Research on Trade Data Encryption of Tobacco Enterprises Based on Adversarial Neural Network

Zhang Yi (✉ hider1986@163.com)
Renmin University of China

Research Article

Keywords: Tobacco enterprise trade data, Adversarial neural network, Data classification, Encryption protocol, Data encryption

Posted Date: September 21st, 2021

DOI: https://doi.org/10.21203/rs.3.rs-898184/v1

Version of Record: A version of this preprint was published at Soft Computing on December 1st, 2021. See the published version at https://doi.org/10.1007/s00500-021-06479-6.

Research on trade data encryption of tobacco enterprises based on adversarial neural network

Zhang Yi
School of Economics, Renmin University of China, 100089,Beijing, China
Email: hider1986@163.com

Abstract: In order to overcome the problems of long encryption process, low data security and poor anti attack rate of traditional tobacco enterprise trade data encryption methods, a trade data encryption of tobacco enterprises based on adversarial neural network is proposed. This method optimizes the traditional neural network by generating countermeasure network, so as to form adversarial neural network. In the adversarial neural network, the encryption processing of tobacco enterprise trade data is completed through data feature classification, design of tobacco enterprise trade data encryption protocol and data encryption channel. The experimental results show that the encryption process of this method takes between 6s-20s, and the encryption efficiency is high, and this method can effectively scramble the original arrangement of the data, so as to effectively hide the effective information in the tobacco enterprise trade data, improve the anti attack rate of the data, and effectively improve the security of the data.

Key words: Tobacco enterprise trade data; Adversarial neural network; Data classification; Encryption protocol; Data encryption

1 Introduction

With the steady progress of the reform and development strategy of the national tobacco industry, tobacco enterprises have gradually become bigger and stronger through combination, reorganization and merger. The industrial organization system structure shows an obvious trend of integration and centralization. This change in the mode of production, operation, organization and management of the industry puts forward higher-level requirements for the construction of tobacco informatization. This requirement is mainly reflected in the use of informatization to support industry centralized management and auxiliary decision support. After years of information construction and rapid market growth, China's tobacco industry has a considerable production and market scale, and has made positive contributions to the national, local finance at all levels and national economic construction. A series of tobacco trade centers have established relatively mature business systems, including three-level batch business system, website system, sales system, etc. these business systems have produced a large amount of data and are gradually developing towards mass. These data are very detailed and stored in different computer systems. With the continuous development of electronic information and the Internet, the transmission and storage methods of these data have also changed greatly [1-2].

However, because the trade data of tobacco enterprises are easy to copy and spread, it will bring many security threats to commerce [3]. At the same time, in order to meet the growing industry demand and the continuous progress of network and information technology, the secure transmission of tobacco enterprise trade data in network and insecure channel has become an urgent problem to be solved [4]. Data security research involves many contents, including security architecture, cryptographic protocol, security theory, security performance analysis, etc. among them, encryption technology is one of the key technologies to ensure information security.

There have been many significant advances in the development of artificial intelligence in recent years. In some cases, AI can learn on its own and perform some tasks better than humans. For example, in reference [5], the key data feedback encryption method based on scheduling model is designed. In this method, the key encrypted data is segmented using 3GR encrypted string segmentation algorithm, and then AEHA comprehensive authentication model is constructed. The truncated string is optimized to complete data encryption on the basis of accelerated authentication. In reference [6], a data homomorphic encryption method based on improved RSA algorithm is designed. This method first uses DES algorithm to encrypt plaintext and RSA algorithm to encrypt key, then calculates ciphertext sequence in plaintext and ciphertext space through addition homomorphism process, and then obtains plaintext data through corresponding decryption operation.

However, it is found in the practical application that the traditional data encryption method has the problems of the encryption process time, data security and anti-attack rate. Therefore, this study designed a tobacco enterprise trade data encryption method based on adversarial neural network.

2 Adversarial neural network analysis

2.1 Traditional neural network

Neural network is a computer model structure with a wide range of applications, which contains a large number of upper and lower related units, and these basic units are named neurons [7]. These neurons are divided into at least two layers according to the logical structure of the network, and different layers are interconnected through weighted connections. These weighted parameters largely determine the characteristics of the two associated neurons [8]. In the structure of neural network, basically each neuron is connected with the neuron at the next level. After the data is input from the input layer, the data processing results are output in the output layer through the action of multiple neurons. The area between the input layer and the output layer is called the intermediate layer and its role is to construct a complex network function to complete data processing.

Neural network can not only realize the function of self-repair, but also has high fault tolerance, which can avoid data vulnerabilities and abnormal nodes. In addition, the neural network model can also infer the correlation between different data nodes, which enables it to have the ability of non-linear data processing [9].

However, although the neural network has so many advantages above, it relies too much on the performance of the system hardware itself, that is, it cannot explain the network behavior, nor can it choose the optimal network structure, nor can it set the training time in the process of network training [10]. These disadvantages also restrict the development of neural network to a great extent.

2.2 Adversarial neural network

Based on the above analysis of neural network, this study optimized the traditional neural network by generating adversarial network, so as to form adversarial neural network and apply it to the encryption process of trade data of tobacco enterprises.

Generative adversarial network can effectively distinguish the distribution of high-dimensional and complex data, and it applies adversarial method to mine the generative algorithm of data distribution [11]. In generative adversarial network, there are two main structures: generative module S and discriminant module P. The purpose of generating modules is to generate false data according to the implicit variable y, while the purpose of discriminating modules is to form real data space based on the input judgment parameters. The discriminant module can output a specific value if it is able to determine that the input information is from a real data space. Because the generative module and the discriminant module compete with each other, they are antagonistic. The network parameters can be optimized in this adversarial learning way, and the process is as follows:

$$\min_S \max_P V(S,P) = \min_S \max_P p_{x-data} \log D(x) + p_y \log(1-D(y)) \quad (1)$$

In formula (1), p_{x-data} and p_y respectively represent the distribution probability of real data and the distribution probability of the hidden variable y in the data space. $V(S,P)$ is a binary cross entropy function. The generating module maps the data into the real data space, and then the discriminant module determines whether the output data x is a fake sample or a real sample.

The real sample x identified through the above process is transmitted to the input layer as the input information of the neural network. From the point of view of data property classification, $V(S,P)$ belongs to a kind of objective function. If x is real data, the discriminant module maximizes its output; If x is the pseudo data formed in the generated module, identify the module to minimize its output. Therefore, formula (1) will yield at least $\log(1-D(y))$ entries. In this process, the generating module tries to maximize the output information of the discriminant module, while the discriminant module tries to maximize the objective function, which results in the minimax relationship in formula (1), thus enabling the neural network to have the ability of classification and discrimination [12].

The structure of adversarial neural network is shown in Figure 1.

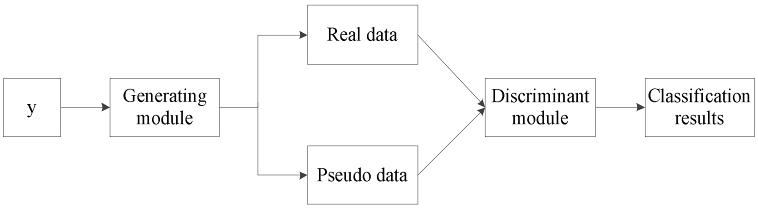


Figure 1 Schematic diagram of adversarial neural network structure

3 Encrypt tobacco enterprise trade data based on adversarial neural network

By inputting the trade data of tobacco enterprises into the adjunctive neural network designed above, real data classification results will be obtained. On this basis, the obtained data will be encrypted to ensure the security of trade data of tobacco enterprises.

3.1 Characteristic classification of trade data of tobacco enterprises

This paper first uses Bayes' theorem to classify the characteristics of tobacco enterprise trade data, which is an important theorem to judge the relationship between prior probability and posterior probability. When judging the probability of an event through known information, the obtained probability is prior probability, and the posterior probability can be obtained by deducing prior probability [13]. Bayes' theorem is as follows:

$$P(A|B)=\frac{P(A)P(B|A)}{P(B)} \quad (2)$$

In formula (2), $P(A)$ and $P(B)$ respectively represent the prior probabilities of event A and event B, and $P(A|B)$ and $P(B|A)$ respectively represent the posterior probabilities of event B and event A when event A and B are known.

Naive Bayes classifier makes full use of Bayes' theorem, which consists of three stages: preparation, training and calculation. In the preparation stage, the attributes of the data are divided; in the training stage, the samples are trained by the attributes of the divided samples; in the calculation stage, the probability of the occurrence of samples of different categories and the conditional probability of the occurrence of different data attributes in samples of different categories are calculated. On this basis, the final data feature classification result is obtained by comparing with the probability obtained in the training stage.

Suppose there is trade data set C of tobacco enterprises, and $C=\{C_1,C_2,\mathrm{L}\ ,C_n\}$, n represents the data category in the set, and $A=\{a_1,a_2,\mathrm{L}\ ,a_m\}$ represents the samples to be classified with the number of attributes m. The naive Bayes classification algorithm is adopted to obtain the maximum value of $P(C_n|A)$, and the process is as follows:

$$P(C_n|A)=\frac{P(A|C_n)P(C_n)}{P(A)} \quad (3)$$

If $P(A)$ of each random classification is in the same state, only the largest molecule can be analyzed in the classification process. All attribute sets in sample set A are in independent states, then formula (3) is transformed into the following form:

$$P(A|C_n)=\prod_{1}^{m}P(A_m|C_n) \quad (4)$$

If sample set A is at $i=q$, $P(A|C_n)$ has the maximum value, then the target classification is q.

3.2 Tobacco industry trade data encryption protocol

The results of feature classification are obtained by using known data information through naive Bayes, and then applied to the subsequent encryption process. In this paper, data encryption protocol and encryption channel are designed to complete encryption design.

Firstly, by chaotic key protocol control process is complete data encryption arithmetic coding and quantitative parameters in the process of optimization, and the sequence to encrypt the data set $X=\{x_1,x_2,...,x_i,...,x_n\}$, where $x_i(1\le i\le n)$ is the characteristics of the i-th a section coding sequence, the characteristics of the trade data distribution in tobacco enterprise is domain, the fuzzy characteristics of the data encryption process sequence are defined as follows:

$$k_{x_i} = \frac{x_i\left(k+1\right)}{pk} \quad (5)$$

In formula (5), k_{x_i} represents the private key of the data sender [14]. On this basis, the spatial quantization coding method is adopted to reconstruct the data features, and the extended ciphertext sequence of the tobacco enterprise trade data to be encrypted is $X\left(n\right) = x_1 + x_2 + ... + x_n$. Then, the proxy reschedule method is used for the block matching of sparse data features, and the block matching model is obtained as follows:

$$f_{k_{x_i}} = \left[sk_{ij} \cdot k_{x_i}\right]_{4\times4} Ck_{x_i} \quad (6)$$

In the block matching model, the convergence control method is adopted to implement linear random control on the trade data of tobacco enterprises, and the decryption key is obtained as follows:

$$S = \sum_i f_{k_{x_i}} \sigma_{k_{x_i}} \quad (7)$$

In formula (7), $\sigma_{k_{x_i}}$ represents the similarity probability of ciphertext distribution in the trade data set of tobacco enterprises. Under the control of the private key, the extended key protocol to encrypt the trade data of tobacco enterprises is:

$$Z\left(k\right) = \frac{f_{k_{x_i}}}{Sp} \quad (8)$$

In formula (8), p represents the fuzzy constraint parameter in the data encryption process.

3.3 Data encryption channel design

Traditional data encryption methods only encrypt data within the local network, ignoring the encryption process in data interaction. Aiming at this problem, the design of data encryption channel is added in this paper. In this part, the data encryption environment is quantified based on the efficient analysis and processing capability of the computing node big data environment.

The conventional data channel state is determined by the runout frequency and runout mode of data node. The influence of multiple signal frequencies is involved [15]. The traditional channel frequency is static and its anti-attack is poor. Based on the analysis results of big data, this study dynamically extracts the characteristics of different signals in the channel, cancelling the corresponding wave frequency, so as to achieve the purpose of quantifying the channel environment. On this basis, the channel after quantization is encrypted and isolated, and the encryption layer and channel are fused together by joint optimization algorithm. The joint optimization process is as follows:

$$G(u) = \int_1^{\partial} u(h)^{\partial} du \quad (9)$$

In formula (9), u represents the channel frequency; h represents the synthesis degree of the encryption layer; ∂ represents the data node coefficient [16]. In formula (9), the value of synthesis degree of encryption layer is related to the encryption quality of the whole data channel. The larger the value of synthesis degree of encryption layer is, the more secure the data channel is. The value range of synthesis degree of encryption layer is dynamically allocated by the hidden layer of adversarial neural network according to data needs. The data encryption channel structure is shown in Figure 2.

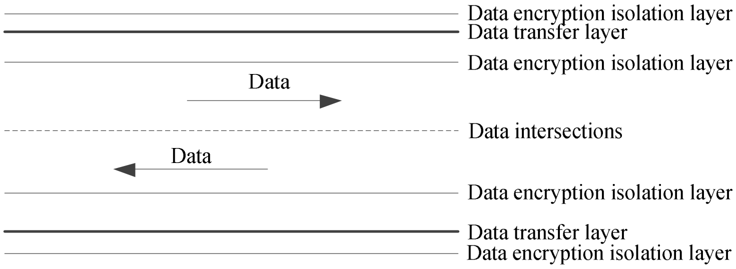

Data encryption isolation layer
Data transfer layer
Data encryption isolation layer
Data
Data intersections
Data
Data encryption isolation layer
Data transfer layer
Data encryption isolation layer

Figure 2 Schematic diagram of data encryption channel structure

At this point, the design of the encryption method of tobacco enterprise trade data based on adversarial neural network is completed.

4 Experiment and result analysis

4.1 Experimental design

In order to verify the practical application performance of the tobacco enterprise trade data encryption method designed above based on adversarial neural network, the following simulation experiment is designed using Matlab platform for verification.

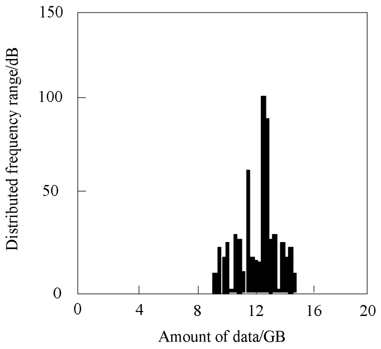
The experiment takes the trade database of a tobacco enterprise as the experimental object, and takes the trade big data generated in the operation process of the enterprise as the experimental sample. The enterprise database contains 20GB of data, including 15 data sets. The experimental environment was configured with two computers, both of which were configured with CUP I5 6200 (main frequency 3.2Hz, memory 8G). The other experimental environments and parameters were shown in Table 1.

Table 1 Experimental environment and parameters

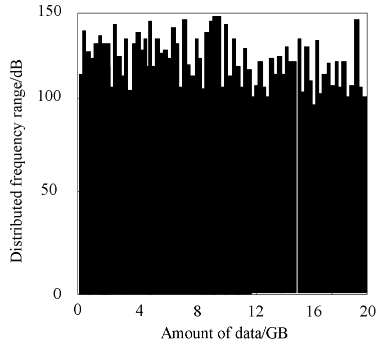
Project	Content
The clustering distribution length of trade data of tobacco enterprises	1200
Displacement of data point and encryption elastic node	12.5
Similarity properties between data	0.41
The dimension of the data sample	220

4.2 Preliminary results

Firstly, the proposed method is used to encrypt the trade information data in the enterprise database, which is used to preliminarily verify the application effect of the proposed method. The output histogram of data before and after encryption is shown in Figure 3.



(a) Histogram distribution of raw plaintext data



(b) Histogram distribution of encrypted data

Figure 3 Output histogram before and after data encryption

By analyzing the results shown in Figure 3, it can be seen that the distribution frequency range of trade information data in the enterprise database is effectively encrypted after the application of the proposed method, indicating that the proposed method has good encryption performance and high data security.

4.3 Compare the experimental results and analysis

Order to avoid the singularity, the experimental results in reference [5] design feedback scheduling model based on key data encryption method and the design in reference [6] data homomorphic encryption method based on RSA algorithm for comparison method, from the encryption process takes three angles, data security, and resistance to attack rate, launches the performance verification together with the method in this paper.

4.3.1 Time verification of different methods of encryption process

Through several iterations, the encryption process time of the method in this paper, reference [5] method and reference [6] method were counted, and the results obtained were shown in Figure 4.

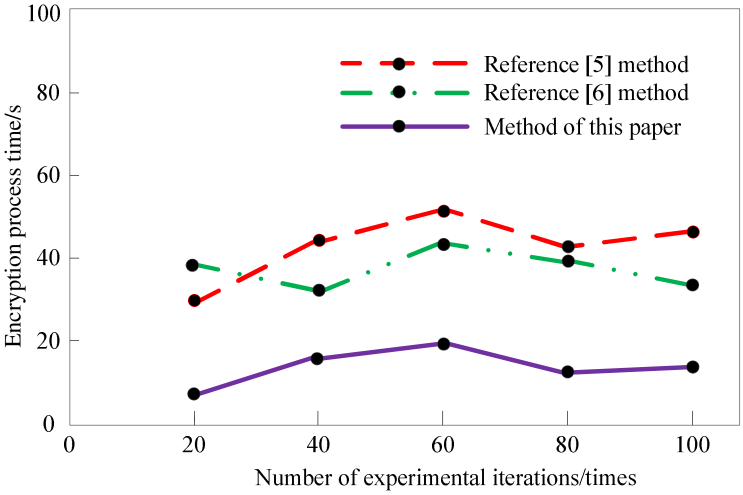
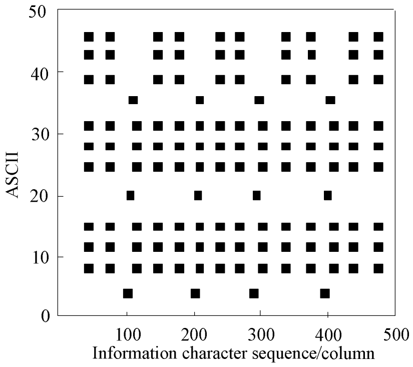


Figure 4 Time comparison of encryption process with different methods

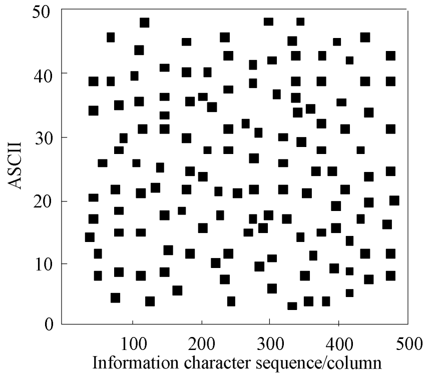
The analysis of the variation trend of line segments in Figure 4 shows that with the constant change of the number of experimental iterations, the encryption process time of the three methods also changes accordingly. Among them, the encryption process time of the reference [5] method varies between 28s-54s, the encryption process time of the reference [6] method varies between 34s-46s, while the encryption process time of the method in this paper varies between 6s-20s, which is significantly less than the two traditional methods. Therefore, it takes less time to verify the encryption process of the system in this paper, indicating that the method in this paper has higher encryption efficiency.

4.3.2 Different methods of data security verification

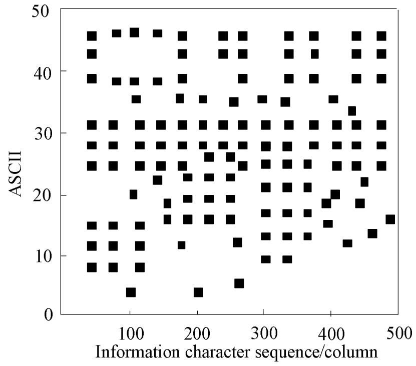
Security is one of the most effective indicators to verify the application effect of data encryption. Therefore, the distribution of standard code values of information exchange (ASCII) was used as an indicator in the experiment to verify the data security after the application of the method in this paper, reference [5] method and reference [6] method. The experimental results are shown in Figure 5.



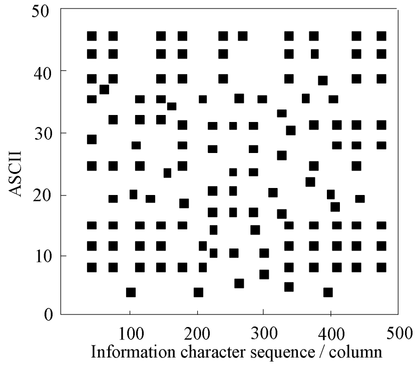
(a) ASCII distribution of data before encryption



(b) The distribution of ASCII values of data encrypted by the method in this paper



(c)ASCII value distribution of data encrypted by the reference [5] method



(d)ASCII value distribution of data encrypted by the reference [6] method

Figure 5 Comparison of ASCII value distribution of data before and after encryption by different methods

Through the analysis of the results shown in Figure 5, it can be seen that the proposed method can effectively scramble the original arrangement of data after encrypting the trade data of tobacco enterprises, thus effectively hiding the effective information involved in the original trade data of tobacco enterprises and improving the security of data. However, the two traditional methods failed to effectively scramble the original arrangement of the data, and the attacker could still extract part of the delay information from the encrypted data. This shows that the encryption effect of the method in this paper is significant, which proves that its security is stronger.

4.3.3 The test of data anti-attack rate by different methods

In order to verify the data anti attack rate of different methods, in this part of the experiment, three attack modes of data tampering attack, forgery attack and injection attack are simulated, and the anti attack times and anti attack rate of data subjected to different external attacks after encryption by this method, reference [5] method and reference [6] method are counted. The results are shown in Table 2.

Table 2 Comparison of anti-attack rates of different methods

Different methods	Aggressive behavior	Number of attacks/times	Hits/times	Resistance to attack rate/%
Method of this paper	Tamper with the attack	200	1	99.50
	Forgery attacks		3	98.50
	Injection attacks		1	99.50
Reference [5] method	Tamper with the attack		27	86.50
	Forgery attacks		15	92.50
	Injection attacks		18	91.00
Reference [6] method	Tamper with the attack		44	78.00
	Forgery attacks		20	90.00
	Injection attacks		16	92.00

By analyzing the data in Table 2, it can be seen that the trade data of tobacco enterprises encrypted by this method can effectively resist three types of attacks, with the highest anti-attack rate up to 99.50%, proving that the trade data of tobacco enterprises encrypted by this method has good security performance. However, the anti-attack rate of the tobacco enterprise trade data encrypted by the reference [5] method and reference [6] method is significantly lower than that of the method in this paper. The above experimental results show that the proposed method can effectively ensure that the encrypted trade data of tobacco enterprises can resist the risk of information disclosure caused by external attacks, and the anti-attack ability of data is significantly improved.

5 Conclusion

Aiming at the problems of traditional data encryption methods, such as poor encryption duration, data security and anti-attack rate, this study designed a new trade data encryption method for tobacco enterprises based on adjunctive neural network. According to the experimental results, the proposed method not only has high encryption efficiency, but also can effectively scramble the original data arrangement, so as to effectively hide the effective information involved in the original trade data of tobacco enterprises, and improve the anti-attack ability of the data, proving that it has achieved the design expectation.

Declarations

Funding

No funding is received.

Research involving Human Participants and/or Animals:

Not applicable

Informed consent:

Not applicable

Conflicts of interests:

The authors declare that they have no conflicts of interest.

Availability of data and material:

No data available.

Code availability

Not Applicable.

Authors' Contributions

Zhang Yi have participated in the designing framework, collecting data, analysis and interpretation of data, drafting, reviewing and approving the manuscript.

References

[1]Kumar P, Bhatt A K. Enhancing multi-tenancy security in the cloud computing using hybrid ECC-based data encryption approach[J]. IET Communications, 2020, 14(18):3212-3222.

[2]Sharma D, Jinwala D. Simple index based symmetric searchable encryption with result verifiability[J]. Frontiers of Computer Science (print), 2021, 15(2):169-174.

[3]Perez-Resa A, Garcia-Bosque M, Sanchez-Azqueta C, et al. A New Method for Format Preserving Encryption in High Data Rate Communications[J]. IEEE Access, 2020, 11(3):1-14.

[4]Ikeda K, Sato Y, Koyama O, et al. Two-dimensional encryption system for secure free-space optical communication of time-series data streams[J]. Electronics Letters, 2019, 55(13):752-754.

[5]Huang Z W.Key data feedback encryption method for ship embedded system based on scheduling model[J].Ship Science and Technology,2019,41(10):124-126.

[6]BAO Haiyan, LU Cailin. Homomorphic encryption of privacy data set based on improved RSA algorithm[J]. Journal of Terahertz Science and Electronic Information Technology,2020,18(05):929-933.

[7]Ahmed, Elhadad. Data sharing using proxy re-encryption based on DNA computing[J]. Soft Computing, 2020, 24(3):2101-2108.

[8]Yoshida K, Fujino T. Countermeasure against Backdoor Attack on Neural Networks Utilizing Knowledge Distillation[J]. Journal of Signal Processing, 2020, 24(4):141-144.

[9]Jss A, Sj B. On the computational power of the light: A plan for breaking data encryption standard[J]. Theoretical Computer Science, 2019, 773(1):71-78.

[10]Saracevic M H, Adamovic S Z, Miskovic V A, et al. Data Encryption for Internet of Things Applications Based on Catalan Objects and Two Combinatorial Structures[J]. IEEE Transactions on Reliability, 2020, 36(9):1-12.

[11]Vuppala A, Roshan R S, Nawaz S, et al. An Efficient Optimization and Secured Triple Data Encryption Standard Using Enhanced Key Scheduling Algorithm[J]. Procedia Computer Science, 2020, 171(1):1054-1063.

[12]Chabanne H, Guiga L. A Protection against the Extraction of Neural Network Models[J]. Arxiv, 2020, 14(5):329-337.

[13]Baagyere E Y, Agbedemnab A N, Zhen Q, et al. A Multi-Layered Data Encryption and Decryption Scheme Based on Genetic Algorithm and Residual Numbers[J]. IEEE Access, 2020, 33(9):1-16.

[14]Masuda H, Nakai T, Yoshida K, et al. Black-Box Adversarial Attack against Deep Neural Network Classifier Utilizing Quantized Probability Output[J]. Journal of Signal Processing, 2020, 24(4):145-148.

[15]Vengala D, Kavitha D, Kumar A. Secure data transmission on a distributed cloud server with the help of HMCA and data encryption using optimized CP-ABE-ECC[J]. Cluster Computing, 2020, 23(5):51-58.

[16]LIU Y, BAI J N. Key Data Encryption Protection Simulation of Complex Electronic

Information System[J]. Computer Simulation, 2019, 36(10):269-272.