Big Data-Based User Data Intelligent Encryption Method in Electronic Case System

Xin Liu(✉)

School of Railway Operation and Management,
Hunan Railway Professional Technology College, Zhuzhou, China
liuxin0011l00@163.com

Abstract. When the user data of the conventional electronic case system was encrypted, there was a shortage of low analysis accuracy. To this end, an intelligent encryption method for user data of the electronic case system based on big data was proposed. Introducing the big data technology, building a framework for intelligent encryption of user data of electronic case system, and realizing the construction of intelligent encryption of user data of electronic case system; Relying on the determination of the data intelligent encryption algorithm, the electronic case system model was embedded to realize the intelligent encryption of the user data of the electronic case system. The experimental data showed that the proposed big data modeling and analysis method was 61.64% more accurate than the conventional method, which was suitable for intelligent encryption of user data in electronic case system.

Keywords: Big data · Electronic case system · Data encryption

1 Introduction

Electronic Medical Record (EMR) is also called a computerized medical record system or Computer-Based Patient Record (CPR). It replaces handwritten paper cases with digitized patient medical records that are stored, managed, transmitted, and reproduced using electronic devices. It is digitally stored, managed, transmitted, and reproduced using electronic devices. Its content includes all information about paper cases [1] Electronic cases are also generated by the network management of hospital computer, the application of information storage media–discs and IC cards, and the globalization of the Internet. Electronic cases are also the inevitable outcome of information technology and network technology in the medical field, are an inevitable trend of modernization of hospital cases, and its preliminary application in the clinic has greatly improved the hospital's work efficiency and medical quality [2], but this is only the beginning of electronic case application. The electronic case system is a set of software and hardware systems that support electronic cases, including data collection and transmission, data storage and extraction, and data processing and display. It can realize the collection, processing, storage, transmission and service of patient information. With the advancement of medicine, medical technology and information technology, the traditional medical model has been challenged as never before. "Digital Hospital" is becoming an inevitable trend in the development of hospital information in the world,

G. Gui and L. Yun (Eds.): ADHIP 2019, LNICST 301, pp. 300–309, 2019.
https://doi.org/10.1007/978-3-030-36402-1_32

and the research and promotion of electronic case systems is one of the important aspects of the development of "digital" hospitals. Especially in hospitals [3], patients have more information during admission, including diagnostic data of clinicians, hospitalization data, family information of patients, and various medical expenses for patients. The amount of information is quite large. If you manage the storage with old paper files in the past, it will take time and effort to lose some information [4].

Electronic cases are the inevitable outcome of the widespread application of information technology and network technology in the medical field, and are also the inevitable trend of modern management of hospital medical records. Its intelligent encryption method can not be underestimated [5].

2 Electronic Case Encryption Technology Design

The electronic case system is medical-specific software. The hospital electronically records the patient's visit information through electronic cases, including: home page, disease record, inspection test results, medical orders, surgical records, nursing records, etc. There are both structured information, unstructured free text, and graphic image information. It involves the collection, storage, transmission, quality control, statistics and utilization of patient information [6].

"Structuralization" is accompanied by electronic cases, that is, "structuring" has been invented in order for computers to be able to handle cases. The core of the case is that the medical staff records the condition, symptoms, medication, and disposal contents of the patient during the diagnosis and treatment. However, due to the complexity of medical services and the differences among individual patients, this record naturally has the dual attributes of "normative" and "differential". Unfortunately, the nature of computers is good at dealing with "normative (structured)" data. For the analysis and utilization of unstructured data, it is limited to late-stage fields such as "full-text search" and "smart segmentation". For example, the library field, or the use of medical big data. However, as a business system application "unstructured" data processing, the computer is powerless. Therefore, the pursuit of "structured" cases is a balance between "human readable" and "computer readable" [7].

The so-called structured, semi-structured, and superior and inferior disputes are basically a pseudo-proposition. The structural maturity evaluation of the electronic case system needs to be carried out in a more sophisticated and multi-dimensional indicator system. Thats is: the background support capability of "node, structured, and coded", and the extent to which the relevant template data in the system achieves the above three capabilities.

2.1 Electronic Case Encryption Components

(1) Clinical documentation: Located at the top of the electronic case data structure, it is a collection of data on the clinical diagnosis and guidance intervention information of patients (or health care subjects) generated and recorded by specific medical service activities (health events). Such as: outpatient cases (emergency), inpatient medical records, consultation records, etc. [8].

(2) Document segment: Structured clinical documents are generally split into logical segments, i.e. document segments. The document segment provides a clinical context for the data that makes up the document segment, i.e. adding specific constraints to the generic definition of the data elements therein. Structured document segments typically consist of data sets and are specifically defined by data sets. Document segments are not explicitly defined in this standard, but the concept of document segments is implicit.
(3) Data group: It consists of several data elements, which constitutes a basic unit of clinical documents as a collection of data elements, and has the characteristics of clinical semantic integrity and reusability. Data groups can have nested structures, and larger data groups can contain smaller sub-data sets. Such as: document identification, complaints, medication, etc.
(4) Data element: Located at the bottom of the electronic case data structure, it is the smallest, non-subdividable data unit that is assigned by a series of attributes such as definition, identification, representation, and allowable values. The allowed values for data elements are defined by the value range. The structure of electronic case data is shown in Fig. 1.

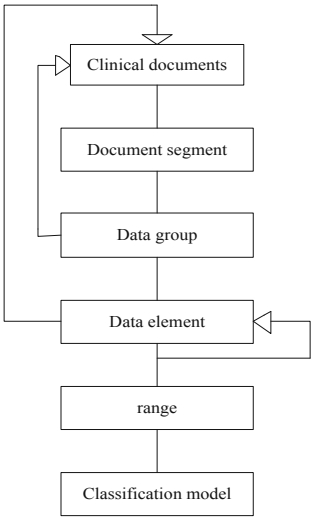

Fig. 1. Electronic case data structure

2.2 Framework for Electronic Case Encryption Technology

In the case editor, the user can see that the segment of the case is split into individual nodes. From the technical backend, each node is saved as a "value pair" so that the database can process the data further. Since 2002, the industry has developed a dedicated electronic case editor to handle nodes in case templates, so this technology threshold has been surpassed by most manufacturers today.

2.2.1 Nodeization of Electronic Case Encryption

A hierarchical relationship between time nodes is required. For example, the previous step "node", after implementation, we can quickly query in the background, for example, query the patient's case "blood pressure diastolic pressure: 90 mm Hg" such a value as reference. But in reality, the questions we need to answer are often: diastolic pressure before medication, diastolic blood pressure after medication, or the range of hypertension in a family history. Therefore, it is meaningless to just save "nodeization" and "value pairing" without knowing the context. The hierarchical relationship definition of case nodes can solve the problem of logical relationship between nodes. At this stage, adding and deleting nodes requires defining the application scenario (document template type, node chapter, etc.) of the node, and giving a unique ID number according to certain rules, so as to facilitate the reuse and query of the node project.

2.2.2 Encoding of Electronic Case Encryption

The definition of the definition of the data content within the node. From the technical background, it is the range of values and data types of node values; from the clinical business point of view, it is the standard dictionary code maintenance for diagnosis, inspection, operation, and medication. Engineers with experience in ICD-10 know that this is not only a question of data coding dictionary compilation, but also the unification of the description dimensions of the same objective subject under different clinical professions and different disease conditions. In real life, it is often necessary to compile and correspond to different coding rules according to the specific purpose of use. The terminology registration service in the hospital information platform can help solve the problem of partial coding and correspondence of coded data. However, in practical applications, it is impossible to achieve 100% data coding. It is often necessary to find an balance of the ratio of input and output between flexibility and normativeness.

2.2.3 Measurement of Electronic User Data

Since the intelligence level of the electronic medical record system is mainly described according to the number of disclosures of the user data by the attacker, the current measurement of the electronic user data is basically measured by the disclosure risk. Disclosure risk is defined as the probability that an attacker may disclose data and other related content provided by an electronic medical record system. Let S be the electronic user encrypted data, SK is defined as the electronic user encrypted data S that the attacker can disclose according to the related content K, then the electronic medical record system user data encryption intelligence degree can be described by the following formula:

$$r(\mathrm{s},\mathrm{k}) = \mathrm{P_r}(\mathrm{S_k}) \tag{1}$$

3 Construction of Intelligent Encryption of Electronic Case System

3.1 Symmetric Encryption Process

Symmetric key encryption is also called Secret Key Encryption, that is, both sides transmitting and receiving data must use the same/symmetric key to encrypt and decrypt plaintext. The most famous symmetric key encryption standard is the Data Encryption Standard (DES). DES is a block encryption algorithm that uses a 56-bit key to operate a 64-bit block of data. The basic flow of symmetric encryption is shown in Fig. 2.

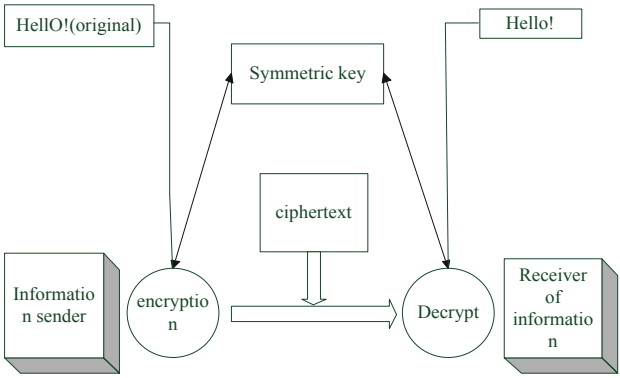

Fig. 2. Basic process of symmetric encryption

Commonly used algorithms in symmetric encryption algorithms are: DES, 3DES, TDEA, Blowfish, RC2, RC4, RC5, IDEA, SKIPJACK, AES, etc. RC5 is a relatively new algorithm, and Rivets designed a special implementation of RC5, so the RC5 algorithm has a word-oriented structure:

$$\mathrm{RC5} - \mathrm{w}/\mathrm{r}/\mathrm{b}, \tag{2}$$

where w is the word length and its value can be 16, 32 or 64. For different word length plaintext and ciphertext blocks, the packet length is 2w bits, r is the number of encryption rounds, and b is the key byte length. Since the RC5 packet length variable cipher algorithm is mainly processed for 64-bit packet w = 32 in this paper, the processing of RC5 encryption and decryption is described in detail below:

3.1.1 Create a Key Group

The RC5 algorithm uses 2r + 2 key-related 32-bit words for encryption: where r is the number of rounds encrypted. The process of creating this key group is very

complicated but straightforward. First copy the key bytes into the array L of 32-bit words (at this time, pay attention to whether the processor is little-endian or big-endian). The last word can be padded with zeros if needed. Then initialize the array S with the linear congruential generator modulo 2:

From I=1 to 2(r+1)-1 :

Where RC5 is a 32-bit group of 16-bit words,

P=0xb7e1 ; Q=0x9e37

For

32-bit words and 64-bit grouped RC5,

P=0xb7e15163 ; Q=0x9e3779b9

For 64-bit words and 128-bit packets, P=0xb7151628aed2a6b
Q=0x9e3779b97f4a7c15

Finally, mix L with S, the mixing process is as follows:

I=j=0 ;

A=B=0 ;

Processed 3n times

3.1.2 Encryption Processing

After the key group is created, the plaintext is encrypted. When encrypting, the plaintext packet is first divided into two 32-bit words: A and B (in the case that the processor byte order is little-endian, w = 32, the first plaintext byte enters the lowest byte of A, and the fourth plaintext byte enters the highest byte of A, the first The five plaintext bytes enter the lowest byte of B, and so on, where the operator <<< indicates the left shift of the loop, and the add operation is the modulus (which should be modulo, w = 32 in this article). The output ciphertext is the contents of registers A and B (Table 1).

Table 1. Symmetric encryption application mode.

Encryption mode (English name and shorthand)	Chinese name
Electronic Code Book (ECB)	Electronic cryptographic model
Cipher Block Chaining (CBC)	Cipher block link mode
Cipher Feedback Mode (CFB)	Encrypted feedback mode
Output Feedback Mode (OFB)	Output feedback mode

3.2 Asymmetric Key Encryption Process

Asymmetric key encryption, also known as Public Key Encryption (Pub1ic Key Encryption), was proposed by Professor Herman of Stanford University in 1977. Unlike symmetric encryption algorithms, asymmetric encryption algorithms require two keys: a public key (pub-lackey) and a private key (private key). The public key and the private key are a pair. If the data is encrypted with the public key, only the corresponding private key can be used for decryption; if the data is encrypted with the private key, only the corresponding public key can be used to decrypt. Since encryption and decryption use two different keys, this algorithm is called an asymmetric encryption algorithm. The basic flow of asymmetric encryption is shown in Fig. 3.

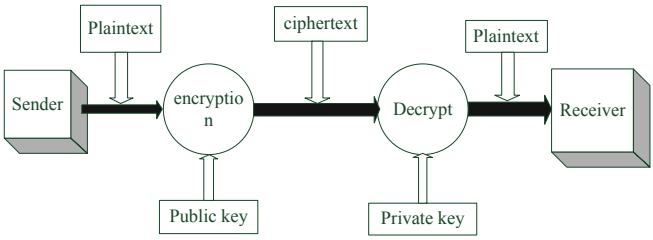

Fig. 3. Basic process of asymmetric encryption

The main algorithms used in asymmetric encryption are: RSA, Elgamal, knapsack algorithm, Rabin, D-H, ECC and so on. The RSA algorithm is the first algorithm that can be used for both encryption and digital signatures, and is also easy to understand and operate. RSA is the most widely studied key algorithm. From the time of its submission to the present, it has been tested by various attacks for more than 30 years. It is widely accepted as one of the best key scheme by 2017. The SET (Secure Electronic Transaction) protocol requires the CA to use a 2048-bit long key, and other entities use a 1024-bit key. The RSA key length increases rapidly as the level of security increases.

4 Big Data Electronic Case System User Data Intelligent Encryption Experiment

The simple process of electronic signature is as follows: when an electronic signature is required, the system will prompt the user to insert the memory card storing the individual into the card reader, and the user inserts and then returns. The system prompts the user to enter the password for reading the memory card information, and the user enters and presses Enter. The system reads the personal key and signs it, prompting the user to remove the memory card to prevent the key from being lost or leaked.

Assuming that User A wants to send a message m [1, p − 1] to B and sign the message m. The first step: User A selects an x [1, p − 1] as the secret key and calculates y = (mod p) as the key. The key y is stored in a public file. The second step: randomly select k [1, p − 1] and calculate r = (mod p). For the general Megamall type digital signature scheme, there is a Signature Equation: ax = b * k + c(mod(p − 1)).

Where (a, b, c) is a permutation of the mathematical combination of (h(m), r, s). The s can be solved by the signature equation. Then (m, (r, s)) is the digital signature of A to message m.

Step 3: A sends (m, (r, s)) to B.

4.1 Preparation of Experimental Data

See Table 2.

Table 2. The length of the key corresponding to the security level

Secrecy level	Symmetric key length (bit)	RSA key length (bit)	ECC key length (bit)	Secrecy years
80	80	1024	160	2010
112	112	2048	224	2030
128	128	3072	256	2040
192	192	7680	384	2080
256	256	15360	512	2120

4.2 Analysis of Experimental Results

The results of encryption analysis are shown in Fig. 4:

According to the test curve results, it can be concluded that asymmetric encryption has obvious advantages for big data electronic case system user data. The asymmetric encryption system does not require the communication parties to transfer the key in advance or has any agreement to complete the secure communication, and the key management is convenient, and the anti-counterfeiting and the repudiation can be realized. Therefore, it is more suitable for the confidential communication requirement in the network communication.

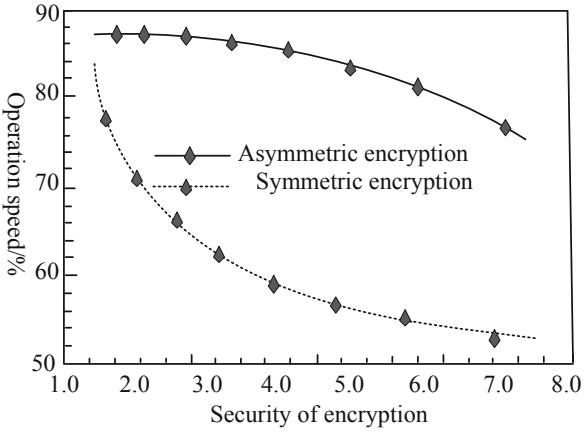

Fig. 4. Comparison of encryption results

5 Conclusions

Once the case is encrypted, the level of confidentiality will be greatly enhanced, and only authorized persons can log into the electronic case system. Moreover, all modifications and saves require identity authentication to be carried out. The details of any changes are also recorded by the computer and can be checked at any time. The purpose is to prevent tampering and prevent "deadbeat."

References

1. Han, W., Yan, H., Wang, Y.: Researchon hidden dangers and security secrets of classified electronic document. 20–23
2. Hu, J.: Analysis of the impact of big data on the development of hospital records management. Manage. Obs. (02), 108–110 (2017)
3. Yu, D.: Application of electronic medical record in hospital information management. Electron. Technol. Softw. Eng. (17), 259 2017
4. Li, A.: Problems and countermeasures of hospital information security. Network Secur. Technol. Appl. (11), 130–131 (2017)
5. Ye, J.: Application of data encryption technology in computer network communication security. Inf. Commun. (06), 173–175 (2018)
6. Wang, W., Shi, X., Liu, X., Xu, K., Yu, H.: Research on smart grid security technology based on symmetric key algorithm. Energy Environ. Prot. (12), 279–281 (2017)
7. Lei, H.Y.: Construction of hospital information resource management platform based on electronic medical record. Inf. Comput. (Theor. Version) (11), 221–224 (2018)
8. Han, T., Xie, J.: RSA encryption and decryption algorithm and related attack methods. Comput. Inf. Technol. (01), 31–32+36 (2018)
9. Liu, T.Y., Lin, K.J., Wu, H.C.: ECG data encryption then compression using singular value decomposition. IEEE J. Biomed. Health Inform. **22**(3), 707–713 (2018)

10. Bokhari, M.U., Shallal, Q.M., Tamandani, Y.K.: Reducing the required time and power for data encryption and decryption using K-NN machine learning. IETE J. Res. (15), 1–9 (2018)
11. Nasution, A.B., Efendi, S., Suwilo, S.: Image steganography in securing sound file using arithmetic coding algorithm, triple data encryption standard (3DES) and modified least significant bit (MLSB). J. Phys: Conf. Ser. **1007**(1), 012010 (2018)
12. Han, C., Yang, X., Hu, W.: Chaotic reconfigurable ZCMT precoder for OFDM data encryption and PAPR reduction. Opt. Commun. **405**, 12–16 (2017)
13. Rasmi, M., Alazzam, M.B., Alsmadi, M.K., et al.: Healthcare professionals' acceptance electronic health records system: critical literature review (Jordan case study). Int. J. Healthcare Manage. (3), 1–13 (2018)
14. Mohammed, A., Franke, K., Boakye, P.O., et al.: Feasibility of electronic health information and surveillance system (eHISS) for disease symptom monitoring: a case of rural Ghana. PLoS ONE **13**(5), e0197756 (2018)
15. Romeromuñiz, C., Nakata, A., Pou, P., et al.: High-accuracy large-scale DFT calculations using localized orbitals in complex electronic systems: the case of graphene-metal interfaces. **30**(50) (2018)