Journal of Computers Vol. 29 No. 3, 2018, pp. 156-164
doi:10.3966/199115992018062903014

Research on Model and Methodology of Big Data Security Situation Assessment Based on Fuzzy Set

Xiao-Lu Han[1*], Yun Liu[1], Zhen-Jiang Zhang[2], Xin Lü[3], Yang Li[3]

[1] Department of Electronic and Information Engineering, Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China
13501165037@139.com, liuyun@bjtu.edu.cn

[2] Department of Software Engineering, Key Laboratory of Communication and Information Systems, Beijing Municipal Commission of Education, Beijing Jiaotong University, Beijing 100044, China
zhjzhang1@bjtu.edu.cn

[3] Postdoctoral Scientific Workstation, State Information Center, Beijing 100045, China
lux@cei.gov.cn, liyang_cas@163.com

Received 13 September 2017; Revised 13 October 2017; Accepted 9 November 2017

Abstract. The traditional security situation assessment is usually based on one-sided data sources and does not consider the multidimensional and dynamic security features, which lead to incomplete, inaccurate and inaccurate situation assessment results, because of the various threats and attacks in cyberspace with a certain degree of concealment and changing rapidly. We propose a model and methodology of big data security situation assessment based on fuzz set, which includes two aspects: big data security situation assessment index system and fuzzy comprehensive evaluation algorithm for big data security situation assessment. Simulation results show the multi dimension and dynamic indexes can reflect the big data security situation more comprehensively and timely than the traditional single dimension static indexes, and compared with the traditional other evaluation algorithm, the fuzzy comprehensive evaluation algorithm can more effectively solve the uncertainty of big data security situation, and can more accurately reflect the situation of big data security situation.

Keywords: big data, fuzzy comprehensive evaluation, security situation assessment, security situation

1 Introduction

Big data security situational awareness is the acquisition and understanding of the security elements that can cause changes the network security situation in big data systems and forecast the future trend of network security situation in big data system. Big data security situation assessment is an important part of big data security awareness system, and a new research hotspot in the field of network security in recent years. The traditional network security situation assessment method can provide network managers with one-sided data sources or static network security situation. In the large-scale complex network space of big data system, there are many kinds of threat and attack data, which involve large network traffic and large data capacity, with the characteristics of concealment and quick change. It is difficult for decision-makers to judge the network security situation of big data system quickly and effectively.

Situational awareness first originated from the needs of military operations, and needed to understand the situation of both sides of the battlefield, to make decisions quickly. Tim Bass proposed the concept of

* Corresponding Author

cyberspace situational awareness (CSA) in 1999, Tadda and Salerno [1] proposed the network situation awareness model based on the general situation awareness model. Situation awareness model fundamentally focuses on data fusion model, including DJL model, Endsley model, etc. Salerno et al. [2] proposed a general framework for situation awareness. Giacobe N A studied the application of JDL model in cyber security [3]. Xi et al. [4] discussed the three main problems in network security situational awareness, which are the extraction of the elements in the network security situation, the comprehension of the network security situation and the projection of future situation. In the framework of situational awareness, the current network security situation assessment method research includes more evaluation method based on mathematical model, based on knowledge inference and based on pattern recognition [5]. Keramati et al. [6] proposed a method that can measure the impact of each shown attack in the attack graph on the security parameters, and can assess network security quantitatively by analyzing attack graphs. Han et al. [7] proposed A network security situation assessment model based on set pair analysis. Kou et al. [8] proposed a network security situation evaluation method based on attack intention recognition from the angle of attacker. The above research has made some new explorations in the network security situation assessment method, but big data security situation assessment model and method for complex network environment need further study.

The method of fuzzy comprehensive evaluation is a comprehensive evaluation method based on fuzzy mathematics, which transforms qualitative evaluation into quantitative evaluation according to the membership degree of fuzzy mathematics. Because of its clear result and strong systematisms, the method is applied well in the field of information security to solve the problems which is fuzzy and difficult to quantify. Xin [9] analyzes the key elements and modeling methods of information security metrics, and proposes a model of information security measurement based on baseline and a fuzzy comprehensive measurement method for information security. Tang et al. [10] proposed a index system construction method of assessment for network security based on situation entropy and gave the assessment index value calculation on the network availability. Mu et al. [11] presented an alert verification approach based on multi-level fuzzy comprehensive evaluation, which is effective to reduce false alerts and irrelevant alerts. Zhang and Yang [12] proposed a mobile Internet security situation assessment model based on improved Fuzzy Analytic Hierarchy Process, which reflects the weights of the indexes more objectively and effectively, and the evaluate results are more accurate. The above researches on the application of fuzzy comprehensive evaluation method in the field of network security are studied, however, it needs further exploration to how to establish and quantify the network security situation assessment indicators of big data system, and to use fuzzy comprehensive evaluation to solve the problems which is fuzzy and difficult to quantify in big data security situation assessment.

Through the above analysis, we can draw the conclusion that further studies is essential to solve the following problems:

(1) The existence of many uncertain factors in big data environment increases the complexity of big data security metrics. Some metrics can be quantified, while others cannot be directly quantified. Therefore, it is necessary to establish a measurable index system for big data security situation assessment, which meets the feasibility of big data security situation quantification calculation.

(2) It is necessary to research an assessment method to assess the security situation timely and accurately.

To solve the above problems, based on the current research, a model and methodology of big data security situation assessment based on fuzzy set is presented in this paper. The contributions of this paper are as follows:

(1) A measurable index system is proposed to support big data security situation assessment in this paper.

(2) the fuzzy comprehensive evaluation algorithm of big data security situation is proposed to help the decision-makers assess the network security situation.

The paper is organized as follow: In the section 2, it is proposed big data security situation assessment index system and fuzzy comprehensive evaluation algorithm. In the section 3, through the analysis and calculation of the experimental data, the experimental results are obtained. Finally, in the section 4, it summed up the conclusions of this study based on the research results, and the future will be carried out to look forward to the work.

2 Big Data Security Situation Assessment Model and Methodology

2.1 Index System of Big Data Security Situation Assessment

Big data security situation assessment model is a distributed, multi-level, hierarchical processing structure in complex network, including data level fusion, feature level fusion and decision level fusion. Multi-source heterogeneous security data fusion is to integrate, simplify and merge the security data, include the collected log data, monitoring data, content data, gathering data, service data, and to extract security feature information. Big data security features fusion is to analyze the security characteristics of big data based on the data fusion, identify the mechanism of security features, divide security features into groups, and recognize important features. Based on big data security evaluation index system to evaluate the security status of big data belongs to decision level fusion.

A index system of big data security situation assessment is shown in the following Fig. 1.

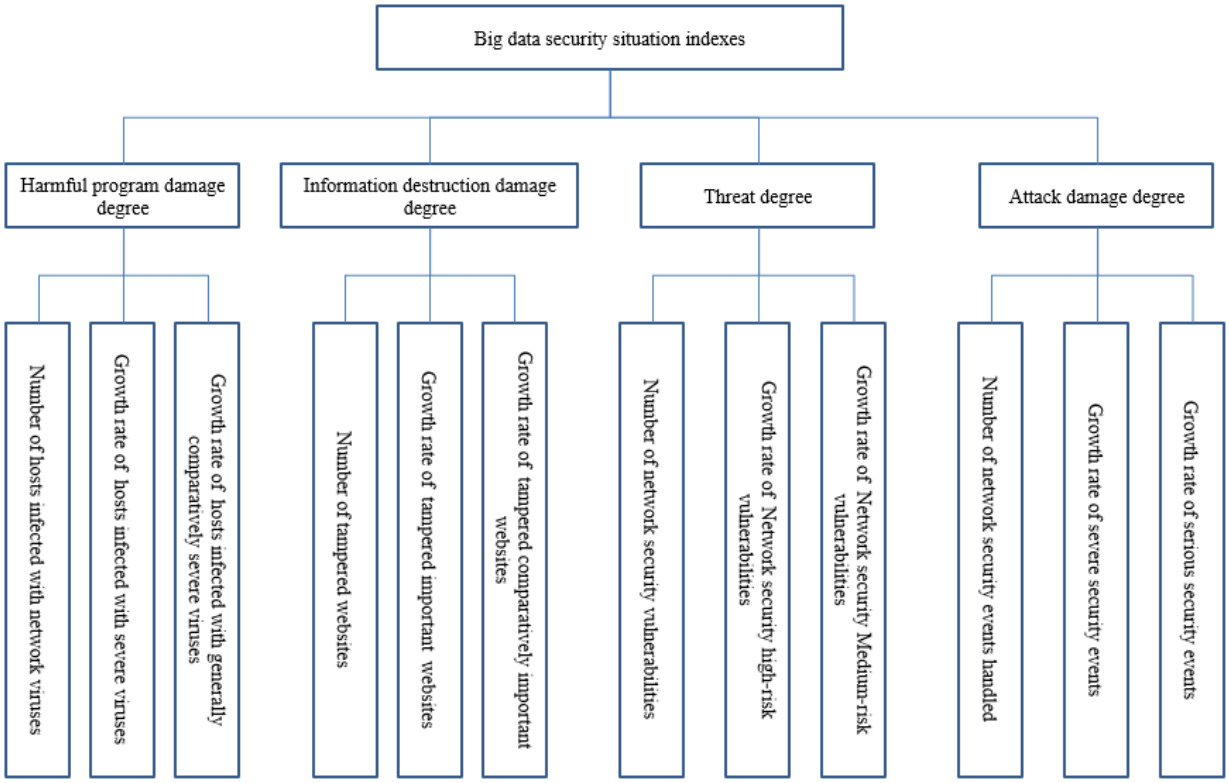

Fig. 1. Index system of big data security situation assessment

The situation index selection takes into account different levels and data sources. For each index level, the situation index reflects different object, scope and evaluation granularity. The first level index of big data security situation reflects the big data security situation from four dimensions in the damage degree of harmful procedures, the damage degree of information destruction, the degree of menace, the damage degree of attack. The second level index is designed according to the first level index, and the granularity is finer.

Definition 1. The damage degree index of harmful procedures, refers to the situation of the network virus damaging network hosts. These states mainly include the number of hosts infected with network viruses, the growth rate of hosts infected with severe viruses, the growth rate of hosts infected with generally comparatively severe viruses.

Definition 2. The damage degree index of information destruction, refers to the information security status of the web site. These states mainly include the number of tampered websites, the growth rate of tampered important websites, the growth rate of tampered comparatively important websites.

Definition 3. The degree index of threat, refers to the number of network security vulnerabilities, These states mainly include the number of network security vulnerabilities, the growth rate of network security

high-risk vulnerabilities, the growth rate of network security Medium-risk vulnerabilities.

Definition 4. The damage degree index of attack, refers to the occurrence of network events, These states mainly include the number of network security events handled, the growth rate of major security incidents, the growth rate of large security incidents.

Each index in the index system is to acquire and analyze the characteristics of different data sources by using different detection or monitoring means, and calculate the index values. The data sources of the damage degree index of harmful procedures are the detection data of the network virus such as the Trojan virus, the zombie virus and the flying worm virus. The data sources of the damage degree index of information destruction are the monitoring data of the web site information being tampered and imitated. The data source of the damage degree index of threat are the detection data of vulnerabilities such as the operating system vulnerability, the database vulnerabilities, and application vulnerabilities, and so on. The data source of the damage degree index of attack are the monitoring data of network events such as denial of service attack, unauthorized access, malicious code, and so on. There are static and dynamic indicators in the index system. Static indicators include the number of hosts infected with network viruses, the number of altered sites, the number of network security vulnerabilities, and the number of network security events handled, etc. Dynamic indicators include the growth rate of host infected, the growth rate of tampered web sites, the growth rate of network security vulnerabilities, and the growth rate of network security events handled, etc.

2.2 The Fuzzy Comprehensive Evaluation Algorithm for Big Data Security Situation Assessment

Fuzzy comprehensive evaluation algorithm is used to assess big data security situation in this paper, which include 6 steps. Step 1 is to identify big data security important features. Step 2 is to establish measurable index factor sets. Step 3 is to establish a measurement level. Step 4 is to establish fuzzy relation matrix. Step 5 is to determine the weight vector of the index. Step 6 is to calculation of big data security situation comprehensive evaluation. It is as follows:

Step 1: identify big data security important features. Enter big data security feature sample matrix X,

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1p} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mp} \end{bmatrix} \tag{1}$$

x_{ij} is the variable value of security feature j at i moment, $i \in (1,2,.....,m)$, $j \in (1,2,.....,p)$.

The judgment matrix is formed by analyzing the relation between the security characteristic factors, establishing hierarchical structure, and comparing the important degree of different feature parameters at the same level with the important degree of a criterion at the upper layer.

$$A = (a_{ij})_{p\times p}, a_{ij} > 0, a_{ij} = 1/a_{ij}, a_{ii} = 1 \tag{2}$$

The relative weights of the feature elements for the criterion are computed by the judgment matrix:

$$w_i = \frac{1}{p}\sum_{j=1}^{p}(a_{ij} / \sum_{k=1}^{p} a_{kj}) \tag{3}$$

Compute conformance ratio, and check the consistency of the judgment matrix:

$$C.R. = C.R./R.I. \tag{4}$$

If $C.R. < 0.1$, it is acceptable to the consistency of the judge matrix, otherwise the judgment should be appropriately modified to maintain a certain consistency. Among them, $C.I. = (\lambda_{\max} - P)(P-1)$ is the consistency index, $R.I.$ is the average random consistency index.

$$\lambda_{\max} = \frac{1}{p}\sum_{i=1}^{p}(\sum_{j=1}^{p} a_{ij} w_j / w_i) \tag{5}$$

Rank all the important degree of the feature indexes at this level against the upper level, and select important feature elements according to the ranking of the importance degree.

$Z=\left(z_1,z_2,.....,z_n,.....,z_p\right)^T$ is the feature vector corresponding to $\lambda_{\max}$, Z_i is the weight of feature element A_i, select the n feature elements in the case of $Z_i \geq \alpha$, α is a constant.

Output the important security feature vector $Z=\left(z_1,z_2,.....,z_n\right)^T$.

Step 2: establish measurable index factor sets. The purpose of this step is to establish a hierarchical analysis model of big data security situation evaluation index system, which is decomposed into layers until the most basic indicators and form a hierarchical structure. The first level evaluation index set is represented as:

$$U=\{U_1,U_2,\ldots,U_n\} \tag{6}$$

For each evaluation indicator U_i, the second level evaluation index is represented as:

$$U_i=\{U_{i1},U_{i2},\ldots,U_{im}\} \tag{7}$$

n is the index number of the first level indicator, m is the index number of the second level indicator. If there are three levels of indicators can be further decomposed, the basic indicators are the important features indicators of big data security which include dynamic and static indicators.

Step 3: establish a measurement level. Divide k fuzzy states $\tilde{e}_1,\tilde{e}_2,\ldots,\tilde{e}_k$ in domain $X(t)$, The fuzzy level of security situation is $E=\{\tilde{e}_1,\tilde{e}_2,\ldots,\tilde{e}_k\}$, $\tilde{e}_1$ represents the fuzzy level of security situation, $\mathrm{i}\in(1,2,.....,\mathrm{k})$.

Step 4: establish fuzzy relation matrix.

$$R=\begin{bmatrix} r_{11} & \cdots & r_{1n} \\ \vdots & \ddots & \vdots \\ r_{m1} & \cdots & r_{mn} \end{bmatrix},(0\leq r_{ij}\leq 1) \tag{8}$$

r_{ij} is the membership relation of index U_i corresponding to fuzzy grade $\tilde{e}_i$ in E.

Step 5: determine the weight vector of the index. The weight reflects the importance of this layer index relative to the upper level index, and it is obtained by the AHP method or entropy weight method. The weight vector of the first level index can be expressed as:

$$W=\{W_1,\ W_1:..,W_n\} \tag{9}$$

W is a fuzzy subset that represents the important degree of the index.

The weight vector of the second level evaluation index can be expressed as:

$$W_i=\{W_{i1},W_{i2},\ldots,W_{im}\} \tag{10}$$

Step 6: calculation of big data security situation comprehensive evaluation. It is called the fuzzy subset of the comprehensive measure that the fuzzy subset W on the index set U transform is the fuzzy set B on the measurement level I by the fuzzy relation R:

$$\mathrm{B}=\mathrm{W}\cdot\mathrm{R}=(\mathrm{b}_1,.....,\mathrm{b}_n) \tag{11}$$

"·"is generalized fuzzy composition arithmetic operator, different measurement models can be obtained according to actual situation. b_j is the membership degree of grade $\tilde{e}_i$ by comprehensive measurement, $\mathrm{j}\in(1,2,.....,\mathrm{n})$.

Normalize the fuzzy metric result of vector B:

$$\mathrm{b_j}'=\frac{b_j}{\sum_{i=1}^{n} b_i} \tag{12}$$

Get the normalized vector $B' = (b_1', b_2', ..., b_n')$ .The fuzzy grade of security situation can be determined according to the principle of maximum membership degree, or get the comprehensive score by the weighted sum.

3 Experiments and Discussions

3.1 Experiment Environment Information

To analyze the big data security evaluation model and method based on time series and multidimensional features, which is more comprehensive, timely and accurate to reflect the security situation than other network situation assessment methods, this experiment selected the data from cyber big data security situation database of a Chinese internet center as the experimental data, and verify the superiority of the model algorithm proposed in this paper comparing with other algorithms.

The experimental environment is analyzed by snort tools and My SQL in the PC server. First, the data packet of tcp replay tool is reproduced simultaneously, then, the data is collected by snort tools and My SQL, and classification data is sent to the database, finally the analysis and calculation of the data in the database is in Matlab2017a environment.

The PC server is Dell Power Edge R730, CPU Intel E5-2600 V3, six cores, memoryDDR4 4G and hard disk1.2T, the operating system is CentOS 7, the database is MySQL 5. PC client is ThinkPad X260, CPU Intel core i5 6200U, memory 4G, and hard disk 500G, the operating system is Windows10, the application system is Matlab2017a.

3.2 Experiment Data Collection

Data source of this paper are relevant record of data of infected mainframe, tampered cyber, security flaw and disposed cyber security events from cyber big data security situation database of an internet data center. The experimental data were observed every half a month from October 18, 2015 to May 8, 2016, through smoothing and noise processing the original situation data set, the data of 30 consecutive periods were obtained after noise processing. Basic experiment data set as Table 1 shows.

Table 1. Basic experiment data set

Data name	Data sources	Number of Experimental sample data	Time interval	Data dimension	Data characteristics
Network security status data	An internet data center	30	2w	4	61

3.3 Experimental Results and Analysis

This experiment selects the index system of big data security situation in section 2.1 to evaluate the security situation of big data in the network environment. The big data security situation is divided into 10 security situation fuzzy levels (0.1 to 1). The lower the situation level, the more secure, the higher the situation level, the more unsafe.

The relative weights of the feature elements are calculated by the judgment matrix, and the consistency is tested. Extract feature elements and calculate the second level index value by fuzzy metric.

The relational matrix of big data security situation index is established, and the index weight is obtained based on AHP method as Table 2 follow.

According to the big data security situation assessment index system and fuzzy comprehensive evaluation algorithm in section 2.2, the rest fuzzy comprehensive evaluation value of the big data security situation for 30 continuous time periods are calculated, big data security situation assessment changes as shown in the following Fig. 2.

The security situation values of the experimental data evaluated by the method in this paper are compared with different indicators, as shown in Fig. 3.

Table 2. The index weight

The first level Index	Harmful program damage			Information destruction damage degree			Threat degree			Attack damage degree		
The first level Index weight	0.09			0.4			0.11			0.4		
The second level Index	Number of hosts infected with network viruses	Growth rate of hosts infected with severe viruses	Growth rate of hosts infected with generally comparatively severe viruses	Number of tampered websites	Growth rate of tampered important websites	Growth rate of tampered comparatively important websites	Number of Network security vulnerabilities	Growth rate of Network security high-risk vulnerabilities	Growth rate of Network security Medium-risk vulnerabilities	Number of network security events handled	Growth rate of serious security events	Growth rate of large security events
The second Index weight	0.06	0.07	0.04	0.11	0.14	0.09	0.06	0.07	0.03	0.14	0.11	0.09

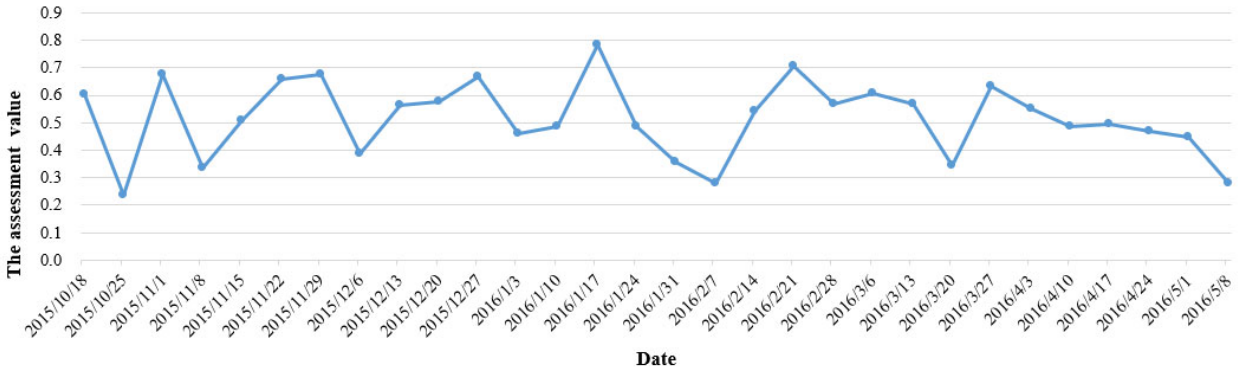

Fig. 2. The changes of big data security situation assessment for 30 continuous time periods

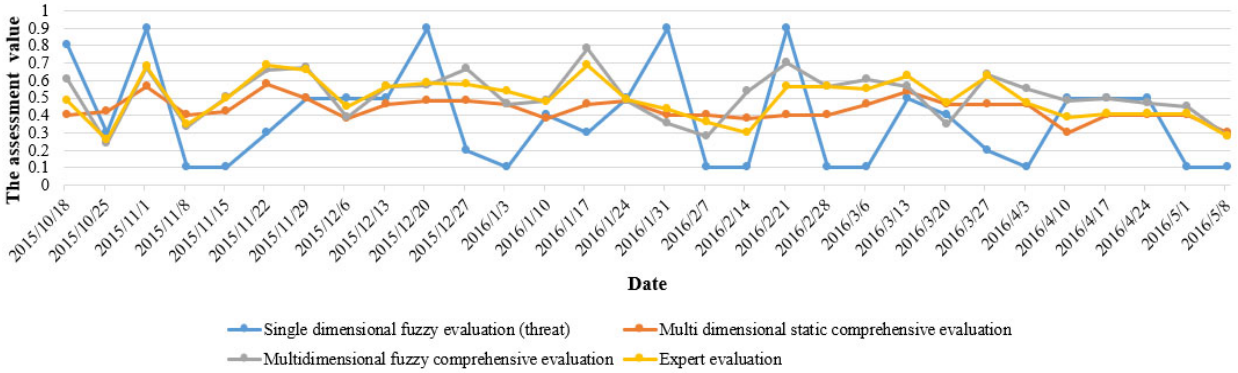

Fig. 3. Comparison of security situation assessment values with different indicators

The mean square error and root mean square error of security situation assessment values with different indicators is shown in Table 3.It analysis shows that the mean square error of the fuzzy comprehensive evaluation method with multidimensional dynamic indicators is 0.0061 in this experiment, the root-mean-square error (RMSE) is 0.0783, which is 0.0238 lower than the RMSE of the comprehensive evaluation method with multidimensional static indicators, is 0.2105 lower than the RMSE of the fuzzy evaluation method with single dimensional (threat) indicators.

The security situation values evaluated by the algorithm in this paper are compared with the values evaluated by the set pair analysis algorithm [7] and the values by expert evaluation, as shown in Fig. 4.

With the value of the expert evaluation as the reference value, the mean square error and root mean square error of security situation assessment values with different algorithm is shown in Table 4. It analysis shows that the mean square error (MSE) of the multidimensional fuzzy comprehensive evaluation algorithm is 0.0085 lower than the MSE of set pair analysis algorithm in this experiment, the root-mean-square error (RMSE) is 0.043 lower than the RMSE of set pair analysis algorithm.

Table 3. Comparison of mean square error and root mean square error of security situation assessment values with different indicators

Experimental data	The number of observed value	MSE	RMSE
Single dimension fuzzy evaluation (threat)	30	0.0834	0.2889
Multidimensional static comprehensive evaluation	30	0.0104	0.1021
Multidimensional fuzzy comprehensive evaluation	30	0.0061	0.0783

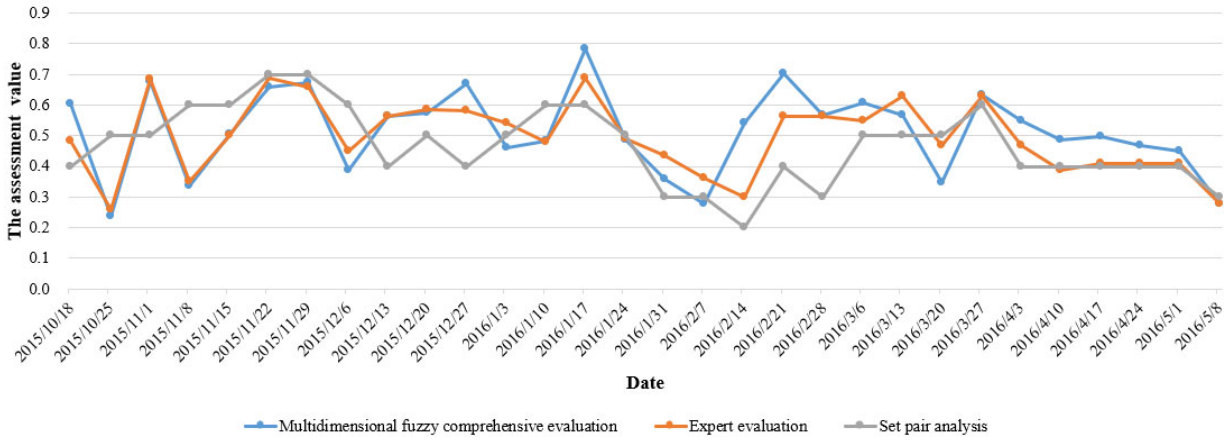

Fig. 4. Comparison of security situation assessment values with different algorithm

Table 4. Comparison of mean square error and root mean square error of security situation assessment values with different algorithm

Experimental data	The number of observed value	MSE	RMSE
Set pair analysis	30	0.0147	0.1213
Multidimensional fuzzy comprehensive evaluation	30	0.0061	0.0783

4 Conclusion

In the complex network environment, the mass security data come from various sources, has multidimensional features, and change rapidly. Previous studies are limited to qualitative or semi quantitative methods, it is particularly important to study how to evaluate the security situation of big data comprehensively, rapidly and accurately. We proposed a a model and methodology of big data security situation assessment based on fuzz set, which construct the evaluation index system of big data security situation, and use the fuzzy comprehensive evaluation algorithm to assess big data security situation. The experimental results show that: (1) In the big data environment, the change of security characteristic factor is relatively fast, and adding dynamic index to big data security situation assessment indexes can better reflect the situation change; (2) Compared with expert evaluation results, the mean square error and the root mean square error of the fuzzy comprehensive evaluation method with multidimensional dynamic indicators are smaller than the method with multidimensional static indicators and single dimensional indicators. (3) Compared with expert evaluation results, the mean square error and the root mean square error of the multidimensional fuzzy comprehensive evaluation algorithm are smaller than the method with Set pair analysis algorithm. (4) The model and methodology proposed in this paper can significantly improve the accuracy and performance of big data security situation assessment.

Acknowledgements

This research is supported by the Fundamental Research Funds for the Central Universities, No. 2017JBZ107.

References

[1] G. Tadda, J.J. Salerno, Realizing situation awareness within a cyber environment, in: Proc. SPIE, 2006.

[2] J. Salerno, M. Hinman, D. Boulware, Building a framework for situation awareness, in: Proc. the Seventh International Conference on Information Fusion, 2004.

[3] N.A. Giacobe, Application of the JDL data fusion process model for cyber security, in: Proc. SPIE, 2010.

[4] R.-R. Xi, X.-C Yun, S.-Y. Jin, Y.-Z Zhang, Research survey of network security situation awareness, Journal of Computer Applications 32(1)(2012) 1-133.

[5] Z.-H Gong, Y. Zhuo, Research on network situational awareness, Journal of Software 21(7)(2010) 1605-1619.

[6] M. Keramati, A. Akbari, M. Keramati, CVSS-based security metrics for quantitative analysis of attack graphs, in: Proc. 2013 International Conference on Computer and Knowledge Engineering, 2013.

[7] M.-N. Han, Y. Liu, Y. Chen, Network security situation assessment based on set pair analysis, Computer Application Research 29(10)(2012) 3824-3827.

[8] G. Kou, G. Tang, X. Ding, S. Wang, K. Wang, A network security situation assessment method based on attack intention perception, in: Proc. 2016 2nd IEEE International Conference on Computer and Communications (ICCC2016), 2016.

[9] L. Xin, Research on the theory and method of information system security metrics, Computer Science 35(11)(2008) 42-44.

[10] C.-H. Tang, X. Wang, R.-X. Zhang, Y. Wang, B.-H. Qiang, Research on the index system of assessment for network security based on situation entropy, Journal of Guilin University of Electronic Technology 31(4)(2011) 270-274.

[11] C.-P. Mu, H.-K. Huang, S.-F Tian, Intrusion detection alert verification based on multi-level fuzzy comprehensive evaluation, International Conference on Computational & Information Science 3801(3)(2005) 9-16.

[12] Y.-Y. Zhang, Q. Yang, Research on mobile Internet security situation assessment based on fuzzy analytic hierarchy process, Computer Engineering and Applications 52(24)(2016) 107-111.