Trust-based Access Control Model for Healthcare Big Data

Rui Liu

School of Logistics and Management Engineering, Yunnan University of Finance and Economics, Kunming 650000, China

Abstract

With the rapid development of the era of big data, it has brought great impact to many fields, and all walks of life are constantly using big data, and the medical industry is one of them. People are using information technology to change the original traditional medical model, and medical big data is gradually being taken seriously by the country and the people. Although this development trend has brought great convenience to the medical industry, it also brings many problems due to its own characteristics, among which the most concerned issue is the leakage of medical data. Therefore, this paper constructs a trust-based access control model for medical big data, quantifies the trust of doctors through their access behaviors, and controls their access behaviors based on trust. Experiments prove that the access control model in this paper works better than the traditional access control model.

Keywords

Healthcare Big Data; Privacy Breaches; Trust; Access Control.

1. Introduction

Medical big data is a branch of big data in the medical field, which refers to the data related to life, health and medical care generated in human health-related activities, mainly from clinical-type data such as electronic health records. Through these massive medical data, it can effectively improve the efficiency of diagnosis and treatment. In terms of medical data, individual users have become an important source of data, and medical privacy information often means negative information such as untold pains and painful experiences to individuals, and the leakage of medical privacy information has become a huge hidden danger in the era of big data. In the past, patients mostly maintained their personality and dignity through self-forgetfulness and privacy of medical institutions. Nowadays, the ubiquitous smart electronic devices and cloud storage and cloud computing functions are like putting users in a transparent glass room, where our every move can be recorded, and the electronic health records generated by the widely used medical devices are making patients' privacy invisible.

Trustwave released a 2015 Healthcare Industry Security Report that surveyed 398 professional healthcare professionals (some technical, including CIOs, CISOs, IT directors, etc., and others general healthcare professionals) and found that 91 percent of survey respondents believe that cyberattacks against the healthcare industry are increasing, yet less than 10 percent of the budget is spent on protecting sensitive patient information. However, less than 10% of the budget is spent on protecting sensitive patient information. By stealing medical data, criminals can easily learn patients' names, home addresses, contact information, test reports, diagnoses, and even health insurance to falsify information for fraudulent purposes or to purchase medical devices. The consequences of data theft in the medical industry are so serious that more than 2 million people in the United States fall victim to it every year, causing losses of up to $13,500 and taking hundreds of hours to solve the problem. 2015 social security system has become the most affected area of personal information leakage, etc. These incidents have seriously violated the privacy and legal rights of users. Nowadays, the public and the government are beginning to pay attention to personal privacy issues. In the existing medical information system, hospitals do not restrict the access rights of doctors, and doctors can access any patient's information, which makes the patient's privacy data very easy to be leaked. Therefore, how to secure the massive data and resources from attacks is the most important issue. In this paper, a trust-based access control model is constructed to address this problem, and trust is quantified based on the user's historical access records on HIS (Hospital Information System), and the corresponding access rights are assigned to the user by the trust amount to control the access behavior of doctors.

The rest of this paper is organized as follows. Part 2 introduces the current status of domestic and international research on access control, and analyzes the research progress and shortcomings of access control models. Part 3 details the construction of a trust-based access control model for medical big data. Section 4 describes the performance analysis of the access control model proposed in this paper. Section 5 concludes the paper.

2. Related Work

With the continuous development of information technology, one of the most basic and important key technologies that can meet the security needs of big data is access control technology. Access control is used to manage access to system resources, and is an important measure to protect information system resources. Access control can restrict users' access to key resources and prevent illegal users from entering the system and illegal access to system resources by legitimate users. Through authentication and authorization, the access control policy can ensure the real identity of users and guarantee that they have the corresponding authority to access resources. The purpose of access control is to prevent unauthorized access to information resources and unauthorized use of information resources. It allows users to access the information repositories they commonly use with appropriate permissions and restricts the behavior or operation of users to delete, modify or copy information files at will. In short, access control techniques are mainly used to determine which users access which information with which permissions, while ensuring the privacy and security of data [1]. Traditional access control mainly includes Discretionary Access Control (DAC), Mandatory Access Control (MAC) and Role Based Access Control (RBAC). Traditional access control is difficult to adapt to the complex and changing healthcare environment due to its limitations such as static and coarse-grained. To solve the above problems, the literature [2-4] introduced the concept of trust into access control by evaluating the trust value of users to grant them the appropriate access rights [2-4].

Trust is a multidisciplinary concept with different definitions in different fields such as sociology, psychology, economics, management, computer science, etc. Currently, there is no widely accepted and uniform definition of trust. Gambetta [5] considers trust as the subjective probability that an agent assesses that another agent or a group of agents will perform a particular behavior. Abual- Rahman [6] et al. consider trust as the assessor's expectation of the likelihood of a particular behavior of the appraisee, depending on the assessor's own experience and constantly revised as the behavior of the appraisee changes. Olmedilla [7] considers trust as a calculation of the trust of one entity A in another entity B based on the behavioral performance of another entity B regarding a particular service in a specific phase and in a specific context. Emphasis is placed on the environment, service domain and computability of trust. By understanding the concept of trust, many scholars have conducted research related to trust-based access control.

Peng Zhang [8] et al. proposed a trust-based dynamic multi-level access control model to solve the system permission scalability problem. Shi Ke [9] calculate the information person of both cloud users and cloud service providers based on their behavioral evidence, forming a two-way access control model based on behavioral trust in the cloud computing environment, which is effective for cloud computing access control control. Deng Sanjun [10] et al. introduce user trust analysis in the role-based access control model, dynamically adjust user access rights through user trust changes, and use blockchain to store access control policies to ensure the security of policy information . Fan Yundong [11] et al. introduced network analysis in the trust calculation process to achieve objective assignment of user behavior factor weights to improve the reliability and security of the model, and added a decay factor to ensure dynamic changes in the integrated trust value of user behavior to achieve dynamic and secure access control in the cloud environment. Lin Guoyuan et al [12] proposed a mutual trust-based access control model. Both the behavioral trust of users and the trustworthiness of cloud service nodes are considered. Only trusted users have access, while users can choose the most trusted cloud service nodes. Yichen Hou [13] et al. propose a data security-enhanced fine-grained access control mechanism for data security in mobile edge computing. Firstly, a dynamic fine-grained trusted user grouping scheme based on attribute and meta-graph theory is designed and combined with a role-based access control model to assign roles to users and grant privileges based on user group trustworthiness. Then user authentication based on attribute matching is used to further verify whether users are allowed to perform access operations.

In summary, it is found that with the continuous acceleration of medical informatization, data privacy protection technology is becoming more and more mature, but access control-based technology in the medical field still needs to be studied, and the literature combining trust and access control technology is even more scarce, and a more mature theoretical system has not yet been formed.

3. Trust Quantification

In this paper, we quantify the trust value of physicians in terms of disease similarity, success rate, and operational behavior of visits through their history of access.

(1) Disease similarity

When a patient is seen, the doctor mainly uses the patient's examination information and certain related medical records (e.g., medical data of other similarly diagnosed patients in the database, etc.) to arrive at a final diagnosis for that patient. There is a risk of compromise when physicians view patient data in the HIS database, so this paper calculates the similarity of patient diseases to similar patient diseases in the accessed case base to measure the trust value of physician access to medical data.

In this paper, we introduce the International Classification of Diseases (ICD) as the code used in the electronic medical record for physician diagnosis. Let a disease in the electronic medical record be represented by the ICD code and then written in the element group as $a_1, a_2, a_3, \dots, a_n$, the physician's access target and the requested access record are labeled according to ICD-10. Usually, a physician will request access to multiple cases because of the same access target, and the requested cases may not be the same for different physicians even though the access target is the same. Therefore, in this paper, we use cosine similarity to calculate the similarity between diseases for measuring the trust of physicians with the following equation [14].

$$DS = \frac{\sum_{i=1}^{n} a_i \times b_i}{\sqrt{\sum_{i=1}^{n} (a_i)^2} \times \sqrt{\sum_{i=1}^{n} (b_i)^2}} \qquad (1)$$

(2) Success rate

In this paper, the visit success rate is used as one of the trust attributes of doctors, and the trustworthiness of doctors is measured by the visit success rate over a period of time. The formula is as follows.

$$SR = \frac{N_s}{N_s + N_f} \qquad (2)$$

Where SR denotes the visit success rate, N_S denotes the number of successful visits requested by doctors, and N_f denotes the number of failed visits requested by doctors. The more successful visits requested by doctors, the higher the visit success rate, and the higher the trustworthiness of doctors.

(3) Operation behavior

There are four main types of doctor's operations on medical records, including viewing patient's medical records, copying electronic medical records, adding medical records and deleting related operations. The viewing, copying, adding and deleting of a doctor's own previous medical records or medical records of other doctors through work objectives may lead to the leakage of patient's privacy. However, the probability that a physician's copying, adding and deleting medical records of patients under the supervision of other physicians generates risk is greater than the probability that a physician's adding medical records of his or her own patients generates risk, so the risk of a physician's manipulation of medical records is expressed as a ratio between the number of times a physician copies, adds and deletes medical records of other physicians and the total number of times a physician copies, adds and deletes medical records.

$$OB = 1 - \frac{\sum_{i \neq j} C_i^j + \sum_{i \neq j} A_i^j + \sum_{i \neq j} D_i^j}{\sum_{j=1}^{K} C_i^j + \sum_{j=1}^{K} A_i^j + \sum_{j=1}^{K} D_i^j} \qquad (3)$$

where $\sum_{i \neq j} C_i^j$denotes the number of times doctor i copied medical records of other doctors, $\sum_{i \neq j} U_i^j$ denotes the number of times doctor i added medical records of other doctors, and $\sum_{i \neq j} D_i^j$ denotes the number of times doctor i deleted medical records of other doctors; $\sum_{j=1}^{k} C_i^j$ denotes the total number of times doctor i copied medical records, $\sum_{j=1}^{k} U_i^j$ denotes the total number of times doctor i added medical records, and $\sum_{j=1}^{k} D_i^j$ denotes the total number of times doctor i deleted medical records.

(4) Total Trust Calculation

Using the trust values already calculated for each trust factor and the entropy weighting method [15] to find the weights λ = {λ1, λ2, λ3}, the total trust AT of the physicians is as follows.

$$TT = \lambda_1 DS + \lambda_2 SR + \lambda_3 OB \qquad (4)$$

4. Access Control Solution

The role of the access control function is to decide whether to allow doctors to access medical records, and after a series of calculations and final diagnosis, doctors who meet the conditions are allowed access and those who do not are denied access. In this paper, we get the doctor's access request trust by trust quantification, which is stored in the doctor trust value database in HIS, read the doctor trust value from the database according to the request to meet a specific time, then calculate the doctor trust value according to the read out doctor trust value, use the trust threshold to derive the doctor trust value for this access request, and finally judge whether to allow the doctor this time according to the doctor trust value and the access control function. The final decision is whether to allow the doctor's access request based on the doctor's trust value and the access control function. The access control model in this paper is set to periodically calculate the trust threshold, which is determined by the average trust value of all physicians over a period of time, as follows[16].

$$\delta(t) = \frac{\sum_{i=1}^{n} TT_{d_i}}{Num(t)} \tag{5}$$

where C(t) denotes the number of all physicians requesting access in time t. Let δ_{di} = TT(d_i, t_j) - δ(t) and the trust control function be:

$$T-Control(d_{\mathrm{i}}) = \begin{cases} 1, & \sigma(d_{\mathrm{i}})>0 \\ 0, & \sigma(d_{\mathrm{i}}) \le 0 \end{cases} \tag{6}$$

5. Simulation Experiment

The purpose of the experimental setup is to verify the validity of the model proposed in this paper, i.e., whether the model proposed in this paper can calculate the trust value of doctors based on historical access records, and to verify whether the model proposed in this paper can control doctors' over-access. Partly, the doctors who may access the electronic medical records outside their own scope of work are called over-access doctors, and the doctors who only access the electronic medical records within their own scope of work are called normal-access doctors.

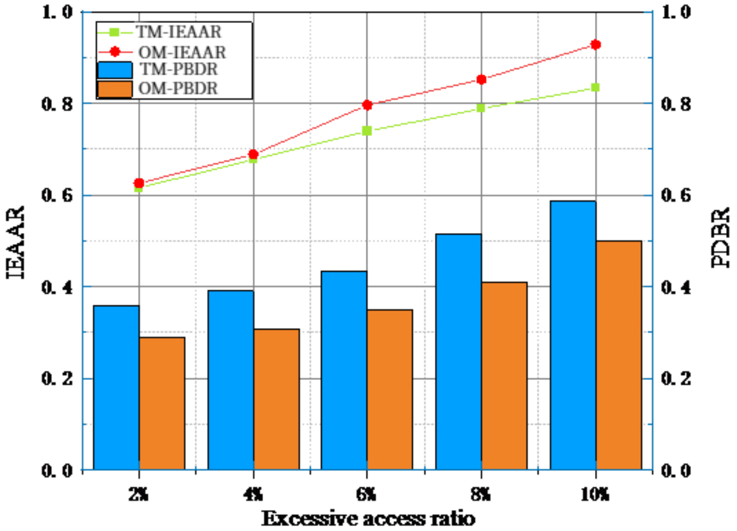

Figure 1. IEAAR and PDBR

In this section, the performance of the access control model proposed in this paper is compared with the traditional access control model. The model is tested by taking 2000 physician users and setting the percentage of excessive access to 2%, 4%, 6%, 8% and 10%, respectively, to observe the model control effect in terms of the correct rate of identifying excessive access behavior and the risk of privacy data leakage.

As in Figure 1, IEAAR indicates the correct rate of identifying excessive access for the access control model; PDBR indicates the privacy leakage risk for the access control model. With the increase of over-access proportion, both IEAAR and PDBR show an increasing trend, but the IEAAR of the model in this paper is higher than that of the traditional model, and the PDBR is lower than that of the traditional access model. The experiment fully proves that the access control model proposed in this paper can effectively identify the over-access behavior of doctors, and at the same time can effectively reduce the risk of privacy leakage of the system.

References

[1] Li FH, Su Q, Shi GZ, et al. Access control model research progress and development trend [J]. Journal of Electronics, 2012, 40(04): 805-13.

[2] Lin G., Wang D., Bie Y., et al. MTBAC: A mutual trust based access control model in Cloud computing[J].China Communications,2014,11(4):154-162.

[3] Takalkar Vedashree, Mahalle Parikshit. Trust-Based Access Control in Multi-role Environment of Online Social Networks[J]. Wireless Personal Communications,2018,100.

[4] Liu, Yuan-Bing, Zhang, Wen-Fang, Wang, Xiao-Min. Access control scheme based on multi-attribute fuzzy trust evaluation in cloud manufacturing environment[J]. Computer Integrated Manufacturing Systems,2018,24(02):321-330.

[5] Gambetta D. Can we trust trust? [J]. Can We Trust Trust?, 1988: 213-37.

[6] Abdul-Rahman A, Hailes S. A distributed trust model; proceedings of the NSPW '97, F, 1998 [C].

[7] Olmedilla D, Rana O, Matthews B, et al. Security and Trust Issues in Semantic Grids [J]. 2005.

[8] Zhang, Peng, Zhou, Liang. Trust-based dynamic multi-level access control model[J]. Computers and Modernization, 2019(07):116-121+126.

[9] Shi Ke. A two-way access control model based on behavioral trust [D]. Nanjing University of Posts and Telecommunications,2018.

[10] Deng Sanjun, Yuan Lingyun, Sun Limei. Research on access control model of Internet of Things based on trust degree [J]. Computer Engineering and Design,2022,43(11):3030-3036.

[11] Fan Yundong, Wu Xiaoping, Shi Xiong. Research on cloud computing access control model based on trust value evaluation[J]. Information Network Security,2016(07):71-77.

[12] Lin G., Wang D., Bie Y., et al. MTBAC: A mutual trust-based access control model in Cloud computing [J]. China Communications,2014,11(4):154-162.

[13] Hou Y, Garg S, Lin H, et al. A Data Security Enhanced Access Control Mechanism in Mobile Edge Computing [J]. IEEE Access, 2020, PP: 1-15.

[14] Mao Yimin, Guo Binbin, Yi Mianbing, Chen Zhigang. Parallel SVM algorithm based on relative entropy and cosine similarity[J/OL]. Computer Integrated Manufacturing Systems:1-26 [2023-01-17].

[15] Sun Yuzhong, Cheng Zhiyang. Research on financial performance evaluation of small and medium enterprises based on entropy method[J]. Mall Modernization,2022(23):180-182.

[16] Hui Zhen, Li Hao, Zhang Min, Feng Dengguo. A risk-adaptive access control model for medical big data [J]. Journal of Communication,2015,36(12):190-199.