

# Research and Application of Agricultural Data Security Exchange based on JSON

Changxia Sun<sup>1</sup>, Xia Zeng<sup>1</sup>, Zheng Zhou<sup>2</sup>, Hongfeng Niu<sup>3</sup>, Shufeng Xiong<sup>1\*</sup>

<sup>1</sup>Department of Data Science and Big Data Technology, College of Information and Management Science, Henan Agricultural University, Zhengzhou, 450046, China

<sup>2</sup>College of Animal Science and Technology, Henan Agricultural University, Zhengzhou, 450046, China

<sup>3</sup>School of Automation, Faculty of Telecommunications, Xi'an Jiaotong University, Xi'an, 710049, China

\*Corresponding Author.

## Abstract:

With the development of agricultural production information in China, a large number of agriculture-related data with complex categories and diverse structures have been produced in the process of agricultural production, operation and management. The data interaction between application systems in the industry presents the characteristics of high frequency, complexity and diversification. These data in the process of exchange is subject to security threats such as interception, counterfeiting, tampering, rejection, etc. In order to ensure the safe, reliable and efficient data exchange, a complete, extensible, high-performance security exchange system is indispensable. In this paper, the data exchange between Henan Province Rural Comprehensive Information Service Platform—the Zhong yuan Information Port and other sites is taken as the research object. Based on cryptographic methods, the data encryption and decryption system architecture is streamlined and materialized, and then the overall architecture of the system implementation is proposed. Using JAVA programming language and RSA encryption algorithm, the agricultural data secure exchange system based on lightweight JSON format is realized.

**Keywords:** Data acquisition, Data classification, Data encryption, Data security transmission, JSON.

---

## I. INTRODUCTION

At present, information security have become the first problem that needs to be considered in the development of various industries with the in-depth development of the Internet industry. In daily life, people generate large amounts of data through the Internet all the time in various aspects such as dining, shopping, housing, and transportation and so on. In different fields, intricate application will also involve multifaceted data aggregation. These large amounts of heterogeneous data not only involve everyone's privacy, but also have huge commercial value. Many Internet companies use large amounts of data to

predict the probability of future events. These data revolve agricultural data, medical data and other data in all aspects of life. At the same time, these data maximize the commercial value and social value of data, and provide great help for the social progress and our lives.

The International Organization for Standardization defines data security as: security is to minimize the possibility of attacks on resources and data. In 2007, the State Council issued the "Administrative Measures for Information Security Grading Protection", which is China's first normative document on information security grading protection, and proposed that the construction and implementation of intranet security is a crucial aspect of information security protection [1]. However, network service attacks, network virus transmission, data and information theft and other malicious network attacks often occur in the process of transmission and storage of large amounts of information on the Internet. Once the important information is stolen by criminals, it will also cause huge losses to the development of the industry and the country's economic development [2]. Therefore, the secure transmission of data has attracted national attention and has become a research hotspot for many scholars.

In this paper, the data between Henan Province Rural Comprehensive Information Service Platform - the Zhong Yuan Information Port and other sites is taken as the research object and is divided into three categories: public data, ordinary data, and private data. For the secret data, the secure transmission of agricultural data is realized using cryptography methods. In addition, to improve the efficiency of data transmission, JSON is chosen to be the data exchange format in the data secure transmission system.

## **II. METHODS**

### **2.1 Cryptography**

Cryptography is the foundation of information security. Cryptography is divided into symmetric cryptography and asymmetric cryptography. Symmetric cryptography is also called traditional cryptography. The encryption key and decryption key are the same in symmetric cryptography. The symmetric cryptography has many advantages as follows, small amount of calculation, high speed of encryption and so on. However, since encryption and decryption use the same key, it may face huge risks. One is that the key owner may leak the key, the other is that the key may be intercepted by the attacker, and thus cause the leakage of the key when it is transmitted in the network environment. Therefore, the security of the symmetric cryptography largely depends on the security of the key. There are many common symmetric cryptography, such as DES, 3DES, AES and so on.

Asymmetric cryptography is also called public key cryptography [3-4]. The encryption key and the decryption key are two different. The encryption key is called the public key, and the decryption key is called the private key. The public key is public and the private key is kept secret. Based on math

problem, a pair of public and private keys is generated, the information that is encrypted by the public key can be only decrypted by the private key. So the public key and the private key come in pairs. In asymmetric encryption, only the public key needs to be transmitted and the public key is for the encryption not for decryption. Therefore, even if the attacker obtains the public key, the data cannot be decrypted, which ensures the security of data transmission. The asymmetric cryptography is precisely used to solve the problems of key leakage and key distribution in symmetric cryptography. But it has many shortcomings as follows, large amount of calculation and the low speed of encryption and decryption. Therefore, the asymmetric encryption is usually not used to encrypt large amounts of data and it commonly is used to transmit the key of the symmetric cryptography. Currently, there are many common asymmetric cryptography, for example RSA, DSA, and etc.

The RSA asymmetric cryptography [5] proposed by Ron Rivest, Adi Shamir and Leonard Adleman in 1978 has been regarded as a cryptography with high security performance until now. The asymmetric encryption algorithm RSA is described as follows:

- (1) Select any two large prime numbers  $p$  and  $q$ ,  $p$  and  $q$  to keep secret.
- (2) Computing  $n = pq$ ,  $\varphi(n) = (p - 1)(q - 1)$ . The  $n$  is public,  $\varphi(n)$  is private.
- (3) Arbitrarily select the positive integer  $e$ ,  $1 < e < \varphi(n)$ , such that  $\gcd(e, \varphi(n)) = 1$ .  $e$  is the public encryption key.
- (4) Computing  $d$ , such that  $de = 1 \pmod{\varphi(n)}$ .  $d$  is the secret decryption key.
- (5) Encryption transformation: for plaintext  $m \in \mathbb{Z}_n$ , cipher text is

$$c = m^e \pmod{n} \quad (1)$$

- (6) Decryption transformation: for the cipher text  $C \in \mathbb{Z}_n$ , the plaintext is

$$m = c^d \pmod{n} \quad (2)$$

The encryption principle of RSA encryption algorithm is based on the difficulty in factorization of a large number which is extremely complicated.

The multiplication of two prime numbers in mathematics,  $e$  and  $n$  are public keys,  $\varphi(n)$  is kept

secretly as the private key. If  $\phi(n)$  is known by others, the security of the RSA asymmetric encryption algorithm will be threatened. In addition, the security of the RSA encryption algorithm is closely related to the length of  $p$  and  $q$ , the longer the length of  $p$  and  $q$ , the the higher security it is. Therefore, when using RSA asymmetric encryption algorithm, the length of  $p$  and  $q$  cannot be less than 512 bits.

## 2.2 Data Exchange Format

To standardize the data format, the data to be transmitted must be encapsulated into a unified message object before the data exchange between the client and the server. At present, the main data exchange formats are Xml, JSON, and etc.

Xml [6] (Xtensible Markup Language, Extensible Markup Language) is a markup language. Using text encoding, text-related information (such as text structure and presentation information) and the original text is combined and marked with a markup language. Xml is usually used to mark electronic documents and data, define data structures, and replace expressions with marked content, and the original files are recovered through the marked content. Xml is not a tool for displaying data, but a tool for exchanging data. It transmits data between various computer applications in a certain format. The huge advantage of Xml is that users can freely define self-identifying labels for specific applications, for instance defining a set of self-identifying labels for internal data exchange and a set of labels that can be recognized by related parties for external data exchange [7]. The Xml data exchange format can provide support for data exchange in different network environments. Xml has the advantages of rigorous grammatical format, good verification mechanism, strong self-description, and cross-platform. However, it also has shortcomings such as difficulty in updating, communication difficulties, and low efficiency. Xml is very efficient for processing structured data information, but it will bring the problem of occupying too much auxiliary space. Therefore, if the amount of data to be transmitted is not very large, the user can choose the Xml data format for transmission. However, when the amount of data is large, other data formats should be considered for transmission due to the low data exchange efficiency.

JSON (JavaScript Object Notation, JavaScript Object Notation) is a lightweight data exchange format [8] which is based on data, easy for humans to read and write, and generated and parsed by computer. JSON is a JavaScript native format, which means that JavaScript doesn't need any special API or toolkits to parse JSON data format, so JSON is especially suitable for applications where JavaScript is used. In addition, many programming languages such as Java, C/C++, PHP and Python provide generators and parsers for JSON data format [9]. At present, JSON format has become the preferred data exchange between major websites, clients and servers. The main reasons are as follows: firstly, compared with other formats, JSON format has the advantages of simply written and easily read because of adopting independent language text format. Especially compared with Xml format, JSON format parsing is more concise and simple just with some square braces, braces instead of complex tags.

Therefore, for a large amount of data, JSON format whose parsing is simple without complicated parsing compared with parsing Xml. As a result, it takes the parsing of JSON significantly less time to parse compared with the parsing of Xml [10].

Therefore, for the advantages and disadvantages of Xml and JSON data exchange format, the data exchange between Henan Province Rural Comprehensive Information Service Platform - the Zhong Yuan Information Port and other sites is taken as the research object in this paper. The lightweight JSON data exchange format is chosen as data exchange format in the data transmission system with effectively improving the high efficiency of data exchange.

### 2.3 Web Crawler

Web crawler [11] are also called web spider, which can automatically grab web content through a set of rules like a machine. Web crawlers are usually used to automatically collect target information, and can complete the verification of the validity of the connection of the crawling target site to avoid generating incorrect data. It is regularly used in search engines for crawling and storing of the required sites. Also, it is usually used to collect amounts of information in data analysis and mining. The web crawler crawls on the web of HTML document links when working like a spider, starting from a certain node, filtering the web content according to the web analysis rule algorithm, filtering out the effective links and stacking them in the crawling queue, then using the search strategy to obtain the target URL for the next step, and repeats this process until the program requirements are satisfied.

## III. DESIGN OF THE SECURE AGRICULTURAL DATA EXCHANGE PLATFORM BASED ON JSON

In this paper, to implement the security of the data transmission system, some of the work done is shown below:

Firstly, some data on agriculture are collected, analyzed and then stored. Secondly, the data on agriculture are divided into three categories: public data, ordinary data, and private data. Then, through the research and analysis of common data exchange format, the lightweight JSON format is selected as the data exchange format in the transmission system. Next, for the private data on agriculture, RSA encryption algorithm is used to ensure secure data transmission. Finally, the secure agricultural data exchange platform based on JSON is designed with Java language. The detailed technical scheme is shown in Fig 1.

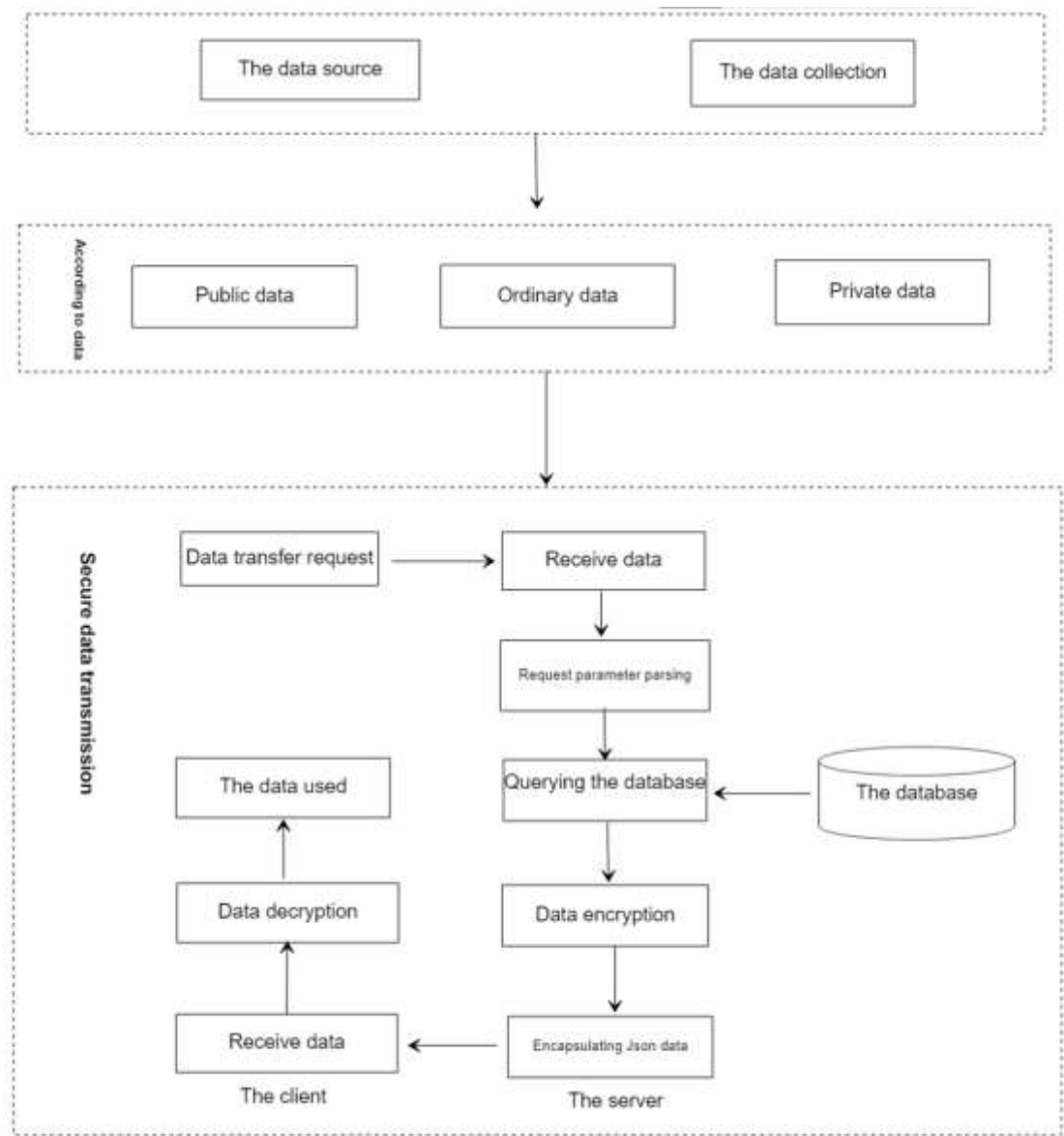


Fig 1: technology roadmap for the secure data exchange platform

3.1 Data Classification

To ensure the security and effectiveness of the encryption algorithm, the data are classified before encryption [12]. According to the security privacy protection in Big Data Classification Guide for Information Technology [13] that was promulgated by The State Administration for Market Regulation and The State Standards Committee in 2020, it is pointed out that the classification of the data security privacy protection is to classified according to the sensitivity of the data content. The following elements

are shown below:

(1) The sensitivity of the data, that is, whether the data itself or its derivative data involve state secrets, corporate secrets or personal privacy.

(2) The privacy of the data, namely, the range of data that can be known.

(3) The importance of the data, in other words, it is the extent of the damage to national security, corporate interests or civil rights after the disclosure, loss, abuse, tampering or destruction of data which are not authorized.

According to the above research content, based on the data sensitivity, the data on agriculture are divided into three categories: public data, ordinary data, and private data. In this paper, users use customized rules to automatically identify the sensitivity of data. Public data refers to data that can be completely open. Ordinary data mainly involves the rights and interests of other people's organizations, but is allowed to disclose slightly sensitive data, such as agricultural pollution, crop diseases and insect pests, meteorological disasters, and water resources information. Private data mainly refers to that many data in the agricultural production process are not allowed to be disclosed in principle, such as planting user information (name, ID card, bank card), and etc. The disclosure of the above information may cause uneasiness and loss of interests of agricultural producers. It is conducive to the development of agricultural production. So, When it comes to relevant information, it should be processed accordingly. These sensitive data should be encrypted and then encapsulated in a data exchange format.

In this paper, the public data and the ordinary data come from the National Bureau of Statistics, which is directly under the State Council. It is mainly responsible for the administration of national statistics and national economic accounting. So the data which were crawled from it are accurate, reliable and can be used. Because the private data involve the sensitivity of data, the data of this article are test data. The classification of the data is shown in Table I.

**TABLE I. Data classification table**

Public data	Ordinary data	Private data
Grain information	Agricultural pollution	ID card
Nursery information	Meteorological disaster	Bank card
Seeding information	Weather forecast	Correspondence address
Pesticide information	Water resources	Date of birth
Fertilizer Information	land resource	mobile phone number
Irrigation information	Biological resources	Mail



Agricultural Machinery Information	Disaster data	Price quotation
Agricultural Information	Market supply and demand information	Market information
Yield information	Cargo circulation information	Price and profit

### 3.2 Designing of the Data Acquisition

The method of the data acquisition is based on the Java crawler technology. The overall framework of the data collection design is shown in Fig 2.

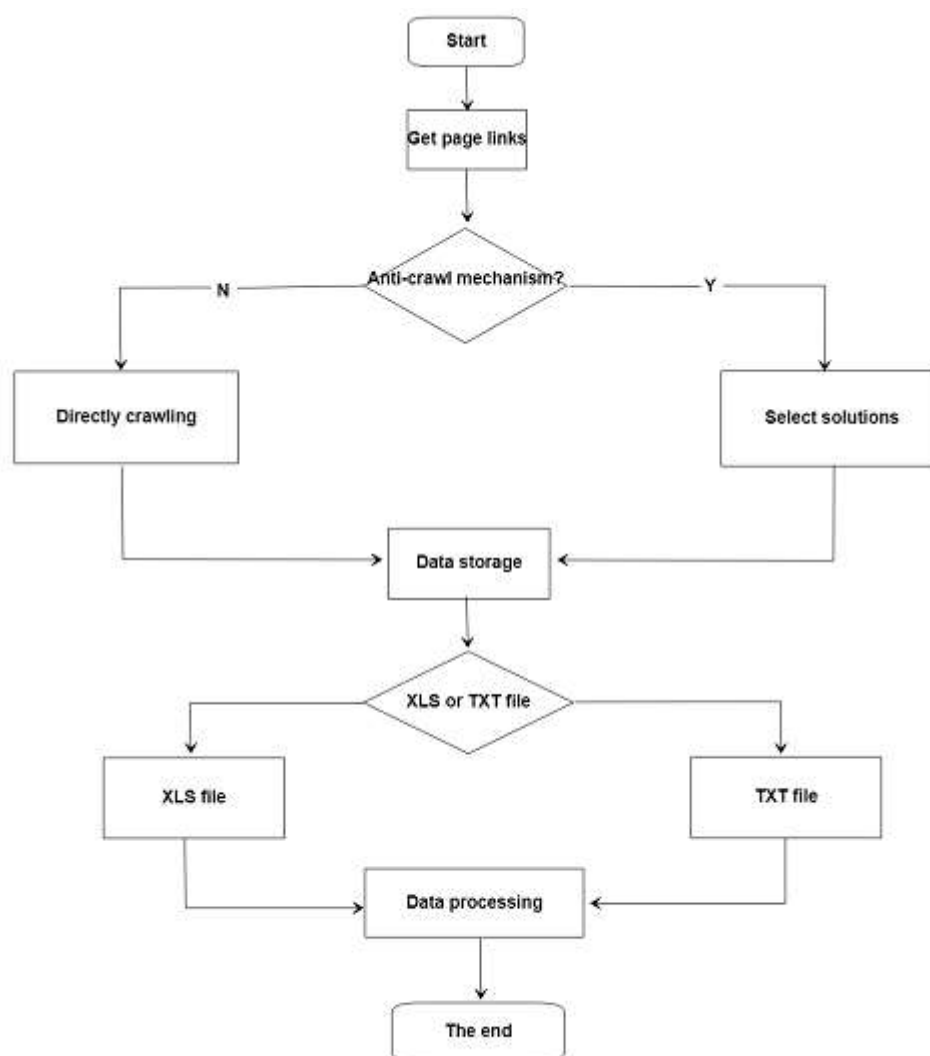




Fig 2: data acquisition overall design framework

When constructing the crawler model, pay attention to the parameters of the data and the single page and paging of the data. The difference between the website of the National Bureau of Statistics and ordinary websites is the identifier of the data to which is obtained according. However, because of the particularity of the National Bureau of Statistics, the method of obtaining data request parameters is adopted in this paper. After obtaining the data, it will be stored in the MySQL database in order of rows and columns to complete the data collection work.

Common solutions to anti-crawl mechanism are shown as follows:

(1) Crawl with a small amount of data (crawling after simulating login, or using cookies to crawl directly);

(2) Use IP agent to crawl the website;

(3) Modify the access frequency of the program.

The third solution to securely obtain data is mainly introduced to this article. First, modify the access frequency of the program. Use the sleep method in the time package to make the program rest for two seconds after each data acquisition, which can effectively avoid the hidden danger of being blocked due to high frequency access to the site. Thus, the stability and reliability of data are ensured during the process of the data acquisition.

The method of request parameters is used to obtain the data in this paper. Taking the total output and index of the agriculture, the forestry, the animal husbandry and the fishery as an example, press F12 on the request page to view the data source URL, and then copy it with the Java code to use requests to request data. The response message is returned in JSON format. After obtaining the response message, the JSON characters need to be parsed to obtain the data needed in the experiment.

After the data is acquired, operations such as data cleaning are usually performed to remove the abnormal values. Because in this data collection, the data has been collected and acquired according to the required data requirements when acquiring the data, so it can be stored and used without data cleaning. After extracting the required data, it is sorted and saved in the MySQL database.

According to the relevant elements of large scale agricultural data collection and the needs of data security transmission system in the later, Java crawler technology is used to crawl the public data and

ordinary data required in this paper and then the data are saved in the MySQL database.

### 3.3 Design of the Secure Transmission System

#### 3.3.1 Key generation

There are many ways to generate public and private keys, such as OpenSSL and key generation tools of the Java language. OpenSSL is a secure Sockets layer password library software package with open source code. It implements ASN.1 certificate and key standards. In the standards, OpenSSL provides encryption and protection for private keys, so that private keys can be stored and distributed securely. Since its development language is C, it is used on a variety of platforms, such as Linux, Windows, Mac, and etc. At the same time, it provides very powerful functions, including cryptography algorithm library, application program and SSL protocol library. For public encryption, it provides DH algorithm, RSA algorithm, DSA algorithm and elliptic curve encryption algorithm. Also, the key generation tool of the Java language can be used through the tool path is %JAVA\_HOME%\bin\keytool.exe in jdk1.4 or later. In the actual development environment, we more use the algorithm tool class provided by JDK to generate by ourselves, which is also convenient for development and debugging. The process of generation is shown in Table II.

**TABLE II. Key generation process**

Process: Key generation
Input: plain text Output: cipher text after public key encryption process: (1) Initialize the RSA algorithm to generate the object Key Pair Generator (2) Initialize the key pair generator and specify the key length of 1024 (3) Generate a pair of public key and private key according to the formula (4) Return the key pair

During development and testing, the length of the key is 1024. Because the key length less than 1024 is not secure. In fact, in order to ensure the security of the data, it is best to use 3072, 4096, and even larger number of bits of key length in the internet environments. Because the time of encryption and decryption will increase with a larger number of bits, the length of the key depends on the actual situation. In this paper, a pair of public and private key is obtained through writing code where the public key is in the server and a private key is in the client. Then encryption and decryption is in the data exchange process.

### 3.3.2 Data encryption

In order to ensure the security of the data, the RSA encryption is used when communicating with the externally provided interface, that is, the data returned by the server is encrypted with the public key, and only those clients with the private key can decrypt the correct data. The specific implementation process as shown in Table III.

**TABLE III. Data encryption process**

Request: Encryption algorithm steps
Input: Plain text data
Output: cipher text data
process:
(1) Initialize the algorithm object Cipher
(2) Initialize the general variables
(3) While the remaining plain text bytes are > 0
(4) Cipher. Do Final ( plain text )
(5) end
(6) Return to cipher text

The specific implementation process of the RSA encryption is as follows: firstly generate two keys. Among which the client holds the private key, the client initiates a request to the server, and the server obtains the public key of the client, and then uses it to encrypt the information. After receiving the information, the client uses the private key to decrypt it into plain text.

### 3.3.3 Data block encryption and decryption

In the actual transmission process, the length of a single message is often bigger than (the byte length of the key-11), and the plain text needs to be encrypted in blocks. Take the 1024-length public key as an example. Since the message that can be encrypted each time is 117 bytes, the message body is split according to the size of each packet of 117 bytes, and then the packet is encrypted one by one, and then the cipher text is synthesized as the final cipher text and exchanged between platforms.

In this way, it is possible to encrypt messages of any size without being limited by the length limit of (the number of bytes in the key length-11). The encryption result is as follows: {"birthday": "2019-12-28", "address": Henan Agricultural University information Management and Communication Office, Zhengzhou City, Henan Province", "data": "R29lsIb0f5sap35/rOXhaW1KtU99ttLPuYkUi3qWgODC3wduF5pccXfluKre/kzw+7vwx4Y8brwUh070jWBTdfjcnDJaWJbX9anPfStbuvnewQynkJocm1kkehbeJuz1IjrE61BFfeFHoBYH2I7

L+4/nZ6XKOCmIsy6ldpDfFY=", "nick,Name": "mahengzhao", "id": 1616980236519, "email": mahengzhao@163.com, "status": "1"}.

After receiving the cipher text, the client must decrypt to get the plain text. The formula used for decryption is as follows:

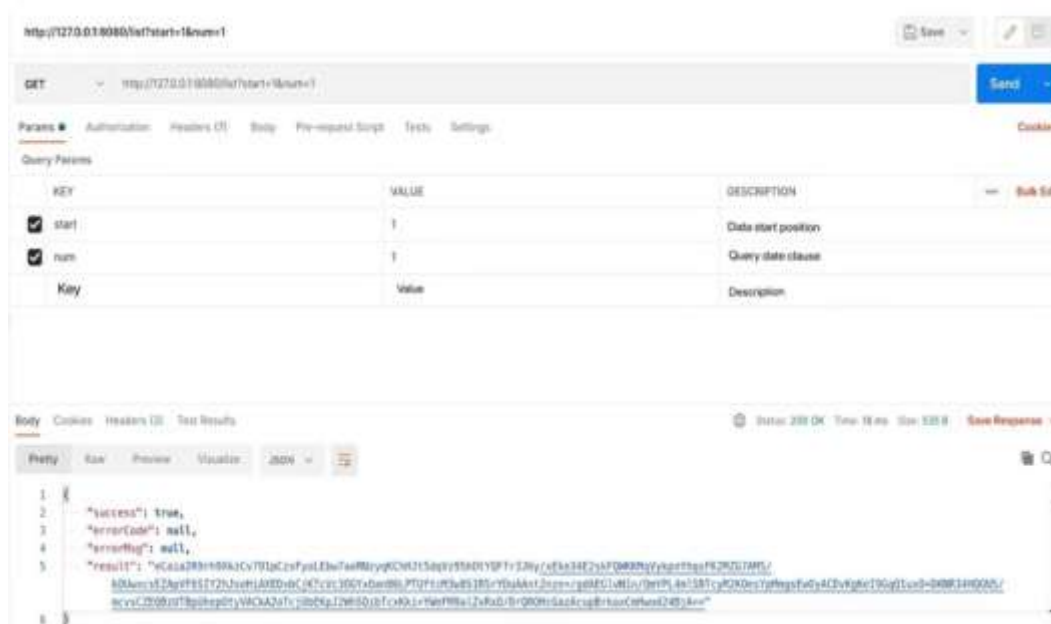
$$C^d = m \pmod{n} \quad (3)$$

That is, the remainder of  $c$  to the power of  $d$  except for  $n$  is  $m$ . However, since the encryption is performed in groups, that is, the number of bytes in each key length is a cipher text, so when decrypting, it is necessary to decrypt part by part, and then combine the results to get the final plaintext.

The decryption process must be consistent with the encryption process. The segmented encrypted cipher text needs to be decrypted into plaintext. Since the data length after RSA encryption is the same as the plaintext length, the decryption is done once every 128 bytes, and finally combine the results to get the plaintext.

#### IV. IMPLEMENTATION AND APPLICATION OF THE SECURE TRANSMISSION SYSTEM

After the data exchange platform is running normally, use Postman to perform a request test, and the returned result is shown in Fig 3.



**Fig 3: results of execution**

In the test, the JSON format returned by the data can be received normally, and it is shown that the data returned by the server has been encrypted, and the plaintext cannot be seen directly. This article deals with the simulation request processing of the entire process of data secure transmission. The client splices the request address according to its own needs, and the server queries the database according to the requirements to obtain data. The data is encrypted and spliced into a standard return format and returned to the client. Because the third party does not know the specific encryption protocol in the communication process, and does not know which encryption algorithm to use and the key used by the algorithm, it is quite difficult to crack the plaintext.

The data security transmission system uses Eclipse3.6 and jdk8.0 as the development technology, and the specific interface of the system is shown in Fig 4-7.



**Fig 4: home page of the system**

Data number	The index name	Index coding	state	Indicators of time	The numerical	unit
2982	Agricultural output refers to AOD0507 number (last		Has been sent	02013	97.5	%
2983	Agricultural output refers to AOD0507 number (last year = 100)		Has been sent	02012	94.9	%
2984	Agricultural Number (= 100 last year)	AOD0507	Has been sent	02011	103.5	%
2985	Agricultural output refers to AOD0507 number (last year = 100)		Has been sent	02010	95.0	%
2986	Forestry output value refers to the AOD0508 number (last		Has been sent	02019		%

Fig 5: public data transmission page

Data number	The title	The author name	content	state	create time	update time
73298965019	number of agricultural	Zeng Xia	According to the	Has been sent	02020-07-15	02020-07-15
2302087	According to the report		65010230206		15:59:10	15:59:10
73298965188	number of ationization		The report 7329892020-07-15, 02020-07-15		6518826065915: all	15:59:11

Fig.6 the ordinary data transmission page

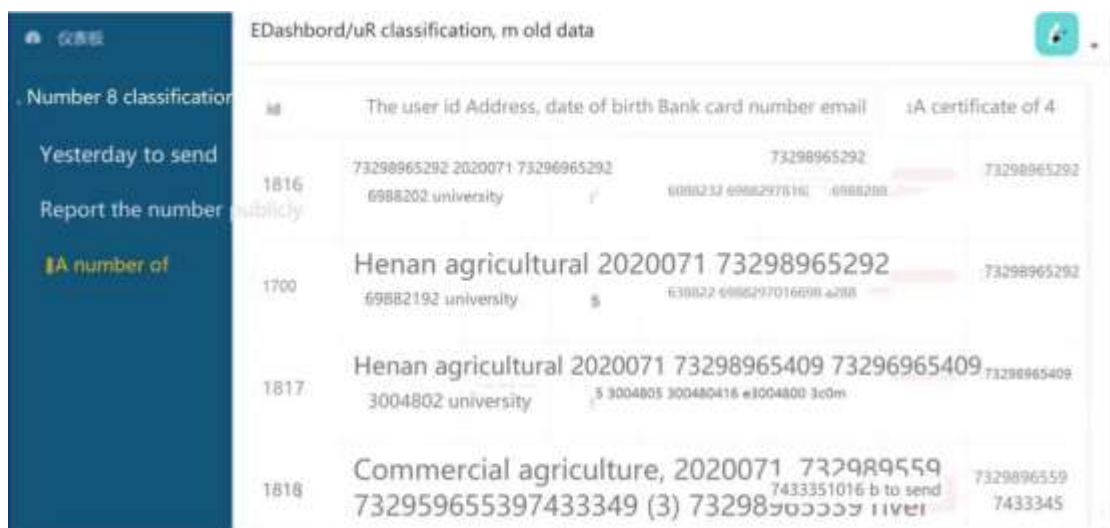


Fig 7: the private data transmission page

The system homepage in Fig 4 mainly displays modules such as registration, login, user management, and data transmission (public transmission, ordinary transmission, and private transmission). The system mainly implements functions such as agricultural data information management and data secure transmission.

For the data transmission page of Fig 5-7, data transmission is the kernel module of the system. After the data are uploaded, the data will be automatically identified according to the sensitivity of the data, the data type. Then the data will be estimated whether the data is private. If the data is not private, it is transmitted normally. If the data is private, it will be encrypted before being transmitted.

## V. CONCLUSION

In this paper, the data exchange between Henan Province Rural Comprehensive Information Service Platform - the Zhong yuan Information Port and other sites is taken as the research object. The agricultural data secure exchange system is realized with lightweight JSON format. During the transmission the data are divided into three categories: public data, ordinary data, and private data. For the secret data. This part of the work is the innovation of this paper. In the secure transmission system the valid information on the development of agricultural production is collected, the private agricultural data are encrypted by RSA, and JSON is selected as the data exchange format. The system is capable of normal operation, the test results in line with expectations, and achieved good results. It can be used in higher data security requirements of agricultural data exchange platform.



Although the data security transmission system designed in this paper can meet the security of agricultural data transmission, it still has some shortcomings. Since the data has to be processed before being encrypted, the transmission speed is too slow. Therefore, it is also necessary to continuously upgrade and reform the algorithms of the agricultural data security transmission system to provide more intelligent, efficient, and secure services for applications, so as to ensure the secure transmission of agricultural data.

## ACKNOWLEDGMENTS

The authors acknowledge the Henan Province Major Public Welfare Projects under Grant No. 201300210300, the Hnan Province Key Science-technology Research Project under Grant No.162102210109, National Science and Technology Resource Sharing Service Platform Project under Grant No. NCGRC-2020-57, and the Science-technology Research Project of Zhengzhou city under Grant No.141PPTGG431.

## REFERENCES

- [1] Wei W T (2010) The security of the internal network and its preventive measures. China Science and Technology Information, (18): 126-127.
- [2] (2007) GB/T22239-2008. Information Security Technology Information System Security Level Protection Basic Requirements. Beijing: China Standard Press.
- [3] Wei R L (2013) Computer network communications security data encryption technology research and application of: China University of Geosciences.
- [4] Gu F F (2009) Research and implementation of asymmetric authentication scheme for wireless local area network. Hefei University of Technology.
- [5] RL RiVest, Shamir A., Adleman L. (1978) A method for obtaining digital signaures and public key cryptosystems. Communications of the ACM, February, 21 (2): 120-126.
- [6] Wang L (2015) Application of XML in Web pages. Electronics and Software Engineering, (05): 14.
- [7] Chun S D, Wang Y F (2002) Design and Application of Network News release management system. Library and Information Technology, (5): 57-59.
- [8] Chen W, Jia Z P (2012) Research on using JSON to reduce XML data redundancy. Computer Applications and Software, 09: 188-190+206.
- [9] San Mao Space. Comparison of the difference between JSON and XML. 2013-06-16[2015-3-27].
- [10] Truică Ciprian-Octavian, Apostol Elena-Simona, Darmont Jérôme, Pedersen Torben Bach (2021) The Forgotten Document-Oriented Database Management Systems: An Overview and Benchmark of Native XML DODB MSes in Comparison with JSON DODB MSes. Big Data Research, 25 (prepublish).
- [11] Zhang L, Chen ZN, Yang SF (2021) Research on the application of crawler technology in machine learning. Journal of Physics: Conference Series, 1865(4)
- [12] Liang Y J, Wei T, Li R X (2021) Classification and hierarchical data encryption method in cloud storage integrating multi-features. Network Security Technology and Application, (02): 35-36.
- [13] T38667-2020, Information Technology Big Data Data Classification Guide.