Research Article

Healthcare Big Data Privacy Protection Model Based on Risk-Adaptive Access Control

Rong Jiang,1,2,3 Shanshan Han,1,2,3,4 Mingyue Shi,1,2,3,4 Tilei Gao,4 and Xusheng Zhao1,2,3,4

1Institute of Intelligence Applications, Yunnan University of Finance and Economics, Kunming, China
2Key Laboratory of Service Computing and Safety Management of Yunnan Provincial Universities, Kunming, China
3Kunming Key Laboratory of Information Economy & Information Management, Kunming, China
4School of Information, Yunnan University of Finance and Economics, Kunming, China

Correspondence should be addressed to Tilei Gao: gtllei@ynufe.edu.cn

Received 20 December 2021; Revised 18 February 2022; Accepted 19 February 2022; Published 15 March 2022

Academic Editor: Honghao Gao

Copyright © 2022 Rong Jiang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Edge computing is playing an increasingly important role in the field of health care. Edge computing provides high-quality personalized services to patients based on user and device data information. However, edge nodes will collect a large amount of sensitive patient information, and patients will also bear the risk of privacy disclosure while enjoying personalized services. How to reduce the risk of privacy disclosure while ensuring that patients enjoy personalized services brought by edge computing is the research content of this paper. In this paper, the work flow and management mode of Hospital Information System (HIS) are investigated on the spot, and the risk-adaptive access control model based on entropy is established. First, we use International Classification of Diseases, Tenth Revision (ICD-10) to mark the information resources accessed by users and use information entropy to measure the correlation “α” between medical information accessed by users and work tasks. Finally, we analyze the relationship between correlation “α” and risk through an example. The results show that users with high correlation “α” have low risk of access behavior, and users with low risk have high correlation “α” of access information resources and work goals. This discovery can help managers predict users’ access behavior in the Big Data environment, so as to dynamically formulate access control policies according to the actual access situation of users and then realize the privacy protection of medical big health data.

1. Introduction

Edge computing is used to extend cloud computing to the edge of the Web. Specifically, edge computing is a new distributed computing mode, in which multiple edge nodes located between cloud servers and local users cooperate to complete outsourced storage and computing tasks [1]. Edge computing is increasingly used in various areas of people’s life, including smart home, healthcare, industrial production, and media entertainment [2]. Especially in healthcare, medical informatization based on edge computing technology can greatly improve medical efficiency. At present, the research of edge computing combined with medical scenarios mainly focuses on the design of network communication protocol and routing algorithm but neglects the research of information security and privacy protection in medical scenarios. As a result, the development of current medical application scenarios based on edge computing technology is blocked. As medical Big Data contains a wealth of patient-sensitive information, the collection, transmission, use, and sharing of data bring great security risks to people.

Medical Big Data is the core asset of the medical field. However, most hospitals lack special privacy protection measures, and the medical industry has become one of the most serious areas of privacy leakage [3]. In 2016, 10 industries with serious data leakage were released in the Internet Security Threat Report. The medical industry ranked first, with 116 data leakage incidents, and the incident incidence rate was 37.2%. The proportion of data leakage is much higher than the second-ranked retail industry. In 2017, 47 GB of medical information was leaked by Amazon server...
and the private information of 150,000 patients was exposed, which brought great negative impact to both the society and individuals. Therefore, it is very necessary to study the security and privacy issues in the medical industry.

At present, the research on data security and privacy protection technology mainly includes access control, data anonymization, data encryption, differential privacy protection, digital watermarking, and so on. Among them, access control technology has become a hotspot of current research, but it mainly targets at the field of operating system, and there are not many researches in the information field, especially the research on the security and privacy protection of medical Big Data. Inspired by previous studies [4–6], this paper proposes a medical Big Data privacy protection model based on risk-adaptive access control. The main contributions are as follows:

(i) We build a set of diagnostic codes that users can access under specific work goals. Based on the ICD-10 diagnostic codes, we cluster the set of diagnostic codes that allow users to access under the target disease.

(ii) We design a method to evaluate user access behavior. According to the user’s historical access data, we use information entropy to evaluate the risk of users’ current request access behavior.

(iii) We propose a new method for calculating risk benchmarks. Based on the user’s information entropy, we use K-means clustering to calculate the baseline value of risk assessment.

The rest of the paper is organized as follows. Section 2 introduces the related work about medical data privacy protection. Section 3 specifically introduces the construction of the medical Big Data privacy protection model. Section 4 is the experimental simulation. Section 5 summarizes the paper and discusses directions of future work.

2. Related Work

Although edge computing improves medical efficiency, medical data may leak and cause serious damage to patients in the process of transmission and access. Therefore, the security and privacy protection of mobile medical information are particularly important. It is urgent to establish privacy security guarantee mechanism and monitoring mechanism to ensure the safe reading and access of medical data. The current research on Big Data security and privacy protection mainly adopts technologies based on cryptography, differential privacy, anonymization, identity authentication, and so on. Gao et al. [7] proposed the Sensitive Data Timed I/O Automata (SDTIOA) model, which was an automatic transformation method for modeling the timed privacy requirements of IoT service compositions. Gao et al. [7] used the SDTIOA model to verify whether the service combination meets user privacy requirements, which can effectively prevent the leakage of user privacy information. In literature [8–19], scholars used differential privacy technology to establish some privacy protection models for medical Big Data security issues. Zhang et al. [20] used encryption technology to study the security of medical data according to the sequence of events; He et al. [21] protected medical information through anonymization and identity authentication. In order to protect the privacy of medical data, Li and Zhang [22] chose to desensitize and anonymize their EMR data before authorizing a third party to use it. Chen et al. [23] proposed an electronic medical record system based on blockchain joint proxy re-encryption, which ensured the security of medical data access and realized fine-grained access to data through attribute-based access control. Xanthidis and Xanthidou [24] designed an error-correction code hash function and constructed a privacy-preserving anonymization algorithm, which to some extent controlled users’ access rights and ensured the safe sharing of data between doctors and patients. Some scholars used VPN, SSL and other technologies to control access to medical data, so as to protect the security of medical data. Malasri and Wang [25] proposed SNAP (sensor network for assessment of patients) scheme to solve the safety problems of wireless sensor monitoring network. In Sun et al.’s study [26], patients’ physiological signals (such as blood pressure and heart rate) were used to generate symmetric keys with patient characteristics, so as to protect the data security of patients. However, these schemes are time-consuming and not practical for medical scenarios requiring high delay.

In recent years, scholars at home and abroad have also carried out research on risk-based access control technology. In Jason et al.’s [27], the author first put forward the concept of risk and gave the primary colors and suggestions related to risk quantification. Dankar and Badji [28] proposed a risk-aware information disclosure model for biomedical data. Model evaluated the risk posed by a data request using all contextual information surrounding the request and feed it into an access control decision module. Ni et al. [29] and Cheng et al. [30] presented specific risk quantification methods according to the safety marks and sensitivity of the subject and object. In addition, a role-based access control (RBAC) model based on risk perception was proposed, which mainly evaluated the trust of users, the relationship between users and roles, and the relationship between roles and permissions. Diep et al. [31] and Sharma et al. [32] mainly conducted risk assessment on the user’s access behavior, and the evaluation basis was whether the user’s access behavior will cause loss of the integrity, availability, and confidentiality of the information. Ding et al. [33] proposed a privacy-preserving multiparty risk-adaptive access control model. This model proposed a privacy quantification method for dynamically accessing data. Furthermore, a multiparty access control evolutionary game model was constructed based on evolutionary game. Yang et al. [34] designed a flexible access control mechanism based on keyword matching, which enabled data to be shared in a fine-grained access control mode, preventing privacy disclosure while not affecting data usage. Line et al. [35] summarized a variety of access control strategies and analyzed risk assessment criteria according to the actual situation of hospitals, and suggested that future work should
focus on expanding access control strategies based on location and situation; Wang and Jin [36] evaluated users’ access behaviors based on their historical access information. The more chaotic the distribution of users’ access behaviors, the greater the risks that users may cause. Shaikh et al. [37] proposed a risk-based decision access control system, which took into account not only the historical access behavior of users but also the recent access behavior of users, and the system was suitable for dynamic and complex environments like the medical industry. Choi et al. [38] constructed a context-aware medical information risk access control framework, which mainly judged whether to grant users access permissions based on permission files, user access logs, context information, etc.; Hui et al. [39] improved on the basis of literature [36] to prevent doctors from stealing patients’ private information by forging work goals. Zhang [40] and Jiang et al. [41] mainly studied the privacy disclosure of medical Big Data in the cloud environment but focused on the analysis of the risk indicator system that may affect privacy disclosure, without involving specific risk quantification model. Few previous studies [42–44] established a risk assessment model for medical Big Data with the help of fuzzy theory.

A comprehensive analysis of relevant research shows that some scholars are currently doing research on the intersection of information and medicine and have made good achievements. The methods mainly focus on cryptography, anonymity, differential privacy, and so on, and some are analyzed from the perspective of management. Although there are also studies from the perspective of risk and access control, they are still in the initial stage of exploration and have not formed a relatively mature system model framework, especially for the research on the privacy protection of medical Big Data based on risk access control is extremely scarce.

3. Medical Big Data Privacy Protection Model

Workflow refers to the automation of part or whole of a business process in the computer application environment. It is an abstract and general description of the business rules between the workflow and each operation step [45]. Before studying the privacy protection of medical Big Data, we should be clear about the user’s workflow, authorization management mode and information use process in HIS, otherwise it will be meaningless to discuss privacy protection apart from the actual situation. Therefore, Section 3.1 first investigates and sorts out the workflow and authorization management mode in HIS, and then establishes the access control model according to the actual situation.

3.1. Workflow and Authority Management Mode in HIS

Through the field investigation of HIS in some hospitals in Kunming, we found that the system generally includes four main modules: outpatient workflow, inpatient workflow, permission allocation, and drug storehouse, while the first three modules are mainly involved in the study of medical Big Data privacy issues. In the outpatient workflow, the outpatient cashier is responsible for logging in the system to fill in the patient’s registration information, outpatient fees, and refund processing; the outpatient doctor selects the department responsible for issuing medical advice. In the inpatient workflow, the inpatient nurse is responsible for the patient’s admission registration, prepayment entry, patient admission, filling in the admission diagnosis information, and patient’s basic health information; the resident is responsible for prescribing short-term or long-term medical advice to the corresponding patient, which is reviewed and implemented by the resident nurse.

In terms of authority allocation, most hospitals adopt role-based authorization management mode. The system administrator first adds the staff’s basic information in the basic information bar of the personnel management module. The hospital mainly includes outpatient department, inpatient department, drug system, clinical department, medical technology department, hospital leader, and so on. The departments are divided into different offices and wards. When adding the staff’s basic information, the system will automatically generate the doctor’s code or the employee number, and the code is needed in the permission allocation. Then the administrator fills in the employee’s department, section, position, title, and other relevant information in the personnel management column, and assigns the account number and password of the system to the employee. The account type includes financial personnel, outpatient financial account, outpatient doctor, system user, medical technology department, hospital office, inpatient care, resident doctor, and so on. User can view the login records of each account type. Finally, the administrator assigns roles to each user. The role management interface includes Administrators, Guests, Public, office staff, financial staff, decision analysis, developer, outpatient registration, outpatient charge, outpatient pharmacy, outpatient doctor, outpatient doctor station, personnel management, data center, resident nurse, resident doctor, medical technology department, and so on. The administrator grants different access modules to different roles. Therefore, the whole process realizes the assignment relationship between user-role-permission in HIS.

From the aforementioned analysis, it can be seen that doctors in different jobs and roles have different work tasks and access authority sets. In general, after a patient is discharged from hospital, the patient’s paper medical record and medical information will be sealed by the medical record room, and junior professional title doctors can only use their own employee number and password to query the patient’s recent medical information. However, in order not to affect the normal work of doctors, some highly qualified doctors or experts will be granted extremely high authority, they can access patient information not only for the whole group, but even for the entire district. In addition, the amount of information that doctors need to complete their respective tasks will vary depending on the patient’s medical history, the number of patients, and other factors. For example, when a patient is diagnosed with cancer, the doctor has access to a lot of sensitive cancer-related information. It is difficult to adapt to the actual situation of the medical environment to evaluate the doctor’s access risk by the amount
of doctor’s access information or the sensitivity of doctor’s access data. This paper is interested in whether the risks caused by users’ access behavior are worth it, that is, whether to grant access to users is determined by measuring the relationship between risks and benefits.

3.2. Entropy-Based Risk-Adaptive Access Control Model. Entropy characterizes the chaos degree of random variable distribution. The more chaotic the distribution is, the greater the entropy is. The essence of entropy is to measure the amount of information. Information entropy, also known as Shannon entropy, is usually used to describe the average amount of information brought by the whole random distribution, and has more statistical characteristics. Since entropy is calculated based on sample data, it is also called empirical entropy. The relevant formula is defined as follows.

Definition 1. Assuming that \( X \) is a random variable and the random distribution of \( X \) is \( P(X) \), the self-information \( I(x) \) of the random variable \( x (x \in X) \) and the information entropy \( H(X) \) of \( X \) are

\[
I(x) = -\log P(x),
\]

\[
H(X) = \sum_{x \in X} P(x) \times I(x).
\]  

(1)

According to Definition 1, entropy represents the chaos degree of random variable distribution, which can be used to formally represent the instability of user access behavior. Therefore, the entropy-based risk-adaptive access control model aims to quantify users’ access behaviors by means of entropy. In HIS, each user will have a corresponding work task, and access the patient’s information resources according to the work task. We believe that if the information resources accessed by the user are not relevant to the work task, or the correlation is very low, the user will have the risk and possibility of snooping the patient information. A user’s access behavior is set to a six-tuple:

\[
(U, R, T, P, M, \alpha),
\]

(2)

where \( U \) is the set of all users, \( R \) represents the set of roles, \( T \) represents the set of tasks, \( P \) represents the set of permissions, \( M \) is the set of medical records, and \( \alpha \) is the correlation between medical records and work tasks. The relationship between factors is as follows: the same user can be granted different roles, and the same role can also be assigned to different users. Users and roles are in a many-to-many relationship. In Figure 1, User1 can be granted Role1, Role2, and Role4; Role1 can be assigned to both User1 and User2. Tasks are assigned to users according to roles, and roles map users to corresponding tasks. A role can be assigned multiple tasks, and a task can also be assigned to multiple roles, tasks, and roles are also in a many-to-many relationship. In Figure 1, Role3 can access the medical information resources required by Task2 and Task4, and the medical information resources provided for Task2 can be accessed by Role2 and Role3. Tasks are assigned to users through roles to perform their permissions. The specific structure and relationship are shown in Figure 1.

The following will analyze in detail how to evaluate the user’s access behavior. As we can be seen from the previous analysis, the correlation between the user’s work task and the medical records accessed by the user is an important basis for evaluating the user’s access behavior. Therefore, we should first clarify two questions before evaluating users’ access behavior: (a) How to mark the medical information accessed by users and (b) How to quantify the risk value of known users’ historical access behavior. The rest of Section 3.2 will focus on the aforementioned two aspects to analyze the user’s access behavior.

3.2.1. Marking Medical Information. Most hospitals use the ICD-10 code to classify and label the diagnosis results of patients. ICD-10 is the 10th revision of the International Statistical Classification of Diseases and Health-related Problems. The disease name corresponds to a single code, and in turn, the disease name can be found through the disease classification code, and then the medical staff can extract the required medical record data. Generally, it is “one disease one code,” which can provide a unified classification standard for medical record management, health statistics, medical information utilization, and scientific research.

Although there is a one-to-one relationship between disease name and coding, in practical problems, in order to fully understand the disease, it is normal for users to access some medical record information related to the disease. Therefore, according to the characteristics of ICD-10 diagnostic code, the paper reasonably summarizes the medical records of similar diseases through clustering. The 6-digit ICD-10 is used in China’s medical system. For example, B18.202, the first three digits are the mixed codes of letters and numbers, which are the category codes with practical significance. The fourth digit or letter code is the subclassification of the first three digits, and the more similar diseases are, the more similar their ICD-10 codes are. Therefore, when the first letter is the same, the second and third digits of the ICD-10 diagnostic code are taken out, and when the second and third digits of the ICD-10 code are the same, they are grouped into one category.

Assume that the diagnostic code of a target disease is B18.902, and the set of diagnostic codes after clustering is as follows: \( S = \{B18.000, B18.001, B18.002, B18.003, B18.103 + N08.0^*, B18.202, B18.204, B18.902, B18.904 + N08.0^*\} \). When the medical information records accessed by a user under a specific work task belong to set \( S \), it indicates that the user’s access behavior is within the scope of the work target. However, when the medical information records accessed by users under a specific work task do not belong to set \( S \), we need to evaluate the instability of their access behavior and specifically quantify the correlation \( \alpha \). Therefore, it is theoretically feasible to mark medical information accessed by users through diagnostic codes.

3.2.2. Risk Quantification. We define the deviation degree between users’ access to medical information and their work tasks as risks. The greater the deviation degree is, the greater
the possibility of privacy snooping. When the risk reaches a certain level, users’ access behavior should be controlled. We use entropy to quantify the user’s access behavior. Before calculating the user’s access behavior entropy, we need to define the probability $P$ of the user’s access to the medical information record set $m_k$.

Let the patient set $D = \{d_1, d_2, \ldots, d_{I_d}\}$ received by user $u_i$ within a period of time, and the medical record set accessed by use $u_i$ for patient $d_j$ is $M = \{m_1, m_2, \ldots, m_{I_m}\}$, then the probability of user accessing medical record set $m_k$ is as follows:

$$P(m_k | u_i, d_j) = \frac{\|f(m_k)\|}{\sum_{m \in M} \|f(m)\|}$$  \hspace{1cm} (3)

where $I_d$ represents the number of patients, $I_m$ represents the number of elements in the set $M$, $d_j \in D$, $m_k \in M$ is a set of diagnostic codes for a certain disease in the medical system, and each element in $M$ represents a set of diagnostic codes, $\|f(m_k)\|$ represents the number of users access to the set $m_k$. Therefore, the information entropy of the medical information accessed by the user $u_i$ for the patient $d_j$ under a specific task within a period of time is

$$H(u_i, d_j) = - \sum_{k=1}^{I_m} P(m_k | u_i, d_j) \times \log P(m_k | u_i, d_j).$$  \hspace{1cm} (4)

Let $\{H(u_i, d_1), H(u_i, d_2), \ldots, H(u_i, d_{I_d})\}$ is the entropy of user $u_i$ access to medical information resources when user $u_i$ treats all patients for a specific task within a certain period of time. By calculating the mean value of $\{H(u_i, d_1), H(u_i, d_2), \ldots, H(u_i, d_{I_d})\}$, we can get the entropy $H(u_i)$ of user $u_i$ accessing medical information resources for specific work tasks in a certain period of time.

$$H(u_i) = \frac{\sum_{j=1}^{I_d} H(u_i, d_j)}{I_d}.$$  \hspace{1cm} (5)

The higher the entropy $H(u_i)$ of the user’s access to medical information, the more unstable the user’s access behavior, the greater the deviation degree between the medical records accessed by the user and their work tasks, the smaller the correlation, and the greater the possibility of disclosing privacy information. Therefore, the entropy $H(u_i)$ of user access to medical information is inversely proportional to the correlation parameter $\alpha$, which can be expressed as follows:
\[ \alpha = 1 - H(u_i). \] (6)

To sum up, we can get the entropy \( H(u_1), H(u_2), \ldots, H(u_I) \) of all users accessing medical information in the HIS within the same time period, where \( I \) represents the number of users in the HIS. Then, we take the entropy of all users’ access to medical information as the input of K-means clustering, and finally obtained two clustering centers \((x_1, y_1)\) and \((x_2, y_2)\) [46]. We averaged the ordinate of the two clustering centers and used them as the benchmark \( \pi \) for risk assessment. Therefore, the access behavior risk of each user \( u_i \) is defined as

\[ \text{Risk} = \max(0, H(u_i) - \pi). \] (7)

4. Case Analysis

We take the inpatient department as an example to analyze the effectiveness of the model in practical applications. That is, whether the model can evaluate the user’s access behavior risk based on the actual situation of the hospital and the existing data resources. According to the previous analysis, when the information resources accessed by users are not related to the work objectives or the correlation \( \alpha \) is low, the user will have the risk of disclosing the patient’s privacy information. Therefore, the validity of the model can be tested by studying correlation \( \alpha \) and risk. When users with high correlation \( \alpha \) have low risk of access behavior, and users with low risk have high correlation \( \alpha \) between access information resources and work objectives, the model in the paper can be considered to be effective.

According to the requirements of this paper, some medical data have been obtained from the Oracle database of a third-class hospital in Kunming, including doctor code table Dmb_ysdm, hospital department code table Dmb_ksdm, patient basic information table Zy_hzjbx, doctor’s order table Zy_yz, and patient inpatient information table Zy_hzzxx. It is worth noting that each table in the hospital has more than 200 fields at most. The table shown in this paper is a regenerated table after extracting fields from multiple tables according to the needs of the model. When a doctor access electronic medical records, an access log will be generated, including the Doctor code, Patient’s identification number, Medical record no., and access time. When users view the log content, they need to retrieve Dmb_ysdm and Dmb_ksdm. Under a specific work task, the doctor needs to obtain the basic information of the patient (Patient’s condition, Clinic diagnosis, Admission diagnosis, ICD-10 code, medical record no., etc.) according to the patient’s medical record number, and then needs to search Zy_hzjbx, Zy_yz, and Zy_hzzxx. The relevant information is shown in Tables 1 to 5.

As medical institutions prohibit public access to data, this paper presents only a part of patients’ data in the basic patient information table Zy_hzjbx, as shown in Figure 2.

According to Section 3.1, each user in HIS has a corresponding role and access module, and the user’s access record can be viewed under the corresponding role, including doctor code, access content, access time and access frequency. In addition, by retrieving the doctor code, you can also query the user’s role, department, job title, and other information. We combine Dmb_ysdm, Dmb_ksdm, Zy_hzjbx, Zy_yz, and Zy_hzzxx to get a user’s work goals and access information resources within a period of time. In the experiment, we randomly selected 30 users from the inpatient department of HIS, and tracked and marked their access within half a year. The marking content mainly includes the patient information received by the user in the past half a year, especially the patient’s ICD-10 diagnostic code and the medical information record set accessed by the user for the patient. Table 6 shows the medical information accessed by the resident \( u_i \) when treating the patient with ID 4151097.

Combining equations (3) and (4), we can calculate that the information entropy of the user accessing medical information for the patient’s condition within half a year is 0.86. In the same way, according to equation (5), we can get the entropy \( H(u_1) \) of the user accessing medical information for a specific task (patient’s diagnostic code is A19.900) within half a year, and the entropy \( \{H(u_2), H(u_3), \ldots, H(u_{30})\} \) of all users accessing medical information within the same period. The specific value of entropy of user access to medical information resources is shown in Figure 3. The entropy in Figure 3 is prepared for us to calculate the risk benchmark and assess the risk of user access behavior. We take the entropy of all users accessing medical information resources as the input data set for K-means clustering. Finally, we get two clustering centers \((x_1, y_1)\) and \((x_2, y_2)\), and the results are shown in Figure 4.

As we can be seen from Figure 3, in the same time period (the time set is half a year in the paper), different users have different access behavior risks for specific work tasks. This experiment proves that it is feasible for us to use the information entropy method to evaluate user access behavior’s risk. It can be seen from Figure 4 that the entropy of 30 users accessing medical information is divided into two categories. The cluster center \( y_1 \) of the first type is 0.45, and the cluster center \( y_2 \) of the second type is 0.60, which are marked with red dots in the figure. Therefore, the risk benchmark \( \pi = 0.525 \) is obtained according to \( y_1 \) and \( y_2 \). The extent to which the user’s access behavior deviates from the risk baseline is shown in Figure 5. Where we set the risk value of users whose access entropy is below the risk baseline to be 0 and default to legitimate users, while the risk value of the user whose access entropy is above the risk baseline can be calculated according to equation (7), as shown in Figure 6.

From Figure 6, it is easy to determine whether to grant access to the user. As we can be seen from Figure 6, users whose access entropy is lower than the risk baseline are considered to be risk-free, and the model grants the user access rights. If the risk value is not 0, that is, the user’s current access request may cause privacy security problems, the user’s access request is rejected. Finally, in order to verify the validity of the model in this paper, we analyzed the change relationship between correlation \( \alpha \) and risk, as shown in Figure 7. The results show that users with high correlation \( \alpha \) have low risk of access behavior, and users with low risk
| Table 1: Dmb_ysdm. |
|---------------------|
| Column | Name         | Type       | Remark                                                                 |
| Ysdm   | Doctor code  | CHAR(3)    |                                                                       |
| Ysxm   | Doctor name  | CHAR(10)   | Get the doctor's department from the ksdm field in mb_ksdm            |
| Yszc   | Doctor title | CHAR(12)   |                                                                       |
| Ksdm   | Subordinate department | CHAR(2) |                                                                       |

| Table 2: Dmb_ksdm. |
|---------------------|
| Column | Name         | Type      | Remark                                                                 |
| Ksdm   | Department code | CHAR(3)   | “1” is the inpatient department                                       |
| Ksmc   | Department name | CHAR(10)  |                                                                       |
| Zybz   | Whether in the hospital | CHAR(12) |                                                                       |

| Table 3: Zy_hzjbxx. |
|---------------------|
| Column | Name                | Type           | Remark                                                                 |
| Bah    | Patient’s identification number | CHAR(8)        | If there is no special description, the default value is used         |
| Hzxm   | Patient name        | CHAR(10)       |                                                                       |
| Blh    | Medical record no.  | CHAR(10)       |                                                                       |
| Xb     | Gender              | CHAR(2)        |                                                                       |
| Srzh   | ID                  | CHAR(21)       |                                                                       |
| Mzd   | Clinic diagnosis    | CHAR(36)       |                                                                       |
| Ryzd   | Admitting diagnosis | CHAR(36)      |                                                                       |
| Kdys   | Billing doctor      | CHAR(3)        |                                                                       |
| Fzys   | Responsibility doctor | CHAR(10)  |                                                                       |
| Bq     | Patient’s condition | CHAR(10)      |                                                                       |
| Hljb   | Nursing degree      | CHAR(10)       |                                                                       |
| Gcys   | Tube bed doctor     | CHAR(3)        |                                                                       |

| Table 4: Zy_yz. |
|----------------|
| Column | Name         | Type           | Remark                                                                 |
| Brid   | Patient Id   | CHAR (10)      | “1” is the inpatient department                                       |
| Zdsj   | Diagnosis time | CHAR(10)   |                                                                       |
| Zdm    | ICD-10 code (diagnosis code) | CHAR(10) |                                                                       |
| Zyh    | Admission number | CHAR(3)   |                                                                       |
| Ksmc   | Department name | CHAR(10) |                                                                       |
| Zybz   | Whether in the hospital | CHAR(12) |                                                                       |

| Table 5: Zy_hzzyxx. |
|---------------------|
| Column | Name                | Type           | Remark                                                                 |
| Bah    | Patient’s identification number | CHAR (8)       | If there is no special description, the default value is used        |
| Hzm    | Patient name        | CHAR(10)       |                                                                       |
| Blh    | Medical record no.  | CHAR(10)       |                                                                       |
| Ryks   | Admission department | CHAR(2)      |                                                                       |
| Zzys   | Attending doctor    | CHAR(3)        |                                                                       |
| Gcys   | Tube bed doctor     | CHAR(3)        |                                                                       |
| Ksmc   | Admission bed       | CHAR(12)       |                                                                       |
| Zybz   | Whether in the hospital | CHAR(12) |                                                                       |
| Xzks   | Current department  | CHAR(2)        |                                                                       |
| Xzcw   | Current bed number  | CHAR(12)       |                                                                       |
### Table 6: User access information table.

| Patient Id | Diagnostic code | Doctor code | Medical records within the scope of the user's work goals | Medical record information actually accessed by the user | Access frequency |
|------------|-----------------|-------------|----------------------------------------------------------|---------------------------------------------------------|-----------------|
| A19.00     | A19.000         | 1           | A19.000                                                  | 1                                                       |
| A19.001    | A19.100         | 1           | A19.100                                                  | 1                                                       |
| A19.100    | A19.200         | 1           | A19.200                                                  | 1                                                       |
| A19.200    | A19.800         | 1           | A19.800                                                  | 1                                                       |
| A19.800    | A19.900         | 2           | A19.900                                                  | 2                                                       |
| 4151097    | A19.900         | 507         | A19.900                                                  | 507                                                     |
| A19.801    | A19.900         | 1           | A19.900                                                  | 1                                                       |
| A19.802    | A20.900         | 2           | A20.900                                                  | 2                                                       |
| A19.803    | A19.901         | 3           | A19.901                                                  | 3                                                       |
| A19.900    |                  |             |                                                          |                                                          |

**Figure 2:** Patients' basic information in Zy_hzjbx.

**Figure 3:** Entropy of user access to medical information resources.
Figure 4: K-means clustering results.

Figure 5: The degree to which users' access behavior deviates from the risk baseline.

Figure 6: User access behavior risk.
have high correlation $\alpha$ between access information resources and work goals. Therefore, managers can predict the risks of users' access behaviors based on the correlation between the information resources accessed by users and the work objectives. In this way, administrators can dynamically formulate access control policies based on users’ access conditions.

5. Conclusion

In this paper, we sort out the current hospital workflow and management mode and establish an access control model based on risk adaptive. By analyzing the correlation between information resources accessed by users and work objectives, we assess the risk of patient information disclosure caused by users’ access behavior. The experimental results show that hospital administrators can predict the risk of privacy disclosure caused by users’ access behavior, this discovery helps them to formulate scientific access control strategies. However, since the HISs of most hospitals are different at present, the privacy-preserving model proposed in this paper cannot be fully adapted to all hospitals. In the future, we will study a more compatible access control privacy protection model, provide new ideas for the hospital’s resource management model, and promote the overall progress of the hospital’s comprehensive management capabilities.

Data Availability

The original data of this article have been signed in a confidentiality agreement with the hospital and are temporarily unavailable, but the processed data (data used to support the research in this article) can be partially shared publicly and submitted with the manuscript.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant nos. 71972165 and 61763048) and the Science and Technology Foundation of Yunnan Province (grant no. 202001AS070031).

References

[1] Y. Yin, Z. Cao, Y. Xu, H. Gao, R. Li, and Z. Mai, "QoS prediction for service recommendation with features learning in mobile edge computing environment," IEEE Transactions on Cognitive Communications and Networking, vol. 6, no. 4, pp. 1136–1145, 2020.
[2] P. Zhang, Y. Zhang, H. Dong, and H. Jin, “Mobility and dependence-aware QoS monitoring in mobile edge computing,” IEEE Transactions on Cloud Computing, vol. 9, 2021.
[3] X. Jin, Big Data in Health Care, People’s Medical Publishing House, USA, 2018.
[4] R. Jiang, Y. Xin, Z. Chen, and Y. Zhang, “A medical big data access control model based on fuzzy trust prediction and regression analysis,” Applied Soft Computing, vol. 117, Article ID 108423, 2022.
[5] Y. Xu, Y. Wu, H. Gao, S. Song, Y. Yin, and X. Xiao, “Collaborative APIs recommendation for artificial intelligence of things with information fusion,” Future Generation Computer Systems, vol. 125, pp. 471–479, 2021.
[6] Y. Huang, H. Xu, H. Gao, X. Ma, and W. Hussain, “SSUR: an approach to optimizing virtual machine allocation strategy based on user requirements for cloud data center,” IEEE Transactions on Green Communications and Networking, vol. 5, no. 2, pp. 670–681, 2021.
[7] H. Gao, Y. Zhang, H. Miao, R. J. D. Barroso, and X. Yang, "SDTIOA: modeling the timed privacy requirements of iot...
service composition: a user interaction perspective for automatic transformation from BPEL to timed automata,” *Mobile Networks and Applications*, vol. 26, pp. 1–26, 2021.
[8] C. Lin, Z. Song, H. Song, Y. Zhou, Y. Wang, and G. Wu, “Differential privacy preserving in big data analytics for connected health,” *Journal of Medical Systems*, vol. 40, no. 4, 2016.
[9] C. Lin, P. Wang, H. Song, Y. Zhou, Y. Wang, and G. Wu, “A differential privacy protection scheme for sensitive big data in body sensor networks,” *Annals of Telecommunications*, vol. 71, no. 9, pp. 465–475, 2016.
[10] J. L. Raisaro, G. Choi, S. Pradervand et al., “Protecting privacy and security of genomic data in i2b2 with Homomorphic Encryption and Differential Privacy,” *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, vol. 15, no. 5, pp. 1413–1426, 2017.
[11] X. Wu, *Privacy Protection and its Key Technologies in Big Data*, Nanjing University, Nanjing, China, 2017.
[12] Y. Bai, “Data Base Technique, Application of differential privacy protection in medical big data,” *Electronic Technology & Software Engineering*, vol. 24, pp. 196–197, 2017.
[13] M. Sung, D. Cha, and Y. Park, “Local differential privacy in the medical domain to protect sensitive information: algorithm development and real-world validation,” *JMIR Medical Informatics*, vol. 9, no. 11, Article ID e26914, 2021.
[14] L. Zhang, X. Zhu, J. Ma, Z. Ma, and D. Yuan, “Medical privacy-preserving service recommendation,” in *Proceedings of the ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, pp. 1–6, Dublin, Ireland, June 2020.
[15] D. R. Harris, “Leveraging differential privacy in geospatial analytics of standardized healthcare data,” in *Proceedings of the IEEE International Conference on Big Data (Big Data)*, pp. 3119–3122, Atlanta, GA, USA, December 2020.
[16] Y. Hu, L. Ge, G. Zhang, and D. Qin, “Research on differential privacy for medical health big data processing,” in *Proceedings of the 20th International Conference on Parallel and Distributed Computing, Applications and Technologies (PDCAT)*, pp. 140–145, Gold Coast, QLD, Australia, December 2019.
[17] A. M. Olawoyin, C. K. Leung, and R. Choudhury, “Privacy-preserving spatio-temporal patient data publishing,” in *Proceedings of the Database and Expert Systems Applications*, pp. 407–416, Bratislava, Czech Republic, September 2020.
[18] J. W. Kim, B. Jang, and H. Yoo, “Privacy-preserving aggregation of personal health data streams,” *PLoS One*, vol. 13, no. 11, Article ID e0207639, 2018.
[19] Z. Lv and F. Piccialli, “The security of medical data on internet based on differential privacy technology,” *ACM Transactions on Internet Technology*, vol. 21, no. 3, pp. 1–18, 2021.
[20] J. Zhang, “Research on security system of healthcare big data based on cryptographic technique,” *Journal of Information Security Research*, vol. 3, no. 7, 2017.
[21] D. He, S. Zeadally, N. Kumar, and J.-H. Lee, “Anonymous authentication for wireless body area networks with provable security,” *IEEE Systems Journal*, vol. 11, no. 4, pp. 2590–2601, 2016.
[22] Z. Li and L. Zhang, “An EMR sharing and privacy protection mechanism based on medical consortium blockchain,” in *Proceedings of the 6th International Conference on Computer and Technology Applications*, pp. 160–164, Antalya, Turkey, 2020.
[23] W. Chen, S. Zhu, J. Li, J. Wu, C.-L. Chen, and Y.-Y. Deng, “Authorized shared electronic medical record system with proxy Re-encryption and blockchain technology,” *Sensors*, vol. 21, no. 22, 2021.
[24] D. Xanthidis and O. K. Xanthidou, “A proposed framework for developing an electronic medical record system,” *Journal of Global Information Management*, vol. 29, no. 4, pp. 78–92, 2021.
[25] K. Malasri and L. Wang, “Design and implementation of a securewireless mote-based medical sensor network,” *Sensors*, vol. 9, no. 8, pp. 6273–6297, 2009.
[26] Y. Sun, C. Wong, G.-Z. Yang, and B. Lo, “Secure key generation using gait features for body sensor networks,” in *Proceedings of the IEEE 14th International Conference on Wearable and Implantable Body Sensor Networks (BSN)*, pp. 206–210, Eindhoven, Netherlands, May 2017.
[27] J. Zhang, “Broader access models for realizing information,” *DomiCorporation*, Report JSR-04-132, MITRE Corporation, McLean, VA, USA, 2004.
[28] F. K. Dankar and R. Badji, “A risk-based framework for biomedical data sharing,” *Journal of Biomedical Informatics*, vol. 66, pp. 231–240, 2017.
[29] Q. Ni, E. Bertino, and J. Lobo, “Risk-based access control systems built on fuzzy inferences,” in *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security*, pp. 250–260, Beijing, China, 2010.
[30] P.-C. Cheng, P. Rohatgi, C. Keser, P. A. Karger, G. M. Wanger, and A. S. Reninger, “Fuzzy multi-level security: an experiment on quantified risk-adaptive access control,” in *Proceedings of the IEEE Symposium on Security and Privacy (SP'07)*, pp. 222–230, Berkeley, CA, USA, May 2007.
[31] N. N. Diep, L. X. Hung, Y. Zhuang, and S. Lee, “Enforcing access control using risk assessment,” in *Proceedings of the Fourth European Conference on Universal Multiservice Networks (ECUM'07)*, pp. 419–424, Toulouse, France, February 2007.
[32] M. Sharma, Y. Bai, S. Chung, and L. Dai, “Using risk in access control for cloud-assisted ehealth,” in *Proceedings of the IEEE 14th International Conference on High Performance Computing and Communication & 2012 IEEE 9th International Conference on Embedded Software and Systems*, pp. 1047–1052, Liverpool, UK, June 2012.
[33] H. Ding, C. Peng, and Y. Tian, “Privacy risk adaptive access control model via evolutionary game,” *Journal on Communications*, vol. 40, no. 12, pp. 9–20, 2019.
[34] Y. Yang, X. Zheng, W. Guo, X. Liu, and V. Change, “Privacy-preserving fusion of IoT and big data for e-health,” *Future Generation Computer Systems*, vol. 86, pp. 1437–1455, 2018.
[35] M. B. Line, I. A. Tøndel, and E. A. Gjære, “A risk-based evaluation of group access control approaches in a healthcare setting,” *International Conference on Availability*, vol. 6908, pp. 26–31, 2017.
[36] Q. Wang and H. Jin, “Quantified risk-adaptive access control for patient privacy protection in health information systems,” in *Proceedings of the 6th ACM symposium on information, computer and communications security*, pp. 406–410, Hong Kong, China, March 2011.
[37] R. A. Shaikh, K. Adi, and L. Logrippo, “Dynamic risk-based decision methods for access control systems,” *Computers & Security*, vol. 31, no. 4, pp. 447–464, 2012.
[38] D. Choi, D. Kim, and S. Park, “A framework for context sensitive risk-based access control in medical information systems,” *Computational and Mathematical Methods in Medicine*, vol. 2015, Article ID 265312, 15 pages, 2015.
[39] Z. Hui, H. Li, M. Zhang, and D.-G. Feng, “Risk-adaptive access control model for big data in healthcare,” *Journal on Communications*, vol. 36, no. 12, pp. 190–199, 2015.
[40] J. Zhang, *The Risk Assessment of Medical Big Data Privacy Security Cloud Environment*, Yunnan University of Finance and Economics, Yunnan, China, 2018.

[41] R. Jiang, M. Shi, and W. Zhou, “A privacy security risk analysis method for medical big data in urban computing,” *IEEE Access*, vol. 7, Article ID 143841, 2019.

[42] J. Li, Y. Bai, and N. Zaman, “A fuzzy modeling approach for risk-based access control in eHealth cloud,” in *Proceedings of the 12th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*, pp. 17–23, Melbourne, VIC, Australia, July 2013.

[43] E. Smith and J. Eloff, “Cognitive fuzzy modeling for enhanced risk assessment in a health care institution,” *IEEE Intelligent Systems and Their Applications*, vol. 15, no. 2, pp. 69–75, 2000.

[44] M. Shi, R. Jiang, X. Hu, and Y. Shang, “A privacy protection method for health care big data management based on risk access control,” *Health Care Management Science*, vol. 23, no. 3, pp. 427–442, 2020.

[45] X. Ma, H. Xu, H. Gao, and M. Bian, “Real-time multiple-workflow scheduling in cloud environments,” *IEEE Transactions on Network and Service Management*, vol. 18, no. 4, pp. 4002–4018, 2021.

[46] Y. Zhu, W. Zhang, Y. Chen, and H. Gao, “A novel approach to workload prediction using attention-based LSTM encoder-decoder network in cloud environment,” *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–18, 2019.