Anti-tampering Monitoring Method of Network Sensitive Information Based on Big Data Analysis

Yi Shen(✉) and Lu Zhang

Anyang Vocational and Technical College, Anyang 455000, China
zhanglu789lll@163.com

Abstract. To improve the security of network sensitive information transmission and storage, it is necessary to design the anti-tampering monitoring of network sensitive information, and a tamper-proof monitoring technology of network sensitive information in big data environment based on big data dimension feature block is proposed. Big data feature space reconstruction method is used to calculate the grid density of network sensitive information distribution, and the network sensitive information to be tampered-proof monitoring is mapped to the divided high-dimensional phase space through the density threshold. The high dimensional phase space of information distribution is divided into dense unit and sparse unit. The coded key is matched to the corresponding network sensitive information block to realize information encryption and covert communication. The simulation results show that the information steganography performance of network sensitive information transmission and storage using this information tampering monitoring technology is better, and the information security transmission ability is improved.

Keywords: Big data · Network sensitive information · Tamper-proof monitoring · Information security

1 Introduction

With the development of big data information storage and transmission technology, the transmission and storage of network sensitive information has attracted more and more attention. The transmission of network sensitive information is to collect information through sensors [1]. Combined with big data network design, the transmission channel model is constructed to realize the communication system of information output and reception. The transmission and storage of network sensitive information has the advantages of simple networking design and large bandwidth. Under the environment of big data, the network sensitive information transmission and storage network can be constructed, which can realize the secure transmission of large-scale information data. The design of tamper-proof monitoring of network sensitive information in big data environment is the key to realize information secure communication and secure communication. It has great significance to study information tamper-proof monitoring technology to improve the security performance of communication [2].

G. Gui and L. Yun (Eds.): ADHIP 2019, LNICST 302, pp. 82–89, 2019.
https://doi.org/10.1007/978-3-030-36405-2_9

In order to achieve encrypted communication on the surface layer, the previously proposed tamper-proof monitoring algorithm for network sensitive information in a big data environment [3]. With the upgrade of hacker attack technology and the improvement of information security communication level, it cannot effectively meet the requirements of digital encryption transmission and communication. In this paper, a tamper-proof monitoring technology of network sensitive information in big data environment based on big data dimension feature block is proposed. Big data feature space reconstruction method is used to calculate the grid density of network sensitive information distribution, and the network sensitive information to be tampered-proof monitoring is mapped to the divided high-dimensional phase space through the density threshold. The coding key of the information tamper-proof monitoring algorithm is obtained, and the coding key is matched to the corresponding network sensitive information block to realize information encryption and covert communication. Finally, the simulation test and analysis are carried out, and the conclusion of effectiveness is obtained.

2 Data Coding and Tamper Proof Monitoring Key Construction

2.1 Network Sensitive Information Coding Based on Spatial Geometry Method

In order to realize the optimal design of anti-tampering monitoring technology of network sensitive information in big data environment, it is necessary to design the coding and encryption of tamper-proof monitoring data at first [4]. The grid density of network sensitive information distribution is calculated by big data feature space reconstruction method. It is assumed that the test sequence of network sensitive information coding pair is X and the training sequence is $P(r_i)$, which are G-dimensional randomly distributed binary data strings. Let $\sum_{i=1}^{n} P(r_i) = 0$ be a discrete distribution map model of order r. the coded linear mapping of network sensitive information transmission is represented as a bit sequence of c generated by X. Through spatial geometry segmentation, according to the principle of breadth first traversing, the probability distribution function of the source can be recorded as:

$$\begin{bmatrix} X \\ P \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & \dots & r_n \\ P(r_1) & P(r_2) & \dots & P(r_n) \end{bmatrix} \tag{1}$$

The data point set $X = \{x_1, x_2, \cdots, x_N\}$ is mapped into geometric space, and the membership function relationship between $P(r_i)$ and r_i is obtained according to the input R_n value:

$$I(r_i) = -logP(r_i) \tag{2}$$

Where, the log base number is set to 2, and the entropy value of network sensitive information $I(r_i)$, which represents the coding of network sensitive information in high dimensional phase space, is calculated to be $\mathrm{H}(X)$, entropy value, which reflects the average amount of data transmitted by network sensitive information:

$$\mathrm{H}(X) = -\sum_{i=1}^{n} P(r_i) \times logP(r_i) \tag{3}$$

According to the above description, the operation of obtaining network sensitive information coding based on spatial geometry is as follows:

(a) Select the initial cell and randomly select:$r_1, r_2 \in Z_q^*$;
(b) The spatial geometric density difference between the high dimensional phase space and the initial element is calculated, and when the $R = g^{r_1}$, coding output is satisfied, the spatial geometric density difference between the high dimensional phase space w_b and the initial element h_1 is calculated:

$$h_1 = \{X, w_b\} \tag{4}$$

$$wb = \{X, u_a, r_2\} \tag{5}$$

In the formula, u_a is the set density difference.

2.2 Construction of Information Tamper Proof Monitoring Key

The network sensitive information to be tampered-proof monitoring is mapped to the divided high-dimensional phase space through the density threshold, and the tamper-proof monitoring key of the network sensitive information is constructed. In the d-dimensional space, according to the grid relative density difference, the tamper resistant monitoring data are traversed deeply [5]. The grid cell sequence $g_0, g_1, \cdots, g_p$, is selected, where G is the relative density difference and g_1 is the initial key. $g_i \in NB(g_{i-1})(i = 2, 3, \cdots, p)$, represents the grid relative density difference between the channel g_1, g_0 transmitted and stored by the network sensitive information. For each i, the construction information tamper proof monitoring key satisfies:

(1) $rgdd(g_0, g_i) \leq \varepsilon$;
(2) For the dense distribution of data points in G, there is an average entropy value $j(1 < j \leq i)$, $rgdd(g_0, g_j) \leq \varepsilon$.

Then $g_i(i = 2, 3, \cdots, p)$ is called the classification key of information tampering monitoring algorithm. Through the above-mentioned design, the signcryption design of the network sensitive information is realized. According to the constructed key, the information encryption and the encrypted communication are carried out [6].

3 Improved Design of Information Tampering Monitoring Algorithm

Based on the technology of network sensitive information coding and key construction based on spatial geometry method, the anti-tampering monitoring technology of network sensitive information is improved. This paper proposes a tamper-proof monitoring technology for sensitive information in big data environment based on big data dimension features [7].

3.1 Sensitive Information Representation Processing

The sensitive information of the network for tamper-proof monitoring is mapped to the divided high-dimensional phase space through the density threshold, and the density threshold is calculated. In the discrete big data network sensitive information transmission system, the indicator function $i(n)$ is defined as:

$$i(n) = a(n)I_{1\{u(n)=1\}} = \begin{cases} 1 & if\ u(n) = 1\,,\ a(n) = 1 \\ 0 & others \end{cases} \tag{6}$$

In the initial adjacent internal unit, the symmetric key S is randomly selected to encrypt k_A to obtain $item_A$, and in the divided high dimensional phase space, the output tamper proof monitoring data is obtained by the sensitive information representation coding.

$$\frac{B \lhd \{\{dsc_A, \{item_A\}_k, w(k)\}_{k_a^{-1}}\}_{k_b}, B| \equiv \overset{k_b}{\mapsto} B}{B \lhd \{dsc_A, \{item_A\}_k, w(k)\}_{k_a^{-1}}} \tag{7}$$

The high-dimensional phase space of information distribution is divided into dense units and sparse units [8]. Data tamper-proof monitoring is performed according to the sensitive information representation model shown in Fig. 1.

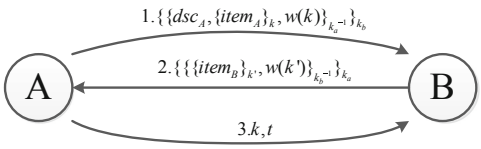


Fig. 1. Representation model of sensitive information

3.2 Code Key Matching for Information Tamper-Proof Monitoring

Because tamper-proof monitoring algorithms usually only change high-frequency data in network-sensitive information. The information classification features are screened based on the spatial geometry method [9]. The normalization of the low frequency

vector of tamper proof monitoring information is obtained by using the coding key matching strategy:

$$\frac{A \lhd \{\{\{item_B\}_{k'}, w(k')\}_{k_b^{-1}}\}_{k_a}, A| \equiv \overset{k_a}{\mapsto} A}{A \lhd \{\{item_B\}_{k'}, w(k')\}_{k_b^{-1}}} \tag{8}$$

The coding key is matched to the corresponding network sensitive information block [10], and the average entropy value of information tampering prevention monitoring is obtained:

$$\mathrm{H}_N(X) = \frac{-\sum_{i=1}^{n} P(x_i) \times \log P(x_i)}{N} \tag{9}$$

After the $\mathrm{H}_N(x_i)$ is calculated, and then the calculated result is divided by N, the variance can be expressed as follows:

$$D(X) = \sum_{i=1}^{N} [\mathrm{H}_N(x_i) - \mathrm{H}_N(X)]^2 \tag{10}$$

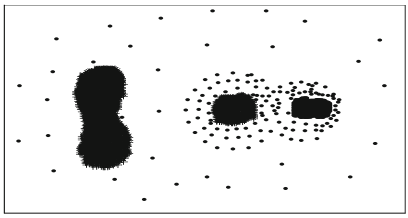
Information tamper-resistant monitoring sequence for binary sequence [11], data tamper-proof monitoring transformation $c_i = E(z_i, m_i)(i = 1, 2, 3, \cdots)$, output plaintext:

$$m_i = D(z_i, E(z_i, m_i)) = m_1 m_2 m_3 \cdots (i = 1, 2, 3, \cdots) \tag{11}$$

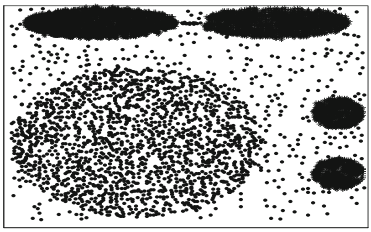
The coding key of the network sensitive information tamper-proof monitoring algorithm determined by the classification feature selection method corresponds to the in-place plane data block, and the tamper-proof monitoring can be realized directly on the data block [12].

4 Experimental Analysis

In order to test the performance of the algorithm in the anti-tampering monitoring of network sensitive information in big data environment, the simulation experiment is carried out. The experiment is simulated with Matlab7 software. The test data set is recorded as DS1 and DS2, parameters as follows: $Rn = 56, \varepsilon = 0.53, \mu_\lambda = 0.71$, the sampling frequency of the data is 1.43 Hz, and the interval of the data coding code is 100 dB. The distribution of the original data is shown in Fig. 2.



(a) DS 1

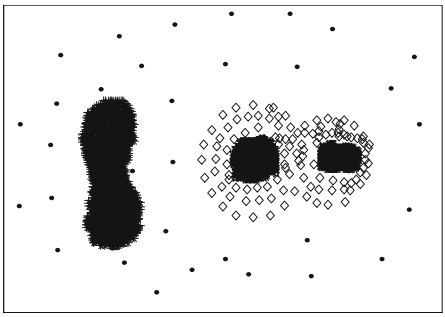


(b) DS2

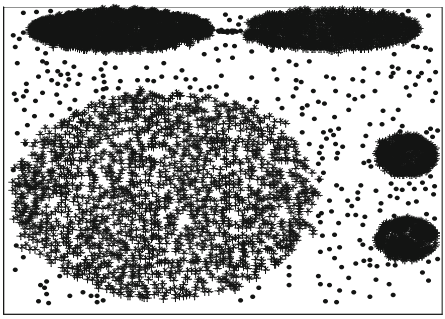
Fig. 2. Original dataset

Taking the above test data as the research object, the high dimensional phase space of information distribution is divided into dense unit and sparse unit, and the data fusion and tamper proof monitoring and processing are carried out combined with similarity feature extraction method. The output results of tamper proof monitoring are shown in Fig. 3.

The results of Fig. 3 show that the data tampering monitoring is carried out by using this method, which hides the characteristics of the original data distribution and realizes the data encryption and encryption transmission. Table 1 gives the statistical performance analysis results of data tamper-proof monitoring of two groups of data by different methods. The results show that the accuracy of this method is high and the throughput of tamper-proof monitoring data is high. The distribution performance of information sparsity is good.



(a) DS 1



(b) DS 2

Fig. 3. Data tamper proof monitoring output

Table 1. Statistical table of performance verification results under different algorithms.

Algorithm	Size of sparse distribution of information/MBit	Tamper proof monitoring accuracy/%	Tamper proof monitoring data throughput/kMB
Algorithm in this paper	60 × 60	89	93.54
	120 × 120	90	94.45
Generalized tamper proof monitoring algorithm	60 × 60	82	66.67
	120 × 120	83	72.87
Bohzmann	60 × 60	69	87.34
	120 × 120	74	80.98

5 Conclusions

In this paper, a tamper-proof monitoring technology of network sensitive information in big data environment based on big data dimension feature block is proposed. Big data feature space reconstruction method is used to calculate the grid density of network sensitive information distribution, and the network sensitive information to be tampered-proof monitoring is mapped to the divided high-dimensional phase space through the density threshold. The high dimensional phase space of information distribution is divided into dense unit and sparse unit. The coded key is matched to the corresponding network sensitive information block to realize information encryption and covert communication. The simulation results show that the information steganography performance of network sensitive information transmission and storage using this information tampering monitoring technology is better, and the information security transmission ability is improved. This method has good application value in tampering prevention and monitoring of data information.

References

1. Han, D., Chen, X., Lei, Y., et al.: Real-time data analysis system based on Spark Streaming and its application. J. Comput. Appl. **37**(5), 1263–1269 (2017)
2. Zhu, Y., Zhu, X., Wang, J.: Time series motif discovery algorithm based on subsequence full join and maximum clique. J. Comput. Appl. **39**(2), 414–420 (2019)
3. Ma, Y., Zhang, Z., Lin, C.: Research progress in similarity join query of big data. J. Comput. Appl. **38**(4), 978–986 (2018)
4. Zheng, N., Wang, J.: Evidence characteristics and attribute reduction of incomplete ordered information system. Comput. Eng. Appl. **54**(21), 43–47 (2018)
5. Yang, L., Kong, Z., Shi, H.: Multi-controller dynamic deployment strategy of software defined spatial information network. Comput. Eng. **44**(10), 58–63 (2018)
6. Luo, H., Wan, C., Kong, F.: Salient region detection algorithm via KL divergence and multi-scale merging. J. Electron. Inf. **38**(7), 1594–1601 (2016)
7. Stoean, C., Preuss, M., Stoean, R., et al.: Multimodal optimization by means of a topological species conservation algorithm. IEEE Trans. Evol. Comput. **14**(6), 842–864 (2010)
8. Liang, J.J., Qu, B.Y., Mao, X.B., et al.: Differential evolution based on fitness Euclidean-distance ratio for multimodal optimization. Neurocomputing **137**(8), 252–260 (2014)
9. Xu, G., Cheng, X.J.: Adaptive reduction algorithm of scattered point clouds based on wavelet technology. J. Tongji Univ. (Nat. Sci.) **41**(11), 1738–1743 (2013)
10. Bi, A., Wang, S.: Transfer affinity propagation clustering algorithm based on Kullback-Leiber distance. J. Electron. Inf. **38**(8), 2076–2084 (2016)
11. Long, M., Wang, J., Ding, G., et al.: Adaptation regularization, a general framework for transfer learning. IEEE Trans. Knowl. Data Eng. **26**(5), 1076–1089 (2014)
12. Bi, A., Dong, A., Wang, S.: A dynamic data stream clustering algorithm based on probability and exemplar. J. Comput. Res. Dev. **53**(5), 1029–1042 (2016)