Adaptive Encryption Model of Internet Public Opinion Information Based on Big Data

Yanjing Lu[1(✉)] and Jiajuan Fang[2]

[1] Zhengzhou Technical College, Zhengzhou 450121, China
luyanjing23233@yeah.net
[2] Department of Software Engineering, Zhengzhou Technical College, Zhengzhou 450121, China

Abstract. The traditional information encryption model takes advantage of the ergodicity of chaotic system, and processes encryption iteratively for many times. Aiming at the above problems, this paper constructs a big data-based network public opinion information adaptive encryption model. Reptiles are used to collect network public opinion information, and the public opinion information is replaced and diffused. After mining the association rules of public opinion information, the information is encrypted by Logistic mapping, and the encryption model is constructed. Compared with the two traditional encryption models, it is proved that the model has the advantages of good encryption effect, high efficiency and low cost, and can be used widely.

Keywords: Big data · Network public opinion · Adaptive encryption · Encryption model · Logistic mapping

1 Introduction

With the rapid popularization of the Internet, the Internet has become one of the main channels for the publication and dissemination of public opinions. At the same time, due to the use and development of various network services, the carriers and content forms of public opinion information also show the characteristics of diversification, including not only traditional e-mail, portal websites, blogs, post bars, forums, but also emerging microblogs, WeChat, etc., and the content of public opinion is not only loaded with news, comments, opinions forwarding, etc., but also various multimedia public opinions, showing great uncontrollability. Government agencies and related research institutions must understand and adhere to public opinion. If Internet public opinion is guided and controlled by lawbreakers, it is likely to endanger social security and stability [1]. However, in the era of big data, the various data generated by the Internet increase in the scale of PB every day, which poses new challenges to the analysis and processing of public opinion information. In addition to mining and distinguishing effective information from network public opinion information, in order to avoid the hidden dangers of social stability, network information security and other problems caused by the leakage of public

W. Fu et al. (Eds.): ICMTEL 2021, LNICST 387, pp. 344–353, 2021.
https://doi.org/10.1007/978-3-030-82562-1_32

opinion information, it is also necessary to adopt effective means to encrypt network public opinion information [2]. The expression forms of public opinion information are various, which can be text, graphics, images, sound, video and so on. Classical DES algorithm, AES algorithm or RSA algorithm is no longer suitable for network information encryption, because these encryption algorithms in the face of large capacity information encryption can not guarantee the encryption effect. The chaotic encryption model based on search mechanism makes use of the ergodicity of chaotic system and controls the number of iterations of chaotic system by the characters of text information. But this encryption model is inefficient and insecure, so it can be cracked by known plaintext attack.

Adaptive technology is a technology that, in the process of processing analytical problems, automatically adjusts the processing method according to the data characteristics of the processed data so as to adapt it to the statistical distribution characteristics of the processed data [3]. Adaptive technology is widely used in many fields such as machinery, manufacture, signal and so on. Therefore, based on the above analysis, this paper will build a big data-based network public opinion information adaptive encryption model, the following is the specific research process. The innovation part is that after mining the association rules of public opinion information, the encryption model of public opinion information is established, and the public opinion information is encrypted by Logistic mapping. Because of the diffusion nature of its association rules, the encryption process of network public opinion information is adaptive, that is, automatic adjustment of parameters, which is also the focus of this paper.

2 Big Data Based Adaptive Network Public Opinion Information Encryption Model

2.1 Internet Public Opinion Information Collection

The crawler completes the crawling task through the Downloader module, the DrawURLer module, and the FilterURLer module in turn, which are iterative, depending on the depth of the crawler site, and after the iteration, executes the DrawTexter module, from and through the following processes:

1) Web crawling: This process is intended to accomplish distributed parallel downloading of HTML text. First, the starting URL of the crawl is stored as a seed in the NEW URL, and then the hash value of the modulus N of the URL domain name (N indicates the number of data nodes in the cluster) is taken. Finally, the seed of the starting URL is distributed to the corresponding node of the starting URL value based on the computed hash value and the different results of the URL to run the subsequent program. During the reduce phase of the crawl process, you can take full advantage of the extensibility of the Big Data platform by assigning downloads to multiple threads at the same time, eventually placing the crawled web file in a file called Pages [4].
2) The process of extracting URL: firstly, the text of the Web page downloaded is parsed in a distributed and parallel manner, and then the original Web page files

of the previous stage are used as input for processing as the extracting URL. The method used to parse this procedure is a regular expression, and the parsed output is then placed in a folder named RAWURL.

3) URL deduplication: This process is run to filter separate URL links, as the extracted new URL may have been used before. To reduce the amount of crawling, the newly resolved URL is compared with the existing URL in the deduplication module to remove the URL that is repeatedly extracted [5].
4) Text extraction: After the iteration of Web page download, URL extraction, and URL deletion, needing to use regular expressions to compute the tag features of all the collected raw HTML to match the various attributes of the specified HTML extraction. Grasping Network Public Opinions by Matching Rule.

After the crawler collects the network public opinion information, the network public opinion information is processed by wavelet transform.

2.2 Public Opinion Information Processing

For the image in the public opinion information, firstly, needing to scramble the plaintext image, so that can change the pixel position of the image without changing the pixel value. The replacement method, also known as permutation transformation, is defined as follows [6]:

A set $A = (a_1, a_2, \cdots, a_n)$ is a finite set, bifurcated from itself as follows:

$$\sigma = \begin{pmatrix} a_1 & a_2 & \cdots & a_n \\ \sigma(a_1) & \sigma(a_2) & \cdots & \sigma(a_n) \end{pmatrix} \tag{1}$$
$$\sigma(a_i) \in A, 1 \leq i \leq n$$

The above bijection process is called permutation in a set A. If $|A| = n$, it is called n meta permutation.

Digital grayscale image is essentially a two-dimensional array of pixel grayscale values, that is, a matrix, if the matrix changes the value of the elements, it becomes another image. The purpose of scrambling is to rearrange the pixels of the original image to produce a new image, that is, to transform the position matrix of the original pixels into that of another permutation [7]. Although the scrambled image looks very different from the original image, the histogram of the image does not change.

In this paper, magic cube transform is used to process public opinion images. Using the rotation of the Rubik's Cube neutron block, the original surface of the Rubik's Cube will be completely disrupted by the idea, so that through the Rubik's Cube transform to achieve image scrambling.

For a matrix $I_{M\times N}$ corresponding to a digital image, the element $i_{l,k}$ of the matrix is the pixel gray value of the position (l, k) in the image, where $l = 1, 2, \cdots, M$, $k = 1, 2, \cdots, N$. The Magic Cube transformation of an image matrix is to rotate the row and column of the matrix. The rotation i_l of the row in the matrix can be seen as a circular movement of the row in a certain direction, and the rotation of the column can be seen as a similar process. The shift sum is obtained by a specific algorithm. When all the rows h_l and h_k columns of the image matrix are rotated once, the new image after

the Rubik's cube transformation is obtained, that is $I'_{M\times N}$, the scrambled image $I_{M\times N}$, which has a relationship with the image [8]:

$$I'_{M\times N} = P(I_{M\times N}) \tag{2}$$

P represents the Rubik's Cube transformation. The scrambling transformation can be carried out not only in the space domain but also in the transform domain. Generally speaking, in terms of the complexity of implementation, the scrambling transformation algorithm in spatial domain is relatively simple to implement and requires less computation. By dividing the image into certain size blocks and performing the global scrambling transformation as a unit, the algorithm can achieve better encryption effect in the subsequent encryption process.

Image diffusion can change the pixel gray value itself, and at the same time ensure the algorithm's sensitivity to plaintext. Row scanning is to scan the corresponding matrix element of the image from the upper left corner, and establish the reversible operation relationship between the former pixel value and the latter pixel value by the diffusion function. Through this method, the effect of complete diffusion can be achieved. The diffusion algorithm uses modulo addition, as shown below, to achieve the value transfer function using summation, so that if the value of the previous pixel changes, this change can be passed to the next pixel. The aim of modulus is to control the result within the range of pixel value [9].

$$c_i = (p_i + c_{i-1}) \bmod L \tag{3}$$

Among them, p_i is the current position of the pixel value, c_{i-1} is the previous pixel value of the encrypted output, L representing the image color level, for a gray image, L takes 256. If introduced the sequence generated by chaotic map into the diffusion function, it can get the diffusion encryption function that usually use. Suppose the elements x_i in the chaotic sequence generated by the chaotic map L represent the color level of the image. A sequence s_i of integers generated from a chaotic sequence can be computed as follows:

$$s_i = \left(x_i \times 10^{15}\right) \bmod L \tag{4}$$

The image is divided into several small blocks sequentially, and each pixel value of the ciphertext block is related to all pixel values of the original plaintext block. But scans usually need to be executed several times in a loop, because the diffusion can only be done inside each small block, so it is necessary to adjust the partition of encrypted blocks to achieve the diffusion of the entire plaintext.

2.3 Mining Public Opinion Information Association Rules

The increasing amount of online public opinion information has caused a lot of problems in protecting the security of public opinion data, and there are a lot of association rules between public opinion information. Alien attackers can collect the relevant public opinion information data by obtaining the association rules in the data. Therefore, before

encrypting the public opinion data, it need to mine and hide the association rules in the public opinion data.

The data set $I = \{i_1, i_2, \cdots, i_m\}$ of network public opinion information is a collection of public opinion information data items D in the public opinion information data set T. For an item set X in the public opinion information dataset, the association rules between the item set and the public opinion information dataset are $X \rightarrow Y, X \subseteq I, Y \subseteq I, X \cap Y = \phi$.

This paper uses FP-tree frequency set algorithm to mine association rules of public opinion information data [10]. Firstly, the public opinion information database is scanned to record the frequent transactions in the public opinion information data set and its corresponding support. The formula for computing support for frequent transaction items is as follows:

$$\sup(X) = \frac{count(X)}{|D|} \tag{5}$$

In Formula (5), $count(X)$ is the number of transactions containing frequent transaction items in the transaction database corresponding to the network public opinion information data set; x is the total number of frequent transaction items in the transaction database. Create the FP-Tree root node based on the frequency items that are first scanned. If the frequent items do not have the same prefix in FP-Tree, a new branch is added; if the frequent items have the same prefix transaction items, only increase the number of nodes, to construct FP-tree. In FP-Tree, the path is traversed to find the minimum path whose support value is the sum of the support degrees of all nodes in the current path. Then the path of FP-Tree with minimum support is the association rules among the internal transaction items in the data set, that is, the association rules of public opinion data are mined. According to the association rules of public opinion information mined, the network public opinion information is encrypted by Logistic chaotic map.

2.4 Building Encryption Model

In this paper, a typical Logistic map is used as a chaotic system for generating encrypted sequences. The Logistic map is as follows:

$$\begin{cases} x_{n+1} = \mu x_n (1 - x_n) \\ 0 \leq \mu \leq 4 \\ x_n \in (0, 1) \end{cases} \tag{6}$$

The Logistic map iterates through the following equations:

$$x(t+1) = rx(t)[1 - x(t)] \tag{7}$$

Among them, t is the synchronization for the iteration time, for any t, $x(t) \in [0, 1]$, r is an adjustable parameter. When parameters $r \in [0, 4]$, you can ensure $x(t)$ that the mapping is always inside [0, 1]. The equation exhibits different dynamic r limiting behaviors when different parameters are changed.

The encryption process steps are described below:

Step 1: Treat the encrypted public opinion information plaintext integer wavelet transform.

Step 2: Select the parameters μ and initial value x_0 of the Logistic map, replace it with the Logistic map, get the chaotic sequence, and discard the previous N_0 point, that is $N_0 + 1$, from the $N_0 + M \times N/4$ value to the end.

Step 3: The chaotic sequence obtained in Step 2 is diffused and converted to obtain the public opinion information after encryption and complete the encryption process.

So far, it have completed the research on the construction of adaptive encryption model of network public opinion information based on big data.

3 Experimental Study

Above, the paper studies the big data-based network public opinion information adaptive encryption model, this section will design an experiment to verify the effectiveness and feasibility of the encryption model.

3.1 Experimental Content

In this simulation, it compare the big data-based adaptive encryption model with the traditional search-based encryption model and the homomorphic encryption model, and compare it with the traditional network information encryption model. Through comparative experiments, the paper validates and analyzes the above model of network public opinion information adaptive encryption based on big data. The simulation experiment to run procedures in Visual Studio 2014 as the development environment, using the C + programming language to achieve efficient experimental content.

In the experimental index, the encryption overhead is represented by the authentication throughput before and after encryption coding, and the encryption effect is represented by the correlation between plaintext and ciphertext.

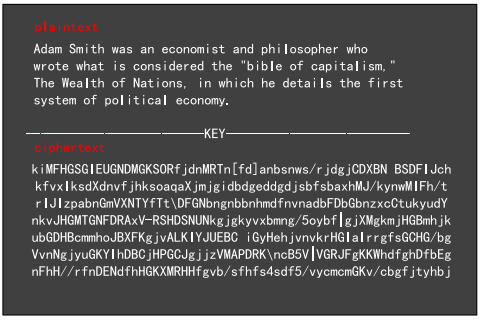
$$r_{xy} = \frac{\operatorname{cov}(x, y)}{\sqrt{D(x) \cdot D(y)}} \tag{8}$$

In the above expression, x is plaintext; y is ciphertext; $\operatorname{cov}(x, y)$ is covariance of plaintext and ciphertext; $D(x)$ is variance of plaintext; $D(y)$ is variance of ciphertext. In the process of experiment, the concrete data of corresponding index is recorded, and the experiment conclusion is obtained after analyzing and processing in MATLAB software. Process and analyze the experimental data, draw the corresponding conclusions, thus completing the simulation experiment preset verification goal.

3.2 Experimental Results

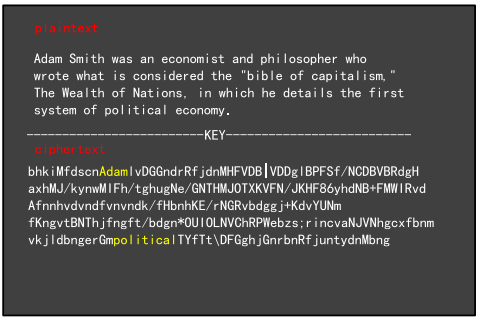
Take the text public opinion information as an example, the encryption effects of the 3 groups of information encryption models are shown in the following Fig. 1.

According to the above analysis, the experimental group encryption model can completely encrypt the text information, while the control group A model and the control

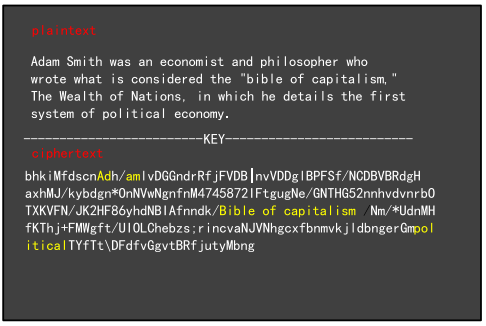


(a) Experimental group model

(b) Contrast group model A



(c) Contrast group model A



(d) Contrast group model B

Fig. 1. Comparison of text encryption

group B model can not encrypt the text completely. Specifically, the encryption effect of the control group A model is better than that of the control group B.

The data of three kinds of network public opinion information encryption models are shown in the following table, and the relationship between the data in the table is analyzed (Table 1).

Table 1. Comparison of experimental data on encryption time consuming and correlation

Experimental data/GB	Experience group		Control group A		Control group B	
	Encryption time/S	Relevance	Encryption time/S	Relevance	Encryption time/S	Relevance
10	1.12	0.0146	1.16	0.0194	1.16	0.0229
20	1.18	0.0161	1.20	0.0199	1.22	0.0251
30	1.15	0.0140	1.19	0.0212	1.18	0.0245
35	1.22	0.0160	1.26	0.0196	1.25	0.0236
40	1.23	0.0162	1.31	0.0205	1.37	0.0253
45	1.24	0.0164	1.35	0.0192	1.43	0.0252
50	1.26	0.0152	1.39	0.0203	1.51	0.0215
55	1.27	0.0155	1.52	0.0201	1.56	0.0222
60	1.26	0.0145	1.68	0.0214	1.73	0.0229
65	1.24	0.0162	1.77	0.0195	1.82	0.0221
70	1.29	0.0151	1.86	0.0207	1.98	0.0244
80	1.28	0.0156	2.01	0.0209	2.24	0.0232

Analysis of the above table shows that when the data is less than 35 GB, the time difference of three encryption models is less than 0.04 s. With the increase of encrypted data, the time of B encryption increased rapidly, but the time of the other two groups increased slightly. When the data amount is less than 50 GB, the difference of encryption time between experimental group and control group A is less than 0.15 s. With the increase of encrypted data, the encrypted time of control group A increased rapidly, which was larger than that of control group B and less than that of experimental group B. The average encryption time of the three encryption models is 1.23 s in the experiment, 1.475 s in the A model and 1.54 s in the B model respectively. Compared with the other two models, the experimental model can reduce encryption time by at least 20% on average. The correlation between plaintext and ciphertext before and after encryption of experimental group is lower than that of the other two groups, which shows that the experimental group has better encryption effect.

Before and after the encryption of the transmission information, the authentication throughput of the public opinion data information in the network is shown in the Fig. 2, and the relationship between curves is analyzed.

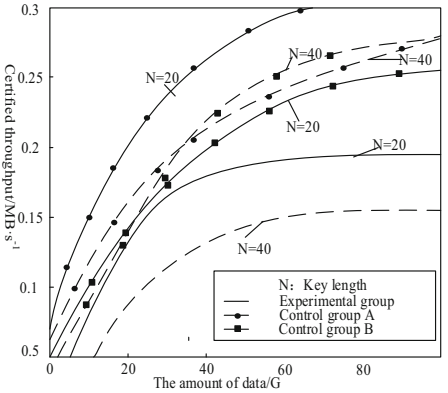


Fig. 2. Comparison of data authentication throughput before and after encryption coding

According to the relationship between the curves in the graph above, the data authentication throughput of the experimental group is faster than that of the other two groups after encrypting different data. Except for the experimental group model, the authentication throughput of the other two groups showed an increasing trend, while the authentication throughput of the experimental group tended to be stable after increasing for a period of time. The higher the authentication throughput is, the greater the encryption overhead of the encryption model is. It can be concluded from the above that the experimental group model can still guarantee a lower encryption overhead when encrypting data.

To sum up, the big data-based network public opinion information adaptive encryption model has the advantages of high encryption efficiency, good encryption effect, low cost, compared with traditional encryption model.

4 Closing Remarks

In order to ensure the security of public opinion information, this paper constructs a big data-based adaptive encryption model of online public opinion information. Compared with the traditional model, the experimental results show that the proposed model has better performance.

References

1. Hamami, F., Dahlan, I.A., Prakosa, S.W., et al.: Implementation face recognition attendance monitoring system for lab surveillance with hash encryption. J. Phys. Conf. Ser. **1641**(1), 012084 (6pp) (2020)
2. Hussain, A., Kiah, M.L.M., Anuar, N.B., et al.: Performance and security challenges digital rights management (DRM) approaches using fog computing for data provenance: a survey. J. Med. Imaging Health Inform. **10**(10), 2404–2420 (2020)
3. Qi, H.: Double encryption method of network privacy information based on dynamic key selection. J. Heilongjiang Univ. Technol. (Compr. Edn.) **20**(03), 89–93 (2020)

4. Chongrui, T., Zhaoxiang, L., Yuxin, L.: Simulation of anonymous privacy protection method based on dynamic data mining. Computer **36**(11), 171–174 233 (2019)
5. Xin, D., Ji, J., Jing, F., et al.: Efficient fully homomorphic encryption scheme using ring-LWE. J. Phys. Conf. Ser. **1738**(1), 012105 (8pp) (2021)
6. Wei, T., Qiping, H., Tangzhi, W.: Research on location big data encryption method based on privacy protection. J. Anhui Electr. Eng. Prof. Tech. Coll. **24**(01), 118–122 (2019)
7. Fu, W., Liu, S., Srivastava, G.: Optimization of big data scheduling in social networks. Entropy **21**(9), 902 (2019)
8. Liu, S., Li, Z., Zhang, Y., et al.: Introduction of key problems in long-distance learning and training. Mob. Netw. Appl. **24**(1), 1–4 (2019)
9. Liu, S., Liu, D., Srivastava, G., et al.: Overview and methods of correlation filter algorithms in object tracking. Complex Intell. Syst. (2020). https://doi.org/10.1007/s40747-020-00161-4
10. Niu, J., Li, X., Gao, J., et al.: Blockchain-based anti-key-leakage key aggregation searchable encryption for IoT. IEEE Internet Things J. **7**(2), 1502–1518 (2020)