Algorithm of ensuring confidential data security of the cloud medical information system

L Babenko[1], *A* Shumilin[1,*], and *D* Alekseev[1]

[1]South Federal University, Institute of Computer Technology and Information Security, 2 Chekhov st., Taganrog, 347922, Rostov region, Russia

Abstract. The objectives of the study are to develop and assess the effectiveness of the structure of a cloud platform for storing, processing and organizing medical data, determining a method of protection, in particular, ensuring confidentiality when transferring and storing examination results. The proposed method for protecting a medical information system involves the use of an original DICOM file and subsequently a converted PNG image, which is subjected to a pixel encryption algorithm. An algorithm based on chaos theory is used to encrypt the image. The capabilities of chaos systems can significantly increase productivity. Hierarchical division of data streams into levels and standardization of data transmission protocols, as well as their storage formats, allow to form a universal, flexible and reliable medical information system. The proposed architecture has the ability to integrate into existing medical systems. In the course of the work, it was found that the considered protection method is an effective way to ensure the confidentiality of medical system data.

1 Introduction

In the age of universal informatization and the active development of information technologies, medical institutions in the course of performing diagnostic studies process and systematize significant amounts of data for the subsequent rehabilitation and treatment of patients. The effectiveness of the provided medical care is directly proportional to the efficiency and ease of use of this information by specialists of medical organizations. The presence of tasks related to the storage, systematization and processing of increasing volumes of data determines the relevance of the development and integration of medical information systems (MIS) into medical institutions. The ability to operate with data in electronic form ensures the promptness of the doctor receiving the necessary information about the patient, which increases the speed of decision-making about the diagnosis and treatment methods [1].

One of the topical directions in the development and implementation of storage systems, systematization and processing of medical data is the use of the capabilities of cloud services.

The main goal of implementing a cloud platform is to create a unified information space for collecting, storing and providing medical research results using a distributed team of

* Corresponding author: shumilin@neurotech.ru

qualified medical specialists. The category of medical research includes the results of medical research carried out using diagnostic equipment from various manufacturers.

The data obtained can be used by both medical institutions and research organizations. The patient can share the results of their own medical research with other users of the cloud platform or groups of qualified medical professionals. The data can be used by medical personnel who provide a range of services for their research, analysis or expertise, after which they provide research results.

Medical organizations by virtue of the law are the operators of the personal data of their patients. They are directly involved in the collection, systematization, accumulation, storage, clarification, updating, modification, distribution and destruction of such information.

One of the problems in the design of medical information systems is the need to integrate mechanisms for protecting confidential information. The category of confidential information includes: last name, first name, patronymic of the patient, month, date and place of birth, series and number of the passport, address of registration and actual residence, taxpayer identification number (TIN), insurance certificate of state pension insurance (SNILS), family, social position, education, profession, position, specialty, series and number of the medical insurance policy and its validity, etc. Due to the fact that this category of information is, as a rule, a text form, its protection is ensured by standard methods and means of encryption. The category of personal medical data that requires unconventional approaches to their protection includes the results of medical examinations of patients stored in the form of signals (for example, electroencephalogram signals).

Due to the fact that the requirements of the legislation establish the need to protect personal data, the key task in the implementation of a cloud storage system, systematization and processing of medical data is to ensure the security of stored information. As part of the work, the purpose of the research is to develop and evaluate the effectiveness of the general scheme of the cloud platform, which ensures the fulfillment of a certain range of tasks, as well as to choose a method to ensure the protection of medical data of patients, in particular, to ensure the protection of examination results stored in electronic form in the form of EEG signals. The aim of the work is to improve the efficiency of security systems (blocking malicious actions) when storing, organizing and transferring information in distributed medical systems with cloud architecture.

To achieve this goal within the framework of the work, it is necessary to solve the following tasks:

- analyze the existing models of information processes and structures in the subject area;
- study the features of the means of accumulating and processing medical data stored in electronic information systems for patient registration;
- to develop an architecture of a cloud platform of distributed data storage that allows interacting with various hardware systems for conducting medical examinations;
- develop an algorithm to ensure the safety of medical data stored in the cloud platform in electronic form in the form of initial physiological signals (EEG, ECG, EMG, EOG, etc.) recorded during patient examinations;
- create an integrated cloud platform for distributed storage, analysis and systematization of medical data and a security system using the developed protection method;
- analyze the effectiveness of the proposed algorithm for protecting confidential medical information in the context of integration into the developed cloud platform;
- to analyze the effectiveness of the proposed algorithm for protecting confidential medical information under conditions of integration into existing medical information systems operating on the basis of various architectures.

2 Analysis of the current state of research

In work [11] Kotyashichev I. A. and Byrylova E. A. consider the possibility of using cloud technologies to improve the efficiency of the implementation of information systems in various branches of medicine. Among the most common methods of ensuring data security, the authors highlight encryption. However, in the course of the work, an inherent problem of symmetric encryption systems is noted - the problem of key distribution, which complicates the process of working with such systems. The problem is that storing keys on a cloud server is impractical, since a user with access to cloud servers gains access to the key, and therefore to the decrypted data.

Kereytova M.R. and Malysh V.N. in [12] note the problem of ensuring information security of confidential data of patients as one of the most important in the creation and design of medical information systems. The issue of information protection is considered on the example of a distributed information system of the Department of Public Health of the Kemerovo Region, covering all medical and preventive institutions (LPI) of the Kemerovo Region. The authors propose an integrated approach to solving the problem: introduce control over workstations for unusually high activity, make full use of anti-virus protection, monitor all updates for existing operating systems, use multi-level user authentication involving the use of USB keys, smart cards, passwords, file keys. However, the approach proposed by the authors does not take into account the mechanisms for ensuring data protection in the aspect of preventing data leakage and / or unauthorized access when transferring and storing information in systems with client-server architecture. Thus, within the framework of this work, methods and means are considered that provide protection at the level of access to workstations for system users, while not taking into account.

Boychenko IV in work [13] notes the importance of the problem of realization of the rights of citizens in the field of protection of personal data of patients. The author considers the possibility of using medical information and analytical centers in the health care structure, focusing only on the legal and legal aspects of the problem. The preliminary analysis carried out by the author allows us to conclude about the great potential of using cloud technologies in solving the problems of modern healthcare. However, their widespread implementation requires a competent technical solution aimed at developing methods to ensure the security of transmitted information and the confidentiality of patients' personal data.

In [14], Rohan Jathanna notes the vulnerability of cloud systems to attacks from malefactors (DDoS attacks, attacks to penetrate the server, unauthorized access to databases). To prevent loss of access to confidential data, the author suggests using the capabilities of backup tools. Combating unauthorized access is achieved by using encryption algorithms. The approaches proposed by the author have significant drawbacks. The backup system requires a large amount of additional computing and memory resources, as well as providing a new object of protection (a resource with a backup copy). The efficiency of the used encryption algorithms decreases due to the presence of the problem of key distribution: it is necessary to provide for the possibility of transferring the key from the client to the server via a secure communication channel. The consequence of a compromised encryption key is the loss of access to confidential data.

In work [15] DA Krivosheev. highlights the main disadvantages of using asymmetric encryption systems in medical cloud platforms: high costs of computing resources, as well as the time it takes to implement computing processes. The author proposes an alternative approach to creating a symmetric encryption key based on the use of the patient's physiological signal as a “physiological” signature. A significant disadvantage of the proposed method is the fact that physiological signals (electrocardiogram, photoplethysmogram, electroencephalogram, etc.) can change during a person's life.

Accordingly, the encryption key generated earlier may become invalid after a certain time and, as a result, access to personal data will become impossible.

An equally important problem of the proposed method is the possibility of accessing data only from the side of their owner (a patient who provided a physiological signal to generate an encryption key). Thus, the possibility of gaining access to examination results by other persons (for example, the attending doctor, patient's relatives, health system analyst, etc.) is difficult or completely excluded.

Summing up, it is worth noting that in the works available in the public domain in scientific literature and electronic libraries, there are various disadvantages, the main of which are: the problem of key distribution, high requirements for computing resources, time and memory resources. The approach proposed within the framework of the current project is aimed at eliminating the above disadvantages by using homomorphic encryption systems, the key feature of which is the ability to process encrypted information without decrypting it.

3 Scheme of a cloud platform for storing and organizing medical data

To solve the problem of storing, organizing and processing medical data, a cloud platform has been developed.

The developed cloud system includes 4 main levels:

Data storage layer: a global data warehouse, which includes a database for storing initial data of examinations and reports, as well as anthropometric, diagnostic, demographic information about patients. The repository contains the full amount of information for research and training machine algorithms, but patient identification is possible only by a secure identifier.

The data consumer layer is a layer that includes systems that receive and process data from the Global Storage or transfer new data to it. This layer is linked to the storage layer through a standardized programming interface (Storage API). Data consumers can be: third-party medical information systems; research systems; data processing information system - contains a database of patients' personal data, complies with the safety and protection requirements of personal data and medical data (Federal Law of the Russian Federation of July 27, 2006 No. 152-FZ "On Personal Data"; Federal Law of 21.11.2011 No. 323-FZ "On the basics of health protection of citizens in the Russian Federation"; Health Insurance Portability and Accountability Act of 1996, HIPAA) [2]. This module provides interaction with end client applications using a distributed interface (REST API).

Application software level - a level containing software for end customers, where medical data are generated and / or displayed (examinations in the form of signals, reporting and personal data of a patient): Windows clients - software for Windows operating systems; Web server - provides the user with the ability to access through the web browser, in accordance with the roles assigned to this user; Mobile client - provides access to the data processing information system using mobile devices (Android, iOS).

Hardware Layer - Physical devices for conducting surveys. In general, there can be different types: electroencephalographs, cardiographs, biofeedback systems, wearable fitness trackers, etc.

4 Implementation of security mechanisms when transferring medical examination data via a cloud platform

In the course of the research, the MIS was developed, one of the mechanisms of which is to ensure the security of transmitted medical data. Information circulating in the system is divided into two types: text information (names of patients, passport data, etc.), the protection of which is achieved through standard encryption mechanisms (symmetric block encryption), as well as the results of medical examinations stored in the form of electroencephalographic signals ... To ensure the protection of the second category of data, an approach based on converting the original digital signals into the image format is proposed.

The developed MIS protection mechanism assumes the use of the original DICOM file and the PNG image file subject to the pixel encryption algorithm.

DICOM (Digital Imaging and Communications in Medicine) file is an object-oriented file with a tagged organization: patient → study → series → image (frame or series of frames) [3, 8].

The file contains structured information, including medical images for later saving as a PNG file and patient data as a text file.

It is intended to use MATLAB to extract medical images from a DICOM file. The JAVA programming language is used to execute MATLAB code and to programmatically implement a medical image encryption algorithm based on chaos theory, based on the traditional cryptographic architecture developed by Friedrich. This algorithm, applied to the resulting medical PNG image, will be executed pixel by pixel: for each pixel of the medical image. The encrypted image will then be uploaded via TCP / IP to the cloud, which will store the patient information (encrypted using a block encryption algorithm) and the encrypted image directly in the file. Steganography methods are used to keep the fact of encrypted information transmission secret.

Due to the complex structure of the DICOM file, as well as the content of heterogeneous information (text, signal image, hospital logo, type of medical imaging device) in it, its processing is a complex process.

DICOM file splitting:

Input data: DICOM file;

Imprint: medical image in .png format and medical text information.

Step 1. Reading a DICOM file;

Step 2. Separating the pixel data of the medical image and associated medical metainformation;

Step 3. Saving medical metainformation in a text file;

Step 4. Save the pixels of the medical image in .png format with 24-bit depth.

Thus, the medical image will be saved in PNG format in order to facilitate pixel processing during the encryption process.

To encrypt a medical image, an algorithm based on chaos theory is used, based on the traditional architecture of cryptography created by Friedrich [4,5]. This algorithm, applied to the resulting medical PNG image, will be executed pixel by pixel: for each pixel of the medical image.

In figure 1 shows an example of processing a medical image with an encryption algorithm.

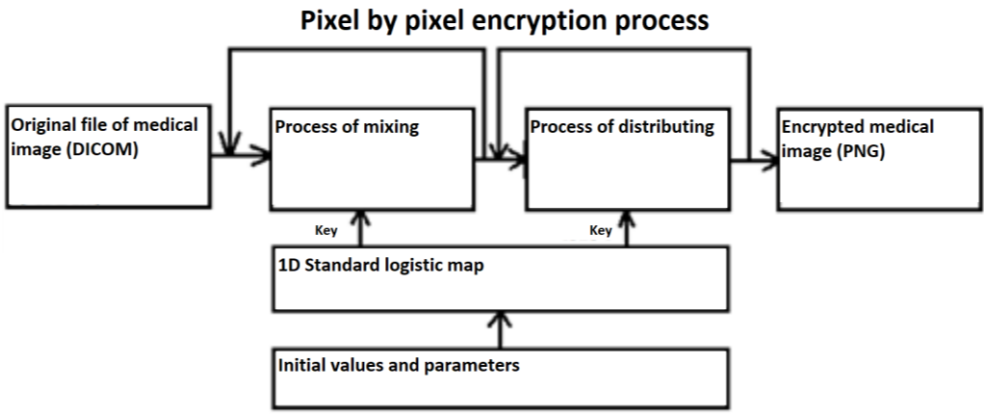

Fig. 1. Diagram of a medical image encryption process.

Pixel mixing means reorganizing the pixel arrangement of the original medical image; the purpose of the step is to reduce the high degree of correlation between neighboring pixels.

The next step is distribution, which refers to changing the pixel values of the medical image by performing some transformations on the pixel values. Therefore, adding new values to the pixels will increase the security of the encryption operation and cancel the correlation between the pixels, resulting in an encrypted image with a uniform histogram. The key generator in this process is one of the well-known one-dimensional maps of chaos theory known as the "1D Standard Logistic Map (SLM)" [7]. SLM has a variable X as an output, an initial condition X_n, and one control parameter μ, which give different results and properties when its value changes as input. Usually this map can be described as follows: $X_n + 1 = \mu X_n (1 - X_n)$ for n = 0, 1, 2, 3.

The experimental results of this map show a chaotic state of the system when $X_n \in [0; 1]$, the control parameter $\mu \in [0, 4]$. For greater accuracy, the logistic map is always chaotic and has an apposite Lyapunov exponent at $3.58 \leq \mu \leq 4$ [8]. Within the framework of the work, SLM is used as a key generator for mixing and distributing pixels of a medical image in a spatial domain, where the use of SLM is repeated for all image pixels to obtain arbitrary values that will be used to encrypt the pixel [9, 10].

Medical image encryption:

Input data: PNG file of a medical image;

Output data: encrypted medical PNG file.

Step 1. Reading the medical image and saving it into a 2-dimensional array of pixels;

Step 2. Using a standard logistic map as a random key generator, its initial state and control parameter as a secret image encryption key;

Step 3. Shuffling the image pixels (permutation of the pixel position) depending on the generated values from the SLM;

Step 4. Distribution of image pixels by changing their values depending on the key generated by SLM;

Step 5. Saving the value of the private key in the same text file that stores the medical metainformation obtained from the section of the DICOM file.

5 Conclusion

Hierarchical division of data streams into levels, standardization of data transfer protocols and formats for their storage ensure the creation of a universal, flexible and reliable medical information system. The developed architecture allows for quick integration into existing

medical systems. A single storage space makes it possible to explore a significant array of classified medical information using machine learning.

The developed MIS protection mechanism assumes the use of the original DICOM file and the PNG image file subject to the pixel encryption algorithm. A chaos-based algorithm is used to encrypt the medical image, based on the traditional cryptography architecture created by Friedrich. This algorithm, applied to the resulting medical PNG image, is performed pixel by pixel: for each pixel of the medical image.

The capabilities of chaos systems that are used to encrypt medical images can significantly improve performance by meeting the demands of digital imaging. Application of the proposed encryption mechanism for medical data is an effective way to protect information in a cloud platform.

Acknowledgements

This work was supported by the RFBR grant No. 20-37-90138 Postgraduates.

References

1. Mitkina P A 2017 *Modern scientific research and innovations* **5** http://web.snauka.ru/issues/2017/05/82546
2. *Health Insurance Portability and Accountability Act* https://en.wikipedia.org/wiki/Health_Insurance_Portability_and_Accountability_Act
3. *DICOM* https://ru.wikipedia.org/wiki/DICOM
4. Li-Chin Huangc L-Y T a M-S H 2013 *The Journals of systems and software* **86** 716-727
5. Jessica Fridrich M G a R D *Detecting LSB Steganography in Color and Gray-Scale Images* (Binghamton)
6. Fatma N A H A-C, Elgamal E-Z A 2013 *International Journal of Advanced Computer Science and Applications (IJACSA)* **4** 130-138
7. Digvijay Singh Chauhan A B K R G a J P S, 2017 *40th International Conference on Telecommunications and Signal Processing (TSP)*
8. *Logistic map* https://en.wikipedia.org/wiki/Logistic_map
9. Alsalmany A 2018 IOSR *Journal of Computer Engineering (IOSR-JCE)* **20(1)** 65-75 https://www.researchgate.net/publication/332571801_Cloud_System_For_Encryption_And_Authentication_Medical_Images
10. Plotnikov A V, Prilutskiy D A, Selishchev S V *DICOM standard in computer medical technologies* https://mks.ru/library/article/1997/dicom.html
11. *Visual cryptography* http://cryptowiki.net/index.php?title=%D0%92%D0%B8%D0%B7%D1%83%D0%B0%D0%BB%D1%8C%D0%BD%D0 % B0% D1% 8F_% D0% BA% D1% 80% D0% B8% D0% BF% D1% 82% D0% BE% D0% B3% D1% 80% D0% B0% D1% 84% D0% B8 % D1% 8F
12. Kotyashichev I A 2015 *Young scientist* **6.4(86.4)** 30-34 https://moluch.ru/archive/86/16357/
13. Kereytova M R, Malysh V N 2012 *NIK* https://cyberleninka.ru/article/n/informatsionnaya-bezopasnost-v-meditsinskih-informatsionnyh-sistemah

14. Boychenko I V 2011 *Doctor and information technologies* **3** https://cyberleninka.ru/article/n/postroenie-it-infrastruktury-zdravoohraneniya-na-osnove-paradigmy-oblachnyh-vychisleniy
15. Rohan Jathanna 2017 *Int. Journal of Engineering Research and Application* **7.6.5** 31-38 www.ijera.com ISSN: 2248-9622
16. Krivosheeva D 2016 *Legal informatics* **3** https://cyberleninka.ru/article/n/model-ugroz-bezopasnosti-v-sistemah-distantsionnogo-monitoringa-sostoyaniya-cheloveka