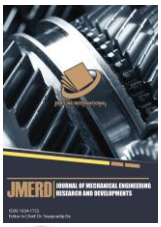
Journal of Mechanical Engineering Research & Developments (JMERD)

ISSN: 1024-1752 (Print)

CODEN: JERDFO

***DOI:** http://doi.org/10.26480/jmerd.02.2019.06.09*



RESEARCH ARTICLE

ALGORITHM DETERMINE TRUST VALUE TO THE DISTRIBUTED INFORMATION SYSTEMS ELEMENTS

Aleksei A. Sychugov[1], Elvir M. Akhmetshin[2*], Vyacheslav M. Grishin[3], Raisa N. Shpakova[4], Andrei V. Plotnikov[5]

[1]Department of Information Security, Tula State University, 92 Lenin Ave., Tula, Russian Federation
[2]Department of Economy and Management, Elabuga Institute of Kazan Federal University, 89 Kazanskaya Str., Elabuga, Russian Federation
[3]Department No.604 "System Analysis and Management", Moscow Aviation Institute, 4 Volokolamskoe shosse, Moscow, Russian Federation
[4]Department of Regional Governance and National Policy, Moscow State Institute of International Relations (MGIMO), 76 Vernadsky Ave., Moscow, Russian Federation
[5]Department of Management, Perm State Agro-Technological University named after Academician D.N. Pryanishnikov, 23 Petropavlovskaya Str., Perm, Russian Federation
[*]Corresponding Author Email: elvir@mail.ru

ARTICLE DETAILS

Article History:

Received 01 February 2019
Accepted 05 March 2019
Available online 08 March 2019

ABSTRACT

In distributed information systems the owner of the information can make a decision only concerning the transfer of information to a particular DIS node for calculations and cannot control the security of information during its processing on the node. Thus, in this case there is a question of information owner's trust to a specific DIS node. The authors describe an experiment conducted in order to verify the quality of approach, which allows evaluating the level of trust based on the featureless pattern recognition. The authors found that this algorithm will allow to prevent the transfer of confidential information on unreliable DIS elements, also it will require more careful attention to be paid to the protection of separate DIS components.

KEYWORDS

Trust, security, featureless pattern recognition, prediction error, information security

1. INTRODUCTION

Nowadays distributed computing technologies are actively developing due to the increasing volume of information to be processed in a short time. In addition to productivity the distributed information systems (DIS) have such advantages as fault tolerance and scalability, achieved by dividing the information system into separate elements with their own software and hardware infrastructure and specific functions [1,2]. This approach allows distributing computational resources optimally and changing the system configuration quickly during the peak loads. As the popularity of the DIS grows the amount of processed information increases and, consequently, the problem of DIS information security becomes essential.

A DIS is defined by the following basic components:

- the functional component providing the solution of tasks;
- the hardware-software platform for fulfilling computing processes;
- the network fulfilling interactions of platforms;
- the information resources;
- the security policy requirements of the secure handling of information resources.

In the DIS the outer zone is blurry, because each element of the distributed systems has its own security zone. Thus, information's owner fails to control and monitor the security features [3]. Today developing methods for assessing DIS elements in terms of security and preserving the confidentiality of the security tools data and DIS element state is the urgent task.

In distributed information systems the owner of the information can make a decision only concerning the transfer of information to a particular DIS node for calculations and cannot control the security of information during its processing on the node. Thus, in this case there is a question of information owner's trust to a specific DIS node [4]. The object of the study was an approach which allows evaluating the level of trust based on the featureless pattern recognition. The aim of the study was to describe an experiment conducted in order to verify the quality of this approach.

2. METHODOLOGY

According to the general definition, trust is a level of subjective probability by which A expects that B performs an action that A can not watch and control and that affects the welfare and benefits of A [5]. In the DIS the concept of the trust may be defined as the probability that the data, transmitted to the DIS node and calculation results will not be discredited or falsified, in other words, their confidentiality, integrity and availability will not be violated during the processing [6,7].

The model of trust in the DIS can be described as follows:

Ω – the set of all nodes of the DIS:

$$\Omega = \{\Omega_0, \Omega_1, \ldots, \Omega_N\}, \qquad (1)$$

Cite The Article: Aleksei A. Sychugov, Elvir M. Akhmetshin, Vyacheslav M. Grishin, Raisa N. Shpakova, Andrei V. Plotnikov (2019). Algorithm Determine Trust Value To The Distributed Information Systems Elements. *Journal of Mechanical Engineering Research & Developments*, 42(2): 06-09.

where *N* – number of elements in the DIS. Every Ω_i at time $t_j \in T$ is in state $s_i^j \in S$, where S – a set of all possible states of DIS element, which determine trust value to the node at the current time and are described by some feature set with metric properties. For each Ω_i the value of trust p_i at time $t_j \in T$ is defined by a set of system conditions:

$$p_i = P(s_i^0, s_i^1, \ldots, s_i^j), \tag{2}$$

where p_i – trust to DIS element in the time period $t = [0 \ldots j]$. Therefore, it is possible to determine the probability of data corruption during processing on the node Ω_i as $q_i = 1 - p_i$.

It can be stated that the trust value depends on the state of the DIS element during observation. To analyze the state of the DIS element, without violating the confidentiality of information about it, it is proposed to analyze the measure of its similarity to a predetermined basis, specifying the state of the element with a known trust value, instead of studying the set of property values that describe the state of the element.

The set of basic objects:

$$B = \{b_0, b_1, \ldots, b_M\}, \tag{3}$$

where $b_i \in S$; M – the number of basic elements.

The function of DIS elements' state similarity:

$$r_{i,j} = \rho(s_i^t, s_j^t), \tag{4}$$

where ρ – function for finding the metrics in relation to the set S.

Then the state of the element Ω_i at time t can be described by the vector:

$$\grave{s}_i^t = \{\rho(s_i^t, b_0), \rho(s_i^t, b_1), \ldots \rho(s_i^t, b_M)\}. \tag{5}$$

Let us substitute (2) into (**Error! Reference source not found.**) and get

$$p_i = P(\grave{s}_i^0, \grave{s}_i^1, \ldots, \grave{s}_i^K). \tag{6}$$

With the proposed approach we can use a mathematical tool, known in machine learning as featureless pattern recognition in which we use projection space samplings based on projection features, represented by similarities to some predefined (space-forming and basic) objects instead of the linear vector space of object properties [8]. In other words, when using featureless pattern recognition for each object of the initial space, the similarity function (a function of distance) is determined. Then a set of basic objects is introduced. Projection features (secondary properties), equal to the measure of object's similarity to the main ones, are calculated for each object. Then it is possible to use the known methods of classification, with the input data being the secondary features of the original objects.

The transition from the direct analysis of DIS elements' states to the analysis of similarity metrics solves the problem of DIS' zone security. If the state of the DIS element can be described with rational numbers, then if the basic objects and similarity metrics are captured the original features can be restored only under the condition that the number of basic objects exceeds the number of features by one. One of the advantages of this approach is that it allows the use of metrics to describe the state of the DIS element with the help of objects of any nature such as the sets or time series. In the latter case it is impossible to restore the original features. Algorithm for calculate the set of basic objects was proposed. This algorithm based on method FastMap.

Let $\Omega^0 \subset \Omega$ – subset for calculate of the set of basic objects, M - quantity of basic objects.

1. Select couple objects with maximum distance;

$$\bar{b}_0 = \langle b_{0,1}, b_{0,2} \rangle = \arg\max(\rho(\omega_i, \omega_j)), \omega_i, \omega_j \in \Omega^0 \tag{7}$$

2. Append this object in the set of basic objects;

$$\bar{B} = \bar{B} \cup \bar{b}_i,$$
$$B = B \cup \{b_{i,1}, b_{i,2}\}.$$

3. If $|B| = M$ then algorithm stopped, else step 4;

4. Search couple points;

$$\bar{b}_i = \underset{\omega_i, \omega_j \in \Omega^0, \notin B}{\arg\min} \left(\sum_{\bar{b}_k \in \bar{B}} proj(\omega_i, \bar{b}_k) - proj(\omega_j, \bar{b}_k) \right) \tag{8}$$

where function $proj$ calculate projection point ω_i on axis through points from set basics $\bar{b}_k$.

$$proj(\omega_i, \langle b_1, b_2 \rangle) = \frac{\rho^2(b_1, \omega_i) + \rho^2(b_1, b_2) - \rho^2(b_2, \omega_i)}{2 \cdot \rho(b_1, b_2)} \tag{9}$$

This algorithm calculates optimal basic set. Along with the advantages of this approach there is a question of choosing the trust functions P. The following principles are suggested to be used as the basis of trust function:

1. There is a time period for which the trust values are known;

2. The trust to the element with repeated and more predictable state is higher.

A numerical experiment was performed to test the quality of the proposed method. Its task was to detect the attack on network nodes maintaining the confidentiality of these nodes' operation. It is assumed that the attack on system element will change its behavior.

3. RESULTS AND DISCUSSION

For the initial experimental data, the following requirements were put forward:

1. The data must include information on the operation of the distributed information system, that is, on the operation of specific components of the system and their interactions.

2. The data must cover the time span of a few weeks. Since information systems have weekly periods, then for algorithm learning several samples of each day of the week are required.

3. The data must be marked. That is, the data should have marked points of attack's start and its end.

Having analyzed the open data, it was found that the requirements concern a small number of datasets [9]. Typically, data sets contain information about one node and the major part of information reflects the time periods of less than a week. This is due to the fact that, until recently, information security was aimed at protecting a single unit, and the systems, analyzing the behavior of the DIS as a whole existed only in theory.

For the experiment we collected data on the computer network of the Los Alamos National Laboratory according to the requirements [10]. The data is represented as the event logs, describing the work of a computer network for 58 days and includes four types of event logs: authorization log, network interaction log, DNS server log and computer processes log. Also, the data contains the activity log, that is, a so-called, red team - potential intruders attacking network computers. The red team log contains starting and ending points of attack, and victim computers' IDs. It should be noted that the statement of the problem for the experiment contains the assumption that the attack on the node changes the behavior of the system. However, there is no information on the type of attack and whether it was successful or not.

For the experiment, we took the data from the networking log, each record of which contains the following fields: time, duration of connection, network address, sender's and receiver's ports and packet size. As the best option between the amount of input data and the expected results, the time period of 3 weeks was selected to reduce the amount of data to be processed. During the first stage, the networking log was divided into several independent logs, each corresponding to a single computer log. Thus, we modeled a situation, when each element has its own security zone and networking information on each computer stays at that computer. Then, the information in each log was aggregated into blocks of data for ten minutes. As a result, DIS element state can be described as two vectors. The first vector $v_{i,1}^t$, with the size *N,* shows how many times the elements interact with each of system elements. The second vector $v_{i,2}^t$ shows which ports from the set of all possible system ports were used and how many times [11]. Thus, in this experiment, the state of the DIS element is determined by the following set of properties:

$$s_i^t = \{v_{i,1}^t, v_{i,2}^t\} \cdot \tag{10}$$

The following element similarity function of the elements has been selected:

$$\rho(s_i^t, s_j^t) = d(v_{i,1}^t, v_{j,1}^t) + d(v_{i,2}^t, v_{j,2}^t) + d(v_{i,3}^t, v_{j,3}^t) \tag{11}$$

where *d* – Euclidean metric.

The state of random DIS elements at the initial time was selected as the basis in the experiment. Then similarity states of DIS and basis elements were calculated. Trust function *P* is proposed to be based on the mean absolute percentage error (MAPE) forecasting time series:

$$P_i = 1 - \frac{1}{j} \sum_{t=0}^{j} \frac{\rho(s_i^t, b_0) - \rho`(s_i^t, b_0)}{\rho(s_i^t, b_0)}, \tag{12}$$

where ρ – the actual value similarity between the current state of the element and the basis, $\rho`$ – the predicted similarity of the current state of the element and the basis.

The formula shows that the higher the prediction error, the lower the trust in a given element in a given time interval. All raw data were divided into two samples: training and testing ones for two and one week respectively. Then a recurrent neural network has been trained on the training sample. The following structure has been chosen for the neural network: an input layer, a layer with long short-term memory (LSTM) elements and an output layer (Figure 1) [12]. Neural networks with such architecture have shown good results in intrusion detection and do not require adjustment of algorithm parameters [13-15].

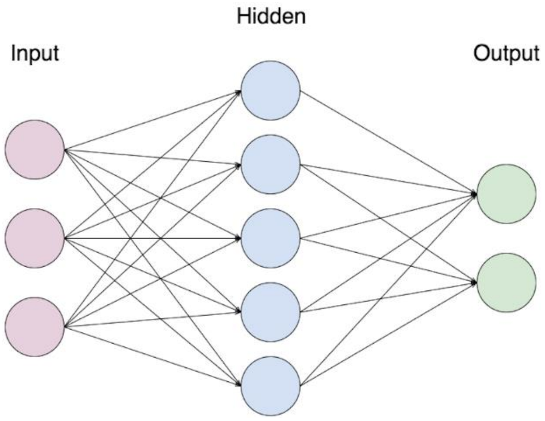


Figure 1: Neural network structure

3. CONCLUSIONS

After the experiment on the entire set we got an error of the second kind concerning the exceed of recognition accuracy and the error of the first kind, from which it can be concluded that for the majority of elements of DIS attacks do not affect the network interaction of those elements. Consequently, the conclusion about the ongoing attack should not be justified by a single networking interaction.

At the same time, the algorithm shows high recognition accuracy when dealing with the separate elements. The data contains simultaneous changes of behaviors in different groups of elements which are responsible for a large number of false positives of the proposed algorithm. This suggests the need to clarify the source data for the experiment or create a new one. Similar results were obtained in. Further research should be aimed at developing an algorithm of choosing the optimal basis.

The proposed method cannot be the only way to protect the information in the DIS. However, this algorithm will allow, on the one hand, preventing the transfer of confidential information on unreliable DIS elements, and, on the other hand, it will require more careful attention to be paid to the protection of separate DIS components. These two aspects increase the difficulty of hacking information systems in general.

ACKNOWLEDGEMENT

The work is performed according to the Russian Government Program of Competitive Growth of Kazan Federal University.

REFERENCES

[1] Hashim, M.J., Kannan, K.N., Maximiano, S. 2017. Information feedback, targeting, and coordination: an experimental study. Information Systems Research, 28(2), 289-308.

[2] Amezquita-Sanchez, J.P., Valtierra-Rodriguez, M., Aldwaik, M., Adeli, H. 2016. Neurocomputing in civil infrastructure. Scientia Iranica, 23(6), 2417-2428.

[3] Rezaeifar, Z., Wang, J., Oh, H. 2018. A trust-based method for mitigating cache poisoning in name data networking. Journal of Network and Computer Applications, 104, 117-132.

[4] Zhang, M., Gable, G.G. 2017. A systematic framework for multilevel theorizing in information systems research. Information Systems Research, 28(2), 203-224.

[5] Mui, L., Mohtashemi, M., Halberstadt, A. 2002. A computational model of trust and reputation. Proceedings of the 35th Hawaii International Conference on System Sciences, 2431-2439.

[6] Liu, F., Wang, L., Johnson, H., Zhao, H. 2015. Analysis of network trust dynamics based on the evolutionary game. Scientia Iranica, 22(6), 2548-2557.

[7] Kalala, K. 2017. Trust and reputation algorithms for hierarchically structured peer-to-peer systems. University of Ottawa, Ottawa.

[8] Seredin, O. 2001. Methods and algorithms of markless recognition of images. TSU, Moscow.

[9] Shiravi, A., Shiravi, H., Tavallaee, M., Ghorbani, A.A. 2012. Toward developing a systematic approach to generate benchmark datasets for intrusion detection. Computers & Security. Elsevier, 31, 357-374.

[10] Kent, A.D. 2015. Comprehensive, multi-source cybersecurity events. Los Alamos National Laboratory. https://csr.lanl.gov/data/cyber1/.

[11] Anan, M., Al-Fuqaha, A., Nasser, N., Mu, T.Y., Bustam, H. 2016. Empowering networking research and experimentation through software-defined networking. Journal of Network and Computer Applications, 70, 140-155.

[12] Schmidhuber, S.H.J. 1997. Long short-term memory. Neural Computation, 9(8), 1735-1780.

[13] Grasel, F.D.S., Fontoura, L.A.M. 2016. Computacional study of the electronic effects in rotational barrier of the N-arilcarbamatos N-CO bond. Periodico Tche Quimica, 13(25), 7-15.

[14] Staudemeyer, R.C. 2015. Applying long short-term memory recurrent neural networks to intrusion detection. SACJ, 56, 1-19.

[15] Turcotte, M.J.M., Heard, N.A., Kent, A.D. 2016. Modelling user behavior in a network using computer event logs. Dynamic Networks and Cyber-Security, 67-89.

Cite The Article: Aleksei A. Sychugov, Elvir M. Akhmetshin, Vyacheslav M. Grishin, Raisa N. Shpakova, Andrei V. Plotnikov (2019). Algorithm Det ermine Trust Value To The Distributed Information Systems Elements. *Journal of Mechanical Engineering Research & Developments*, 42(2): 06-09.