{tb} = P_h \cdot P_{ist} \quad (7)$$

Then MD UAM risk before the proposed method introduction:

$$R_b = I \cdot K \cdot F \cdot P_{tb} = I \cdot K \cdot F \cdot P_h \cdot P_{ist} . \quad (8)$$

When added to the protection system method based on artificial neural networks (ANN) proposed in [1, 2, 3] the model takes the view shown in Figure 6.

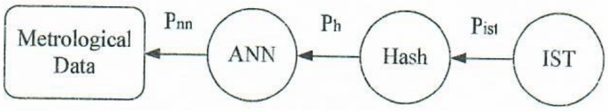

Fig. 6. Model of MD UAM threat with hash function and ANN

In Fig. 6 P_{nn} - the probability of the method based on ANN overcoming. Denote the total probability of the MD UAM threat after the proposed method introduction as P_{ta}. By figure 6 and equation (2) P_{tb} is calculated as follows:

$$P_{ta} = P_h \cdot P_{nn} \cdot P_{ist} . \quad (9)$$

Then MD UAM risk after the proposed method introduction:

$$R_a = I \cdot K \cdot F \cdot P_{ta} = I \cdot K \cdot F \cdot P_h \cdot P_{nn} \cdot P_{ist} . \quad (10)$$

The effectiveness of the proposed integrity tool E_f can be calculated as follows:

$$E_f = \frac{R_b - R_a}{R_b} = 1 - \frac{R_a}{R_b} . \quad (11)$$

Using formulas (8) and (10) we get the following term to determine the effectiveness of the proposed solutions:

$$E_f = 1 - P_{nn} . \quad (12)$$

Denote the total time required for training the ANN as a t_{nn}, and substitute the formula (4) to (12) we get the effectiveness of proposed solution that can be defined as:

$$E_f = 1 - e^{-\lambda t_{nn}} . \quad (13)$$

Thus, it can be concluded that the effectiveness of offered method depends on the time required for training the ANN. And the longer this time is the more effectiveness. is the method. Above the λ coefficients were determined for the three attacker skills levels. Plots of E_f on t_{nn} for these levels are shown in Figure 7. Figure 7 shows that the greater λ corresponds to the higher plot of E_f on t_{nn}. That is, when attacker has lower qualification the proposed method effectiveness is larger. For example, when training time t_{nn} = 1 hour and λ = 0,2 the proposed method effectiveness is 0.19. In other words the MD UAM risk reduces to 19%. Based on formula (13) we can conclude that to increase the proposed method effectiveness it is necessary to increase the training time of ANN. In [1, 2, 3] multilayer perceptron is proposed as the ANN. Training time of such ANN type depends on the following parameters:

- The structure of the ANN (hidden layers number and neurons number in them);
- Learning rule and its parameters;
- Patterns count in the training set;
- Training error.

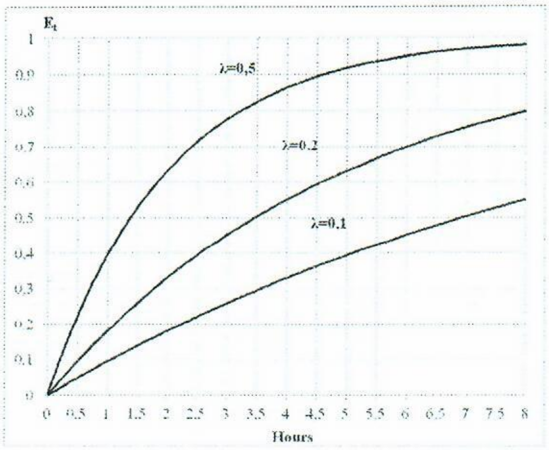

Fig. 7. Plot E_f on t_{nn} for different λ

In this paper ANN structure and training error are proposed as the parameters that influence on training time. Thus, the problem of estimating the influence of the ANN structure and the training error on training time (consequently, on proposed solution effectiveness) and the problem of choosing the ANN structure are arisen.

4. The ANN structure and training error influence on proposed method effectiveness

The influence of the ANN structure with one hidden layer and different neurons counts C_{nn} in it on training time is analyzed in problem described.

The results of ANN training time evaluation depending on neurons count in the hidden layer and training error ΔE_{nn} are shown in the Table 2. Training time is measured in hours.

Table 2

Training time depending on neurons count in the hidden layer and training error

C_{nn} \ ΔE_{nn}	10E-5	10E-6	10E-7
4	0,008	0,123	0,306
6	0,012	0,150	0,404
10	0,015	0,190	0,557
16	0,018	0,271	0,899
20	0,022	0,344	1,238
30	0,032	0,619	2,756
40	0,048	1,117	6,133

The results of effectiveness evaluation depending on neurons count in the hidden layer and training error are shown in the Table 3. Effectiveness is shown in %.

During experiments it was found that if the training error is less than 10E-6 or the number of neurons in the middle layer is more than 40 the ANN trains unstable. Therefore, according to Table 3 we can conclude that maximum efficiency of the proposed method 20% is achieved when there are 40 neurons in the middle layer and the training error is equal to 10E-6.

Table 3

Effectiveness depending on neurons count in the hidden layer and training error

C_{nn} \ ΔE_{nn}	10E-5	10E-6	10E-7
4	0,16	2,5	6,3
6	0,25	2,98	7,76
10	0,30	3,77	10,53
16	0,37	5,36	16,46
20	0,43	6,76	21,94
30	0,64	11,99	42,37
40	0,96	20,26	70,67

5. Comparative assessment of MD UAM risk before and after proposed method introduction

To evaluate the proposed solution effectiveness it is necessary to give an example of information security risks calculation before and after this solution introduction. Let consider pipeline with oil flow as technological object and oil flow rate as parameter in which an attacker wants to modify metrological data. Assume there is corporative information system automating oil account processes. Information system includes workstation with installed client software for processing metrological information. Information system has two-tier architecture. Metrological data are stored on the Oracle Database 10. For each metrological value hash function is calculated by the SHA-256 algorithm. Hash values are checked when metrological values are processed in the application. The hashes are also stored in the database. Users must make authentication by login and password to get access to application. There is no possibility to modify data directly came from the automated measurement systems, i.e. metrological data remains in the database unchanged. However, the login and password used for authentication in the application can also be used to authenticate directly to the database (for example, by using utilities to work with the database).

Assume that an attacker with middle skills level (λ = 0,2) has the following access and rights:

- Access to the corporate network;
- Access to the operating system;
- Access to the application;
- Rights to modify data including the hashes.

Suppose the attacker wants to reduce the flow rate to 10%. For some MD set company loss for one day in this case is 1238 tons. Suppose the one oil ton cost is 70$. Then we can calculate the annual loss of the company:

$$I_{sut} = 1238 \cdot 70 \cdot 365 = 31630900\,\$. \qquad (14)$$

According to described system and the attacker model we can say that for the MD UAM implementation attacker must overcome only the integrity control tool. Then MD UAM threat model shown in Figure 5 takes the following form:

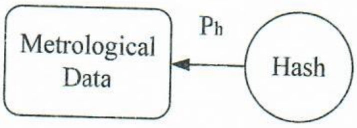

Fig. 8. Model of MD UAM threat before proposed solution introduction

Consequently, the threat probability will be equal to the probability of overcoming the integrity control tool P_h. Experiments result that the time required to recalculate 1440 hashes and to save them in a database is about 144 seconds or 0.04 hours. Then P_h, calculated by the (4):

$$P_h = e^{-0,2 \cdot 0,04} = 0,992 . \qquad (15)$$

The coefficient K in this case is equal to 1. Assume the attacker intends to implement MD UAM once a month. This assumption is based on the fact that the reporting frequency in the company is one by month and after reporting MD loses their relevance. In this case the value of F is equal to 0.033. By formula (8) we can calculate the risk before the proposed method introduction:

$$R_b = 31630900 \cdot 0,992 \cdot 0,033 \cdot 1 = 1035469\$. \qquad (16)$$

Let's calculate the risk after proposed solution introduction. After proposed method based on ANN is added to the system the model shown in Figure 6 takes the view like in Figure 9.

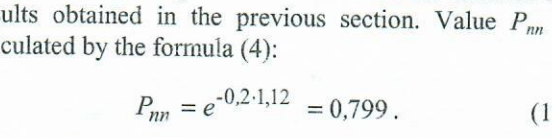

Fig. 9. Model of MD UAM threat after proposed solution introduction

The time required for ANN retraining and saving it in the database we take equal to 1.12 hours on the basis of the results obtained in the previous section. Value P_{nn} is calculated by the formula (4):

$$P_{nn} = e^{-0,2 \cdot 1,12} = 0,799 . \qquad (17)$$

The risk for this case (by the formula (10)):

$$R_a = 31630900 \cdot 0,992 \cdot 0,799 \cdot 0,033 \cdot 1 = 827339\$. \; (18)$$

Then we can calculate reduction of the risk:

$$\Delta R = R_b\text{-}R_a = 1035469 - 827339 = 208130\$. \; (19)$$

The efficiency is (by the formula (11)):

$$E_f = \frac{208130}{1035469} = 0,20 . \qquad (20)$$

6. Conclusion

- There are shown the ways how the integrity control tools influence on the MD UAM total probability.
- The probability of integrity control tool overcoming in proposed method depends on the time required to overcome this tool.
- So, the effectiveness of previously proposed method depends on the time required to overcome it. And this time depends on the ANN structure and training error. It is shown that use of the ANN with 40 neurons in the middle layer and training error about 10E-6 allows getting the best results.
- As an example it is cited MD UAM risk calcula-tion before and after introduction of integrity control method based on ANN. This example shows that the method allows to reduce MD UAM risk to 20% or approximately 210,000 $.

References

1. Fazliakhmetov T.I. "The method of rapid diagnosis of flowmeter condition and calculation of its calibration coefficients using artificial neural networks". *Automation, telemechanization and communications in the oil industry. STJ.* 2009; 4:58-61.
2. Fazliakhmetov T.I., Frid A.I. "The method of effectiveness estimation of metrological data integrity providing in gas-oil production systems". *Problems of information protection: SIJ.* 2011; 2:31-36.
3. Frid A. I., Fazliakhmetov T. I. "Metrological information integrity providing algorithm in industrial automated systems". In: *Proc. of 13th International Workshop on Computer Science and Information Technologies (CSIT'20011)*, Vol. 1. Garmisch – Partenkirchen, Germany, 2011, pp. 91-94.
4. "Protecting Essential Information: Securing the Foundation of the Internet Business Platform", Websence, USA, 2008.
5. Mashkina I.V., Rakhimov E.A., Vasilyev V.I. "The method of model design for complex estimation of threat to information circulating in the object of informatization". *Informational opposition to terrorism threats: South Federal University.* 2006; 7: 269-280.