How to ensure the confidentiality of electronic medical records on the cloud: A technical perspective

Zongda Wu, Shaolong Xuan, Jian Xie, Chongze Lin, Chenglang Lu

PII: S0010-4825(22)00504-2

DOI: https://doi.org/10.1016/j.compbiomed.2022.105726

Reference: CBM 105726

To appear in: *Computers in Biology and Medicine*

Received Date: 21 May 2022

Revised Date: 8 June 2022

Accepted Date: 11 June 2022

Please cite this article as: Z. Wu, S. Xuan, J. Xie, C. Lin, C. Lu, How to ensure the confidentiality of electronic medical records on the cloud: A technical perspective, *Computers in Biology and Medicine* (2022), doi: https://doi.org/10.1016/j.compbiomed.2022.105726.

How to ensure the confidentiality of electronic medical records on the cloud: A technical perspective☆

Zongda Wu[a], Shaolong Xuan[a,*], Jian Xie[a,**], Chongze Lin[b], Chenglang Lu[c]

[a]*Department of Computer Science and Engineering, Shaoxing University, Shaoxing 312000, Zhejiang, China*
[b]*Zhejiang Economics Information Centre, Hangzhou 310006, Zhejiang, China*
[c]*Zhejiang Institute of Mechanical and Electrical Engineering, Hangzhou 310053, Zhejiang, China*

Abstract

From a technical perspective, for electronic medical records (EMR),, this paper proposes an effective confidential management solution on the cloud, whose basic idea is to deploy a trusted local server between the untrusted cloud and each trusted client of a medical information management system, responsible for running an EMR cloud hierarchical storage model and an EMR cloud segmentation query model. (1) The EMR cloud hierarchical storage model is responsible for storing light EMR data items (such as patient basic information) on the local server, while encrypting heavy EMR data items (such as patient medical images) and storing them on the cloud, to ensure the confidentiality of electronic medical records on the cloud. (2) The EMR cloud segmentation query model performs EMR related query operations through the collaborative interaction between the local server and the cloud server, to ensure the accuracy and efficiency of each EMR query statement. Finally, both theoretical analysis and experimental evaluation demonstrate the effectiveness of the proposed solution for confidentiality management of electronic medical records on the cloud, i.e., which can ensure the confidentiality of electronic medical records on the untrusted cloud, without compromising the availability of an existing medical information management system.

Keywords: Disease Diagnosis; Medical Diagnosis; Cloud Computing; Electronic Medical Record; Medical Image; Confidentiality; Availability

1. Introduction

Cloud computing, through pay-per-use, enables an organization to access their required resources easily from a configurable computing resource sharing pool, anytime and anywhere [1, 2, 3], thereby, reducing the investment of an organization on business operation and data management, and then improving the efficiency of organization management. To this end, hospitals are vigorously implementing the Cloud First strategy [4, 82], i.e., giving priority to the cloud computing mode in the process of medical informatization, such that more and more medical data (such as medical diagnosis data and disease diagnosis data) are generated, stored and managed on the cloud. It has become the general trend for medical data management on the cloud [5]. However, although medical data management on the cloud can effectively reduce the management cost of hospitals on medical data and in turn improve the management efficiency, it also results in some negative impacts, among which, the most prominent is the confidentiality problem of medical data on the cloud [7].

Electronic Medical Record (EMR) is the digital record of the whole process of patient medical diagnosis (or disease diagnosis) and treatment in hospitals. It is the most common medical data, and it contains rich personal sensitive information and has great economic and social value [8]. Fig. 1 illustrates a basic architecture of an EMR management system under a cloud computing environment. It can be seen that under a cloud computing environment, electronic medical records are not stored on a trusted local server of hospital information center, but stored and managed by an untrusted cloud server, resulting in the separation of data from its owner, and then a serious threat to the confidentiality of electronic medical records [9, 10]. Such a threat mainly comes from two aspects: (1) external threat, i.e., hacker attacks on cloud service providers (it has been demonstrated by endless hacking incidents) [11]; and (2) internal threat, i.e., inside jobs from cloud service providers (driven by interests, it is common for managers of service providers to maliciously steal medical data) [12]. In short, the confidentiality of medical data on the cloud (i.e., how to ensure the confidentiality of sensitive medical data on the untrusted cloud) has become an important obstacle restricting the development and application of the cloud computing technology in a medical information system, which has attracted increasingly widespread

☆The work is supported by the key project of Humanities and Social Sciences in Colleges and Universities of Zhejiang Province (2021GH017), Humanities and Social Sciences Project of the Ministry of Education of China (21YJA870011), Philosophy and Social Science Planning Project of Zhejiang Province (22ZJQN45YB), and National Social Science Foundation of China (21FTQB019).

*Corresponding Author
**Corresponding Author
Email addresses: zongda1983@163.com (Zongda Wu), zjsxxsl@163.com (Shaolong Xuan), chenlingyimin@qq.com (Jian Xie), linchongze123@qq.com (Chongze Lin), playnet107@163.com (Chenglang Lu)

attention.

Scholars in the area of humanities and social sciences attempted to solve the problem of data protection of network users from the view of laws and regulations. However, it has been pointed out in [4, 13] that making laws and regulations can mitigate the problem of data security to a certain extent, but it cannot solve the problem fundamentally. In other words, the confidentiality management of medical data on the untrusted cloud requires not only the support of laws and regulations, but also the support of technical methods [14, 15, 16]. In general, to ensure the confidentiality of medical data, a medical information system has deployed multiple technical methods, such as identity authentication, access control and encryption. Both identity authentication and access control can prevent external users from illegal access to sensitive medical data in a medical information system, and thus alleviate the security problem of electronic medical records, but they assume the server side credibility, i.e., they are only targeted for external illegal attackers, and cannot prevent internal staff on the untrusted server side from accessing electronic medical records [36, 17, 18, 19]. However, because of being away from local control, the cloud server side is not credible, which is the main source resulting in data security. Therefore, the problem of confidentiality management of electronic medical records on the cloud cannot be solved by traditional access control and identity authentication. Data encryption is an important means to solve the problem of data confidentiality under a cloud environment [38, 39, 40, 41]. However, in an EMR management platform, there exist a large number of query operations defined on electronic medical records (e.g., querying by patient names). Once the electronic medical records stored in the cloud database are encrypted strictly, the original EMR related query operations in the system no longer can be correctly executed on EMR ciphertext on the cloud, which seriously damages the medical data query accuracy [42]. Therefore, the problem of confidentiality management of electronic medical records on the cloud cannot be directly solved by traditional data encryption.

To this end, aiming at the confidentiality management problem of electronic medical records under a cloud computing environment, this paper provides an effective solution, which can effectively improve the security of electronic medical records on the untrusted cloud, without compromising the availability of an existing EMR management system. Its basic idea is to deploy a trusted local server between the untrusted cloud and trusted clients of an EMR management system, responsible for running an EMR cloud hierarchical storage model and an EMR cloud segmentation query model. Specifically, the contributions of this paper are threefold.

(1) An EMR hierarchical storage model is proposed to store light EMR data items (such as patient basic information) on the local server, and encrypt heavy EMR data items (such as patient medical images) and store them on the cloud, to ensure the confidentiality of EMRs on the cloud.
(2) An EMR segmentation query model is proposed to perform EMR related query operations by the collaborative interaction between the local server and the cloud, to ensure the accuracy and efficiency of each EMR query statement.
(3) Finally, both theory analysis and experimental evaluation are performed to demonstrate the performance of the proposed solution. The results show that the proposed solution can effectively improve the security of electronic medical records on the untrusted cloud, without compromising the availability of an existing EMR cloud management system.

The rest of this paper is organized as follows. Section 2 reviews related work. Section 3 presents a proposed solution for confidentiality management of electronic medical records on the cloud, which includes a system framework, an EMR cloud hierarchical storage model and an EMR cloud segmentation query model. Section 4 analyzes the impact of the proposed solution on an existing EMR management platform on the cloud in terms of confidentiality, accuracy, efficiency and availability. In Section 5, we conclude this paper.

2. Related Work

Privacy right is the personality right enjoyed by a natural person. The rapid development of network technologies such as cloud computing poses a great challenge to the protection of citizens' privacy rights, and the problem of data protection of network users has attracted more and more extensive attention [13, 20, 21]. In recent years, China has issued three laws related to network users' data security, i.e., Network Security Law, Data Security Law and Personal Information Protection Law, which also plays an important role in ensuring the confidentiality of medical data on the cloud [4, 22]. However, the endless number of user privacy incidents shows that making laws and regulations can mitigate the problem of data security to a certain extent, but it cannot solve the problem fundamentally. In other words, the confidentiality management of medical data on the untrusted cloud requires not only the support of laws and regulations, but also the support of technical methods [14, 15]. In general, to ensure the confidentiality of medical data, a medical information management system has deployed multiple technical methods or strategies in advance, such as identity authentication, access control and encryption. Below, we briefly describe the characteristics of these technical methods, and analyze their application limitation in confidentiality management of medical data on the cloud.

Identity authentication is a process of user authentication to prevent illegal user access to system resources [23, 24]. It can be divided into two categories, i.e., single-factor authentication [26, 27] (such as user name authentication, smart card authentication, dynamic password authentication, and biometric authentication), and dual-factor authentication (it further strengthens the effect of identity authentication by combining two single-factor authentications) [28, 29]. Access control is a process of restricting users' access to unauthorized resources or use of unauthorized functions, based on their specific identities [30]. It can be divided into two categories, i.e., discretionary access control [31] and mandatory access control [32]. The identity authentication and access control technologies have been widely used in operating system [33], database system [34]

and management information system [35]. Although the two kinds of technologies can prevent external users from illegal access to sensitive medical data in a medical information system, and thus alleviate the security problem of electronic medical records, they cannot leave the support of the server side (they assume the server side credibility), i.e., they are only targeted for external illegal attackers of a medical information system, and cannot prevent internal staff on the untrusted server side (or hackers who break through the server side) from accessing electronic medical records [36]. However, because of being away from local control, the cloud server side is not credible, which is the main source resulting in data security. In conclusion, the problem of confidentiality management of electronic medical records on the cloud cannot be solved by traditional access control and identity authentication.

Data encryption refers to the strict encryption of user sensitive data stored on the untrusted server side, making the encrypted data difficult to be understood even if being leaked, and in turn improving data security [25, 37, 88]. It is an important means to solve the problem of data confidentiality under a cloud environment [38, 39, 40, 41]. However, in an EMR management platform, there exist a large number of query operations defined on electronic medical records (e.g., querying by patient names). Once the electronic medical records stored in the cloud database are encrypted strictly, the original EMR related query operations in the system no longer can be correctly executed on EMR ciphertext on the cloud, which seriously damages the medical data query accuracy [42]. In order to solve the ciphertext query problem, we can first transmit all the ciphertext medical data on the cloud back to a local server, and then decrypt them and perform each query operation on the decrypted data [43]. However, because almost the whole query process is completed locally, such a way not only completely loses the cost and efficiency advantages of management of medical data on the cloud, but also seriously reduces the efficiency of medical data query (the cost of network transmission and data decryption is very huge). Therefore, the problem of confidentiality management of electronic medical records on the cloud cannot be directly solved by a traditional data encryption means.

Aiming the problem of medical data security on the cloud, some effective technical methods have been proposed. For example, in [5], a security-preserving approach and a privacy-preserving approach, were proposed, where the former can be used to ensure the security and integrity of electronic health records, and the former can be used to protect the privacy of personal health records. In [48, 49, 50], based on the blockchain technology, some privacy-preserving mechanisms for electronic medical records were proposed, which can well ensure the security of electronic medical records on the cloud. In [51], aims at discussing the security of medical data in the Internet era, the authors proposed an algorithm model combining k-anonymity and differential privacy. In [52], based on the watermarking technology, the authors proposed a new data-sharing framework and a data access control mechanism for medical data sharing in smart healthcare. In [53], the authors proposed an appropriate method to improve accessibility of the Australian My Health Records while preserving privacy and security of the system. In addition, in [54, 55, 56, 57, 58], the authors reviews some security and privacy issues of medical data on a cloud computing environment. Overall, for most of the existing technical methods for medical data security, it is required to change or reconstruct the architecture of an existing management platform of medical data on the cloud, consequently, reducing their practical availability.

In addition, aiming at the problem of cloud data security, scholars in the field of computer sciences have also conducted some in-depth systematic research, and proposed many effective technical methods [44, 45, 46, 47]. However, many of the proposals are developed on top of the traditional technical methods mentioned above (i.e., identity authentication, access control and data encryption), making them difficult to meet the actual needs of confidentiality management of medical data on the cloud. Furthermore, many of the proposals are not specifically proposed for medical information systems, making that they still cannot satisfy the practical application requirements of confidentiality management of medical data on the cloud in terms of effectiveness, availability and confidentiality.

From above, we can conclude that under the existing architecture of a management platform of medical data on the cloud, it still needs to be further studied how to effectively improve the security of electronic medical records on the cloud, under the basic premises of not compromising the availability of an existing medical information system and the effectiveness of each medical data query operation. To this end, this paper aims to provide an effective solution, to improve the security of electronic medical records on the untrusted cloud, without compromising the availability of an existing EMR management system.

3. Solution

In this section 3, we describe our proposed solution for confidentiality management of electronic medical records on the cloud. First, in Section 3.1, we describe a system model for confidentiality management of EMRs on the cloud, and then the constraints of a solution constructed based on the system model. Second, in Section 3.2, under the system model, we describe an EMR cloud hierarchical storage model. Finally, in Section 3.3, under the system model, we describe an EMR cloud segmentation query model.

3.1. System Model

A basic framework of the proposed solution for confidentiality management of electronic medical records on the cloud is shown in Figure 2. It is developed on top of an existing EMR management information system under a cloud environment (which is shown in Figure 1). It can be seen that the system framework mainly includes four roles, i.e., EMR input clerk, EMR query clerk, local server and cloud server, whose functions can be described as follows.

(1) An EMR input clerk (usually, which is a doctor) submits an EMR through a trusted EMR input interface. Here, EMR

data is divided into EMR light data (such as patients' background information) and EMR heavy data (such as patients' medical images).

(2) An EMR query clerk (usually, which is a doctor) submits an EMR query request through a trusted EMR query interface to obtain the target EMR data. Here, each query is often defined on EMR light data.

(3) The cloud server is deployed on the untrusted cloud, responsible for storing EMR ciphertext heavy data from the local server, and performing EMR heavy data query requests from the local server.

(4) The local server is deployed on the trusted local, responsible for 1) dividing the EMR data submitted by input clerks into heavy data and light data, and encrypting and storing the heavy data in the cloud database (i.e., running a hierarchical storage model for Steps 1 to 2); and 2) rewriting each EMR query request submitted by query clerks into a light data query and a heavy data query, and ensuring the accuracy and efficiency of each EMR query through the collaborative interaction between the local server and the cloud server (i.e., running a segmentation query model for Steps 3 to 6).

Also, it can be seen that the local server plays an important role in the system framework, actually, which is responsible for running the proposed solution for confidentiality management of electronic medical records. In order to meet the practical application needs of EMR cloud management in terms of availability, effectiveness and confidentiality, for any EMR confidentiality management solution constructed based on the framework of Figure 2, it should meet the following four constraints.

(1) How to ensure the confidentiality of EMR data, which includes how to ensure the confidentiality of EMR light data on the trusted local server, and how to ensure the confidentiality of EMR heavy data on untrusted cloud server.

(2) How to ensure the accuracy of each EMR related query, i.e., before and after the introduction of the proposed solution, the execution results of each original EMR query in the EMR management system remain unchanged.

(3) How to ensure the efficiency of each EMR related query, i.e., before and after the introduction of the proposed solution, the execution efficiency of each original EMR query statement in the EMR management system should not be significantly reduced.

(4) How to avoid the performance bottleneck of the local server, i.e., under the system framework of Figure 2, it is required to deploy the local server, but it should not lead to a storage performance bottleneck or a computation performance bottleneck for the local server.

As can be seen from Figure 2, the confidentiality management of EMRs on the cloud mainly includes two models, i.e., an EMR hierarchical storage model (Steps 1 to 2) and an EMR segmentation query model (Steps 3 to 6), both of which are deployed on the trusted local server. Below, in Section 3.2 and Section 3.3, we describe the two models, respectively. Then, in Section 4, according to the four constraints mentioned above, we demonstrate the effectiveness of the proposed solution for EMR confidentiality management, by theoretical analysis and experimental evaluation.

3.2. Storage Model

The EMR hierarchical storage model corresponds to Step 1 and Step 2 shown in Figure 2, whose workflow can be further described by Figure 3. It can be seen that the EMR hierarchical storage model includes the following four steps.

Step 1.1 EMR Release. An EMR input clerk submits an electronic medical record by a trusted EMR input interface, to the local server. An electronic medical record can be denoted by

```
EMR = (data[i][1], data[i][2], ...)
```

, where `data[i][j]` denotes a simple data item (such as patient name, doctor name, medical picture and diagnostic description) of an electronic medical record.

Step 1.2 EMR Division. The local server divides each EMR into EMR light data items and EMR heavy data items, which are respectively denoted by

```
EMR_L = (dataL[i][1], dataL[i][2], ...)
EMR_H = (dataH[i][1], dataH[i][2], ...)
```

, where `dataL[i][j]` $\in$ `EMR` and `dataH[i][j]` $\in$ `EMR`, so `EMR` = `EMR_L` + `EMR_H`. Then, all the EMR light data items are stored in the form of plaintext to a local database (called an EMR plaintext light database) of the local server.

Here, each EMR light data item is of small capacity, whose data type is numeric type (such as examination date and patient age) or short text type (such as patient name and doctor name). Generally, most of EMR related query operations are defined on light data items. For example, querying all electronic medical records from the patients 'Zhang San' (patient name) who were seen in the respiratory department between January and December 2022. Here, each EMR heavy data item is of large capacity, whose data type is long text type (such as diagnosis description) or multimedia picture type (such as medical pictures), and it generally does not directly support EMR related query operations (i.e., each EMR query conditional item is generally unrelated to EMR heavy data items), but it is the key factor to determine the capacity of an electronic medical record. According to statistics, the EMR heavy data usually accounts for about 90% of a total EMR database capacity. Therefore, storing EMR heavy data on the cloud can well reduce the pressure of local data storage and management, and then the management cost and efficiency of medical data.

Step 1.3 Heavy Encryption. The local server randomly generates a secret key (denoted by `keyH[i]`), and then uses a traditional encryption algorithm to encrypt the EMR heavy data to obtain the EMR ciphertext heavy data, denoted by

```
EMR_E = (
E(keyH[i], dataH[i][1]),
E(keyH[i], dataH[i][2]), ...)
```

, where **E** denotes an encryption function. Then, the local server submits the EMR heavy data in the form of ciphertext to the cloud server for storage.

Step 1.4 Heavy Storage. The cloud server stores the EMR heavy data in the form of ciphertext submitted from the local server to the cloud database (called an EMR ciphertext heavy database).

Now, each EMR submitted by an EMR input clerk is divided into light data items (`EMR_L`) and heavy data items (`EMR_H`), and stored into the EMR plaintext light table (`EMR_L_Table` consisting of `EMR_L`s) on the local server and the EMR ciphertext heavy database on the cloud server (`EMR_E_Table` consisting of the ciphertext of `EMR_H`s). Let `EMR_H_Table` denote the plaintext corresponding to `EMR_E_Table`. Then, we have that `EMR_Table` = `EMR_L_Table` + `EMR_H_Table`.

3.3. Query Model

The EMR segmentation query model corresponds to Steps 3 to 6 shown in Figure 2, whose workflow can be further described by Figure 4 (where N_1 or N_2 in curly brackets represents the size of a dataset). It can be seen that the EMR segmentation query model includes the following six steps.

Step 2.1 Issue Query. An EMR query clerk submits an EMR query request through a trusted query interface to the local server. Let **q** [`EMR_Table`] represent an EMR query statement, where **q** is the query itself, and `EMR_Table` is the data table associated with **q**. Here, an EMR query statement consists of a group of basic conditional items defined on EMR data items, where one conditional item is generally associated with only one data item. For example, for querying all electronic medical records from the patients 'Zhang San' (patient name) who were seen in the respiratory department between January and December 2022, its SQL statement can be described as follows

```
SELECT * FROM EMR_Table EMR WHERE
EMR.patientName = 'Zhang San' AND
EMR.departmentName = 'Respiratory' AND
EMR.seeDate
BETWEEN '2022-01-01' AND '2022-12-31'
```

Step 2.2 Rewrite Query. The local server rewrites the EMR query statement **q** [`EMR_Table`] into an EMR light query clause (consisting of all the conditional items defined over EMR light data items) and an EMR heavy query clause (consisting of all the conditional items defined over EMR heavy data items), which are respectively denoted by **qL** [`EMR_L_Table`] and **qH** [`EMR_H_Table`].

Note that each EMR query is generally unrelated to EMR heavy data items. Thus, the above EMR heavy query clause is generally empty, and the above EMR light data query clause is equal to the original EMR query statement (i.e., **qL** = **q** and **qH** = $\oslash$), and only the query table is changed from an EMR data table (`EMR_Table`) to an EMR light data table (`EMR_L_Table`).

Step 2.3 Light Query. The local server performs the light data query clause **qL** on the EMR plaintext light data table (`EMR_L_Table`) to obtain an EMR light dataset (denoted by `EMR_L_Dataset`, and the set composed of the ID attributes of all records is denoted by `EMR_L_Dataset.ID`). Then, the local server generates a heavy data query statement on the EMR ciphertext heavy data table (`EMR_E_Table`) accordingly, which is denoted by **qE** [`EMR_E_Table`]. Finally, the local server submits it to the cloud database for execution. Here, the SQL heavy data query statement corresponding the previous example can be described as follows

```
SELECT * FROM EMR_E_Table WHERE
EMR_E_Table.ID IN EMR_L_Dabaset.ID
```

Step 2.4 Ciphertext Query. The cloud server performs the ciphertext heavy data query statement **qE** submitted by the local server on the EMR ciphertext heavy database `EMR_E_Table`, to obtain the EMR heavy dataset in the form of ciphertext (denoted by `EMR_E_Dataset`), and then return it to the local server.

Step 2.5 Heavy Query. Based on the secret keys stored locally, the local server first decrypts the EMR heavy dataset (`EMR_E_Dataset`) returned from the cloud server. Then, the local server executes the EMR heavy data query clause **qH** generated in Step 2.2 (if the heavy data query clause is not empty) on the decrypted dataset, to obtain the EMR heavy dataset (denoted by `EMR_H_Dataset`) in plaintext.

Step 2.6 Join Query. The local server performs an equivalent join query on the EMR light dataset (`EMR_L_Dataset`) and the EMR heavy dataset (`EMR_H_Dataset`), to obtain a target EMR dataset and return it to the client. Here, for the heavy-light data join query, its SQL statement corresponding the previous example can be described as follows

```
SELECT * FROM
EMR_L_Dataset, EMR_H_Dataset WHERE
EMR_L_Dataset.ID = EMR_H_Dataset.ID
```

It can be seen that for a given EMR query statement submitted from client, the EMR cloud segmentation query model, by performing EMR related query operations through the collaborative interaction between the local server and the cloud server, can ensure the accuracy and efficiency of the query statement.

4. Analysis

In this section, we analyze the impact of the proposed solution on an existing EMR management platform on the cloud in terms of confidentiality, accuracy, efficiency and availability.

Observation 1: The proposed solution can well ensure the confidentiality of EMRs, specifically, including that:

(1) It can well ensure the confidentiality of EMR light data on the trusted local server.
(2) It can well ensure the confidentiality of EMR heavy data on the untrusted cloud server.

Explain: (1) The local server is considered as honest and credible, i.e., its threat comes from outside (not itself), so it is trusted. However, the confidentiality of EMR light data on the trusted local server can be well ensured by the traditional data security strategies deployed by the EMR light database (such as identity authentication and access control). Therefore, the confidentiality of EMR light data can be well ensured on the trusted local server. (2) The cloud server is considered as honest but curious, which means that it follows cloud service-related protocol specifications, but remains curious about EMR data submitted from clients. Thus, the cloud server is not trusted, and its security threat comes from not only outside but also inside (i.e., itself). The confidentiality of EMR heavy data stored on the untrusted cloud can be well ensured by the traditional data encryption technology, because the EMR heavy data stored in the cloud database has been strictly encrypted by the local server and stored in the form of ciphertext, where the key is stored on the trusted local (i.e., it cannot be obtained by the cloud). However, the effectiveness of traditional encryption technology has been demonstrated by a lot of practice, i.e., without secret key, it is almost impossible for an attacker to directly know plaintext from ciphertext. Therefore, the confidentiality of EMR heavy data can be well ensured on the untrusted cloud server. (**End**)

Observation 2: The proposed solution can well ensure the accuracy of each EMR related query statement, i.e., before and after the introduction of the proposed solution, the execution result of each original EMR query statement in the EMR cloud management system remains unchanged.

Explain: As can be seen from Figure 3, the proposed solution divides the original table (`EMR_Table`) of electronic medical records of a medical management system into a light data table (`EMR_L_Table`) and a heavy data table (`EMR_H_Table`), whose sizes are the same to each other, and `EMR_Table` = `EMR_L_Table` + `EMR_H_Table`.

As can be seen from Figure 4, the proposed solution divides each EMR related query statement (**q** [`EMR_Table`]) into a light query clause defined over the light data table (**qL** [`EMR_L_Table`]), and a heavy query clause defined over the heavy data table (**qH** [`EMR_H_Table`]). It can be seen that the result of performing the two query clauses on the EMR table is consistent with that of performing the original EMR query on the EMR table, i.e.,

q [`EMR_Table`] = **qH** [`EMR_Table`] + **qL** [`EMR_Table`]

However, since the light query clause is only related to the EMR light data items, and the heavy query clause is only related to the EMR heavy data items, we have that:

qH [`EMR_Table`] = **qH** [`EMR_H_Table`]
qL [`EMR_Table`] = **qL** [`EMR_L_Table`]

Therefore, for the execution result of the light query clause on the light data table, and the execution result of the heavy query clause on the heavy data table, their join result is certainly consistent with the result of performing the two query clauses on the EMR table, and then the result of performing the original EMR query statement on the original EMR table, i.e.,

q[`EMR_Table`] = **qL**[`EMR_L_Table`] + **qH**[`EMR_H_Table`]

Now, we conclude that the solution can well ensure the accuracy of each EMR related query statement. (**End**)

Observation 3: The proposed solution can well ensure the efficiency of each EMR related query statement, i.e., before and after the introduction of the proposed solution, the execution efficiency of each original EMR query statement in the EMR cloud management system should not be significantly reduced.

Explain: Before the introduction of the proposed solution, the execution time of an EMR query statement on the cloud database is denoted by $\Gamma[\mathbf{q}]$, and the transmission time of medical record records returned from the cloud to the client is denoted by $N_2 \cdot \Gamma[\mathbf{e}]$ (where $\Gamma[\mathbf{e}]$ denotes the network transmission time of one record, and N_2 denotes the number of returned records). After the introduction of the proposed solution, the total execution time of three related query statements (see Figure 4) is denoted by $\Gamma[\mathbf{qL}] + \Gamma[\mathbf{qE}] + \Gamma[\mathbf{qH}]$, and the transmission and decryption time of returned medical records is denoted by $N_1 \cdot (\Gamma[\mathbf{e}] + \Gamma[\mathbf{d}])$ (where $\Gamma[\mathbf{d}]$ denotes the decryption time of one record, and N_1 denotes the number of records returned from the cloud to the local server). Then, according to the comparison before and after the introduction of the proposed solution, we know that the execution efficiency ratio of an EMR query statement can be calculated as follows

$$\frac{\Gamma[\mathbf{q}] + N_2 \cdot \Gamma[\mathbf{e}]}{\Gamma[\mathbf{qL}] + \Gamma[\mathbf{qE}] + \Gamma[\mathbf{qH}] + N_1 \cdot (\Gamma[\mathbf{e}] + \Gamma[\mathbf{d}])} \tag{1}$$

Since the total execution time of three related query statements is basically in the same order of magnitude with the execution time of the original query statement, we have that

$$\Gamma[\mathbf{qL}] + \Gamma[\mathbf{qE}] + \Gamma[\mathbf{qH}] = \alpha \cdot \Gamma[\mathbf{q}] \tag{2}$$

In the above formula, $1 < \alpha < 3$. Note that each conditional item of an EMR query statement is generally defined on an EMR light data item (not on an EMR heavy data item). Thus, we have that $N_2 < N_1 \wedge N_2 \approx N_1$. Also, it is noted that the decryption and transmission time of one record in the form of ciphertext is almost fixed. Therefore, we have that

$$N_1 \cdot (\Gamma[\mathbf{e}] + \Gamma[\mathbf{d}]) = \beta \cdot N_2 \cdot \Gamma[\mathbf{e}] \tag{3}$$

In the above formula, $1 < \beta < 2$. After substituting Formulas (2) and (3) into Formula (1), we have that

Table 1: An experimental evaluation result of the efficiency of EMR query operations

EMR Capacity	Heavy Ratio	$N_2 = 2$	$N_2 = 4$	$N_2 = 8$	$N_2 = 16$	$N_2 = 32$	$N_2 = 62$	$N_2 = 128$
10 GB	88%	0.5747	0.5426	0.5164	0.4987	0.4881	0.4824	0.4793
10 GB	90%	0.5682	0.5344	0.5069	0.4884	0.4775	0.4715	0.4684
10 GB	92%	0.5618	0.5263	0.4977	0.4786	0.4673	0.4611	0.4579
20 GB	88%	0.5263	0.4977	0.4786	0.4673	0.4611	0.4579	0.4562
20 GB	90%	0.5216	0.4924	0.4729	0.4614	0.4551	0.4518	0.4501
20 GB	92%	0.517	0.4872	0.4673	0.4556	0.4492	0.4459	0.4442
30 GB	88%	0.502	0.4762	0.4615	0.4533	0.449	0.4467	0.4456
30 GB	90%	0.495	0.4708	0.4559	0.4476	0.4432	0.4409	0.4398
30 GB	92%	0.4902	0.4655	0.4505	0.442	0.4375	0.4352	0.4341

$$\frac{\Gamma[\mathbf{q}] + N_2 \cdot \Gamma[\mathbf{e}]}{\Gamma[\mathbf{qL}] + \Gamma[\mathbf{qE}] + \Gamma[\mathbf{qH}] + N_1 \cdot (\Gamma[\mathbf{e}] + \Gamma[\mathbf{d}])} = \frac{\Gamma[\mathbf{q}] + N_2 \cdot \Gamma[\mathbf{e}]}{\alpha \cdot \Gamma[\mathbf{q}] + \beta \cdot N_2 \cdot \Gamma[\mathbf{e}]} > \frac{1}{3} \tag{4}$$

Now, we can conclude that before and after the introduction of the proposed solution, the execution efficiency of an EMR related query statement is in the same order of magnitude, i.e., after the introduction of the proposed solution, the execution efficiency of each EMR related query statement on an EMR cloud management system should not be significantly reduced. (**End**)

To further evaluate the impact of the propose solution on the query efficiency, we also designed a simple set of experiments. In the experiments, the capacity of an EMR database was set to about 20 GB, the number of electronic medical records in the database was set to about 20,000, the cloud server uses Inspur SA5212M5, and the local server uses a desktop computer. The experimental evaluation results are shown in Table 1, where the heavy ratio factor refers to the ratio of the heavy data capacity to the total EMR database capacity. Figure 5 is a graphical description of Table 1, where the abscissa represents the size of an EMR query dataset (i.e., N2 in Formula (1)), the ordinate represents the performance efficiency ratio of an EMR related query statement (i.e., Formula (1)), and each subgraph represents the capacity of an EMR database. As can be seen from Table 1 and Figure 5, the experimental results once again demonstrates the previous theoretical analysis result, i.e., after the introduction of the solution, the execution efficiency of each EMR related query statement should not be significantly reduced.

Observation 4: The proposed solution can well ensure the availability of an original EMR cloud management system, specifically, including that:

(1) It requires no change to each user interface program of the original system (such as EMR input interface and EMR query interface) and users' habits.
(2) It requires no change to the architecture of the cloud server and EMR service algorithm.
(3) It requires no change to the accuracy and efficiency of each EMR related query statement of the original system.

Explain: (1) As can be seen from Figures 1 to 3, the proposed solution is deployed on a local server, which requires no change to each user interface program (only needs to redirect each client request from the cloud server and to the local server) and user usage habits. (2) As can be seen from Figure 3, although the solution requires the cloud server to modify its database schema (so as to store ciphertext medical records and related heavy data), it requires no change to the cloud database management system and the information service algorithm, so it requires no change to the architecture of the cloud server. (3) From Observations 2 and 3, we have known that the solution does not impact the accuracy and efficiency of each EMR query statement in the original system. Therefore, the solution can well ensure the availability of an original EMR cloud management system. (**End**)

Based on Observation 4, we can conclude that the solution proposed in this paper for confidentiality management of EMRs on the cloud is constructed on top of an original EMR cloud management system. It does not require secondary development for the original system, so it can realize a seamless connection with the original system, resulting in good actual availability. However, the biggest problem of the proposed solution is the introduction of a local server, and it not only needs to store some medical record data, but also needs to perform some query statements, which may result in a storage performance bottleneck or a computation performance bottleneck.

Observation 5: For the proposed solution, although it requires the deployment of a local server, it requires not high storage performance and computation performance of the local server, i.e., the solution would not cause a storage performance bottleneck or a computation performance bottleneck.

Explain: (1) The storage performance analysis for the local server. The local server only stores EMR light data (while EMR heavy data are stored on the cloud). A large number of statistics shows that EMR light data usually accounts for only about 10% of a total EMR database capacity. For 1024 GB of EMR data, the local database only needs to cost 100 GB of stor-

Table 2: The effectiveness comparison between our solution and other related ones (where $\surd$ denotes "good", $\odot$ "general" and $\otimes$ "not good")

Candidates	Confidentiality	Accuracy	Efficiency	Availability
Our Proposed Solution	$\surd$	$\surd$	$\surd$	$\surd$
Identity Authentication [26, 28]	$\otimes$	$\surd$	$\surd$	$\surd$
Access Control [31, 32, 35]	$\otimes$	$\surd$	$\surd$	$\surd$
Encryption [37, 39, 42]	$\surd$	$\surd$	$\surd$	$\otimes$
Blockchain [48, 49, 50]	$\surd$	$\surd$	$\otimes$	$\otimes$
Watermarking [52]	$\surd$	$\surd$	$\odot$	$\otimes$

age space (the rest of the data are stored on the cloud). Thus, the solution leads to no storage performance bottleneck for the local server. (2) The computing performance analysis for the local server. As can be seen from Figure 4, the calculation required by the local server mainly includes a light data query operation and a heavy data decryption operation. For the former, because the storage capacity of an EMR light database is smaller, so the time overhead of its query operation is not high. For the latter, according to the previous analysis, the number of ciphertext medical records returned from the cloud to the local server is basically equal to the number of the records finally returned to a client, thus the number of ciphertext medical records required to be decrypted by the local server is usually not large, i.e., the time cost required for decrypting the ciphertext medical records is not high. In summary, the solution leads to no computing performance bottleneck for the local server. (**End**)

According to the above observations, we conclude that the proposed solution can well improve the confidentiality of EMR data on the untrusted cloud, under the premises of not compromising the availability of an original medical information system, and not compromising the accuracy and efficiency of each original EMR query statement. In addition, the proposed solution does not cause the storage performance bottleneck or computing performance bottleneck for the local server. In summary, the solution is simple and easy to use, which can realize an effective connection with an existing EMR cloud management system (without the need for a secondary development of the existing system), so can better meet the actual needs of the availability and confidentiality of EMR management under a cloud computing environment.

In addition, from the related work section, we have known that: (1) for identity authentication and access control, although it has good availability, it is difficult to resist the internal threat from the cloud, consequently, making it unable to meet the constraint of confidentiality; (2) for data encryption, although it has good confidentiality, it cannot meet the constraints of efficiency and confidentiality; and (3) for a solution based on blockchain or watermarking, it may has good confidentiality, but it requires to reconstruct the architecture of an existing EMR cloud management system, consequently, making it unable to meet the constraint of availability. In Table 2, we present a brief comparison between our proposed solution and other related ones, where $\surd$ denotes "good", $\odot$ "general" and $\otimes$ "not good". From the table, we see that compared with others, our solution has better overall performance in terms of confidentiality, accuracy, efficiency and availability, which demonstrates again that our solution can well meet the four constraints presented in Section 3.1. At last, it should be noted that the solution proposed in this paper is targeted for the confidentiality management of electronic medical records on the cloud, but the solution can be transferred to other problems of data confidentiality management as well, such as digital libraries [59, 60, 61, 62], archives management [63, 64, 65], location-based services [66, 67, 68, 69], personalized information services [70, 71, 72, 73], multimedia data management [74, 75, 76, 77], knowledge management [78, 79, 80, 81, 82], and series management [83, 84, 85, 86, 87].

5. Conclusion

In this paper, we propose an effective solution to the confidentiality management of the electronic medical records (EMR) on the cloud. Its basic idea is to deploy a trusted local server between the untrusted cloud and the trusted clients, responsible for running an EMR hierarchical storage model and an EMR segmentation query model. The former is responsible for storing EMR light data on the local server, while encrypting and storing EMR heavy data to the cloud server to ensure the confidentiality of EMRs. The latter is responsible for performing each EMR related query operation through the collaborative interaction between the local server and the cloud, to ensure the accuracy and efficiency of each EMR query statement. Finally, both theoretical analysis and experimental evaluation demonstrate the effectiveness of the proposed solution: (1) the solution can effectively improve the security of electronic medical records on the untrusted cloud without compromising the availability of an existing medical information system; and (2) the solution requires small storage performance and computation performance for the local server, to avoid causing the performance bottleneck. Therefore, the proposed solution is simple and easy to use, which can realize an effective connection with an existing EMR management platform on the cloud, consequently, meeting to the practical application requirements of confidentiality management of EMRs under a cloud environment in terms of security, availability, efficiency and accuracy.

Reference

References

[1] Liu J, Wang X, Shen S, et al. Intelligent jamming defense using DNN Stackelberg game in sensor edge cloud[J]. IEEE Internet of Things Journal, 2021

[2] Liu J, Wang X, Shen S, et al. A bayesian q-learning game for dependable task offloading against ddos attacks in sensor edge cloud[J]. IEEE Internet of Things Journal, 2020, 8(9): 7546-7561.

[3] Shen S, Huang L, Zhou H, et al. Multistage signaling game-based optimal detection strategies for suppressing malware diffusion in fog-cloud-based IoT networks[J]. IEEE Internet of Things Journal, 2018, 5(2): 1043-1054.

[4] Thapa C, Camtepe S. Precision health data: Requirements, challenges and existing techniques for data security and privacy[J]. Computers in Biology and Medicine, 2021, 129: 104130.

[5] Sahi A, Lai D, Li Y. Security and privacy preserving approaches in the eHealth clouds with disaster recovery plan[J]. Computers in Biology and Medicine, 2016, 78: 1-8.

[6] Cui Z, Wu Z, Zhou C, et al. An efficient subscription index for publication matching in the cloud[J]. Knowledge-Based Systems, 2016, 110: 110-120.

[7] Heidari A, Toumaj S, Navimipour N J, et al. A privacy-aware method for COVID-19 detection in chest CT images using lightweight deep conventional neural network and blockchain[J]. Computers in Biology and Medicine, 2022, 145: 105461.

[8] Remeseiro B, Bolon-Canedo V. A review of feature selection methods in medical applications[J]. Computers in Biology and Medicine, 2019, 112: 103375.

[9] Wu Z, Xu G, Lu C, et al. An effective approach for the protection of privacy text data in the CloudDB[J]. World Wide Web, 2018, 21(4): 915-938.

[10] Mei Z, Zhu H, Cui Z, et al. Executing multi-dimensional range query efficiently and flexibly over outsourced ciphertexts in the cloud[J]. Information Sciences, 2018, 432: 79-96.

[11] Wu Z, Shen S, Lian X, et al. A dummy-based user privacy protection approach for text information retrieval[J]. Knowledge-Based Systems, 2020, 195: 105679.

[12] Wu Z, Shen S, Li H, et al. A basic framework for privacy protection in personalized information retrieval[J]. Journal of Organizational and End User Computing, 2021, 33(6): 1-26.

[13] Wu Z, Xie J, Zheng C, et al. A framework for the protection of user behavior preference privacy of digital library[J]. Journal of Library Science in China, 2018, 44(2): 72-85.

[14] Chen Z, Xu W, Wang B, et al. A blockchain-based preserving and sharing system for medical data privacy[J]. Future Generation Computer Systems, 2021, 124: 338-350.

[15] Lu C, Wu Z, Liu M, et al. A patient privacy protection scheme for medical information system[J]. Journal of Medical Systems, 2013, 37(6): 1-10.

[16] Cui W, Ye J. Logarithmic similarity measure of dynamic neutrosophic cubic sets and its application in medical diagnosis[J]. Computers in Industry, 2019, 111: 198-206.

[17] Nosouhi M R, Yu S, Sood K, et al. UCoin: An Efficient Privacy Preserving Scheme for Cryptocurrencies[J]. IEEE Transactions on Dependable and Secure Computing, 2021.

[18] Ye J, Cui W. Modeling and stability analysis methods of neutrosophic transfer functions[J]. Soft Computing, 2020, 24(12): 9039-9048.

[19] Cui W, Ye J, Fu J. Cotangent similarity measure of single-valued neutrosophic interval sets with confidence level for risk-grade evaluation of prostate cancer[J]. Soft Computing, 2020, 24(24): 18521-18530.

[20] Zhang H, Shen S, Cao Q, et al. Modeling and analyzing malware diffusion in wireless sensor networks based on cellular automaton[J]. International Journal of Distributed Sensor Networks, 2020, 16(11): 1550147720972944.

[21] Cheng Z, Yue D, Shen S, et al. Secure frequency control of hybrid power system under DoS attacks via lie algebra[J]. IEEE Transactions on Information Forensics and Security, 2022, 17: 1172-1184.

[22] Li H, Zhu Y, Wang J, et al. Consensus of nonlinear second-order multi-agent systems with mixed time-delays and intermittent communications[J]. Neurocomputing, 2017, 251: 115-126.

[23] Liu J, Shen S, Yue G, et al. A stochastic evolutionary coalition game model of secure and dependable virtual service in sensor-cloud[J]. Applied Soft Computing, 2015, 30: 123-135.

[24] Feng S, Shi H, Huang L, et al. Unknown hostile environment-oriented autonomous WSN deployment using a mobile robot[J]. Journal of Network and Computer Applications, 2021, 182: 103053.

[25] Feng S, Wu C, Zhang Y, et al. WSN deployment and localization using a mobile agent[J]. Wireless Personal Communications, 2017, 97(4): 4921-4931.

[26] Kumari A, Jangirala S, Abbasi M Y, et al. ESEAP: ECC based secure and efficient mutual authentication protocol using smart card[J]. Journal of Information Security and Applications, 2020, 51: 102443.

[27] Shen Y, Shen S, Li Q, et al. Evolutionary privacy-preserving learning strategies for edge-based IoT data sharing schemes[J]. Digital Communications and Networks, 2022.

[28] Shen Y, Shen S, Wu Z, et al. Signaling game-based availability assessment for edge computing-assisted IoT systems with malware dissemination[J]. Journal of Information Security and Applications, 2022, 66: 103140.

[29] Abuarqoub A. D-FAP: dual-factor authentication protocol for mobile cloud connected devices[J]. Journal of Sensor and Actuator Networks, 2019, 9(1): 1.

[30] Wang T, Bhuiyan M Z A, Wang G, et al. Preserving balance between privacy and data integrity in edge-assisted Internet of Things[J]. IEEE Internet of Things Journal, 2019, 7(4): 2679-2689.

[31] Wu B, Chen X, Wu Z, et al. Privacy-guarding optimal route finding with support for semantic search on encrypted graph in cloud computing scenario[J]. Wireless Communications and Mobile Computing, 2021, 2021.

[32] Wu B, Zhao Z, Cui Z, et al. Secure and efficient adjacency search supporting synonym query on encrypted graph in the cloud[J]. IEEE Access, 2019, 7: 133716-133724.

[33] Wang S, Cong Y, Zhu H, et al. Multi-scale context-guided deep network for automated lesion segmentation with endoscopy images of gastrointestinal tract[J]. IEEE Journal of Biomedical and Health Informatics, 2020, 25(2): 514-525.

[34] Mocrii D, Chen Y, Musilek P. IoT-based smart homes: A review of system architecture, software, communications, privacy and security[J]. Internet of Things, 2018, 1: 81-98.

[35] Fan C, Hu K, Feng S, et al. Heronian mean operators of linguistic neutrosophic multisets and their multiple attribute decision-making methods[J]. International Journal of Distributed Sensor Networks, 2019, 15(4): 1550147719843059.

[36] Cao Y, Sun Y, Min J. Hybrid blockchainCbased privacy-preserving electronic medical records sharing scheme across medical information control system[J]. Measurement and Control, 2020, 53(7-8): 1286-1299.

[37] Wu Z, Shi J, Lu C, et al. Constructing plausible innocuous pseudo queries to protect user query intention[J]. Information Sciences, 2015, 325: 215-226.

[38] Liu J, Yu J, Shen S. Energy-efficient two-layer cooperative defense scheme to secure sensor-clouds[J]. IEEE Transactions on Information Forensics and Security, 2017, 13(2): 408-420.

[39] Li T, Wang H, He D, et al. Blockchain-based privacy-preserving and rewarding private data sharing for IoT[J]. IEEE Internet of Things Journal, 2022.

[40] Li Q, Zhang Q, Huang H, et al. Secure, efficient and weighted access control for cloud-assisted industrial IoT[J]. IEEE Internet of Things Journal, 2022.

[41] Li T, Wang H, He D, et al. Synchronized provable data possession based on blockchain for digital twin[J]. IEEE Transactions on Information Forensics and Security, 2022.

[42] Renardi M B, Basjaruddin N C, Rakhman E. Securing electronic medical record in near field communication using advanced encryption standard (AES)[J]. Technology and Health Care, 2018, 26(2): 357-362.

[43] Dai Y, Wu J, Fan Y, et al. MSEva: A musculoskeletal rehabilitation evaluation system based on EMG signals[J]. ACM Transactions on Sensor Networks, 2022.

[44] Wu Z, Li G, Liu Q, et al. Covering the sensitive subjects to protect personal privacy in personalized recommendation[J]. IEEE Transactions on Services Computing, 2016, 11(3): 493-506.

[45] Zhang S, Ren W, Tan X, et al. Semantic-aware dehazing network with adaptive feature fusion[J]. IEEE Transactions on Cybernetics, 2021.

[46] Qiu W, Xie J, Shen Y, et al. Endoscopic image recognition method of gastric cancer based on deep learning model[J]. Expert Systems, 2022, 39(3): e12758.

[47] Kumar P R, Raj P H, Jelciana P. Exploring data security issues and solu-

tions in cloud computing[J]. Procedia Computer Science, 2018, 125: 691-697.
[48] Fu J, Wang N, Cai Y. Privacy-preserving in healthcare blockchain systems based on lightweight message sharing[J]. Sensors, 2020, 20(7): 1898.
[49] Esposito C, De Santis A, Tortora G, et al. Blockchain: A panacea for healthcare cloud-based data security and privacy?[J]. IEEE Cloud Computing, 2018, 5(1): 31-37.
[50] Jin H, Luo Y, Li P, et al. A review of secure and privacy-preserving medical data sharing[J]. IEEE Access, 2019, 7: 61656-61669.
[51] Lv Z, Piccialli F. The security of medical data on internet based on differential privacy technology[J]. ACM Transactions on Internet Technology, 2021, 21(3): 1-18.
[52] Fang L, Yin C, Zhu J, et al. Privacy protection for medical data sharing in smart healthcare[J]. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), 2020, 16(3s): 1-18.
[53] Vimalachandran P, Liu H, Lin Y, et al. Improving accessibility of the Australian My Health Records while preserving privacy and security of the system[J]. Health Information Science and Systems, 2020, 8(1): 1-9.
[54] Adamu J, Hamzah R, Rosli M M. Security issues and framework of electronic medical record: A review[J]. Bulletin of Electrical Engineering and Informatics, 2020, 9(2): 565-572.
[55] Keshta I, Odeh A. Security and privacy of electronic health records: Concerns and challenges[J]. Egyptian Informatics Journal, 2021, 22(2): 177-183.
[56] Hathaliya J J, Tanwar S. An exhaustive survey on security and privacy issues in Healthcare 4.0[J]. Computer Communications, 2020, 153: 311-335.
[57] Sun W, Cai Z, Li Y, et al. Data processing and text mining technologies on electronic medical records: a review[J]. Journal of Healthcare Engineering, 2018, 2018.
[58] Wang L, Alexander C A. Big data analytics in medical engineering and healthcare: methods, advances and challenges[J]. Journal of Medical Engineering & Technology, 2020, 44(6): 267-283.
[59] Wu Z, Shen S, Lu C, et al. How to protect reader lending privacy under a cloud environment: a technical method[J]. Library Hi Tech, 2020.
[60] Wu Z, Xie J, Pan J, et al. An effective approach for the protection of user privacy in a digital library[J]. Libri, 2019, 69(4): 315-324.
[61] Wu Z, Lu C, Zhao Y, et al. The protection of user preference privacy in personalized information retrieval: challenges and overviews[J]. Libri, 2021.
[62] Wu Z, Shen S, Li H, et al. A comprehensive study to the protection of digital library readers' privacy under an untrusted network environment[J]. Library Hi Tech, 2021.
[63] Wu Z, Xie J, Lian X, et al. A privacy protection approach for XML-based archives management in a cloud environment[J]. The Electronic Library, 2019.
[64] Wu Z, Li R, Zhou Z, et al. A user sensitive subject protection approach for book search service[J]. Journal of the Association for Information Science and Technology, 2020, 71(2): 183-195.
[65] Kusumawardhani D, Masyithah D C. Security and privacy cloud storage as a personal digital archive storage media[J]. Record and Library Journal, 2018, 4(2): 167-173.
[66] Wu Z, Wang R, Li Q, et al. A location privacy-preserving system based on query range cover-up for location-based services[J]. IEEE Transactions on Vehicular Technology, 2020, 69(5): 5244-5254.
[67] Wu Z, Li G, Shen S, et al. Constructing dummy query sequences to protect location privacy and query privacy in location-based services[J]. World Wide Web, 2021, 24(1): 25-49.
[68] Wu B, Chen X, Zhang C, et al. Privacy-protection path finding supporting the ranked order on encrypted graph in big data environment[J]. IEEE Access, 2020, 8: 214596-214604.
[69] Zhang P, Gan P, Kumar N, et al. RKD-VNE: Virtual network embedding algorithm assisted by resource knowledge description and deep reinforcement learning in IIoT scenario[J]. Future Generation Computer Systems, 2022.
[70] Wu Z, Shen S, Zhou H, et al. An effective approach for the protection of user commodity viewing privacy in e-commerce website[J]. Knowledge-Based Systems, 2021, 220: 106952.
[71] Wu Z, Zheng C, Xiejian J, et al. An approach for the protection of users book browsing preference privacy in a digital library[J]. The Electronic Library, 2018.
[72] Zhou H, Shen S, Liu J. Malware propagation model in wireless sensor networks under attack defense confrontation[J]. Computer Communications, 2020, 162: 51-58.
[73] Shen S, Zhou H, Feng S, et al. HSIRD: A model for characterizing dynamics of malware diffusion in heterogeneous WSNs[J]. Journal of Network and Computer Applications, 2019, 146: 102420.
[74] Zhao L, Lin T, Zhang D, et al. An ultra-low complexity and high efficiency approach for lossless alpha channel coding[J]. IEEE Transactions on Multimedia, 2019, 22(3): 786-794.
[75] Zhou Q, Zhao L, Zhou K, et al. String prediction for 4: 2: 0 format screen content coding and its implementation in AVS3[J]. IEEE Transactions on Multimedia, 2020, 23: 3867-3876.
[76] Chen L. Road vehicle recognition algorithm in safety assistant driving based on artificial intelligence[J]. Soft Computing, 2021: 1-10.
[77] Wu Z, Xu G, Zhang Y, et al. GMQL: A graphical multimedia query language[J]. Knowledge-Based Systems, 2012, 26: 135-143.
[78] Wu Z, Zhu H, Li G, et al. An efficient Wikipedia semantic matching approach to text document classification[J]. Information Sciences, 2017, 393: 15-28.
[79] Pan J, Zhang C, Wang H, et al. A comparative study of Chinese named entity recognition with different segment representations[J]. Applied Intelligence, 2022: 1-13.
[80] Xu G, Wu Z, Li G, et al. Improving contextual advertising matching by using Wikipedia thesaurus knowledge[J]. Knowledge and Information Systems, 2015, 43(3): 599-631.
[81] Li Q, Li L, Wang W, et al. A comprehensive exploration of semantic relation extraction via pre-trained CNNs[J]. Knowledge-Based Systems, 2020, 194: 105488.
[82] Xu G, Zong Y, Jin P, et al. KIPTC: a kernel information propagation tag clustering algorithm[J]. Journal of Intelligent Information Systems, 2015, 45(1): 95-112.
[83] Li Q, Cao Z, Ding W, et al. A multi-objective adaptive evolutionary algorithm to extract communities in networks[J]. Swarm and Evolutionary Computation, 2020, 52: 100629.
[84] Yan W, Li G, Wu Z, et al. Extracting diverse-shapelets for early classification on time series[J]. World Wide Web, 2020, 23(6): 3055-3081.
[85] Li Q, Cao Z, Zhong J, et al. Graph representation learning with encoding edges[J]. Neurocomputing, 2019, 361: 29-39.
[86] Bai B, Li G, Wang S, et al. Time series classification based on multi-feature dictionary representation and ensemble learning[J]. Expert Systems with Applications, 2021, 169: 114162.
[87] Wu Z, Jiang T, Su W. Efficient computation of shortest absent words in a genomic sequence[J]. Information Processing Letters, 2010, 110(14-15): 596-601.
[88] Wu Z, Xu G, Yu Z, et al. Executing SQL queries over encrypted character strings in the Database-As-Service model[J]. Knowledge-Based Systems, 2012, 35: 332-348.

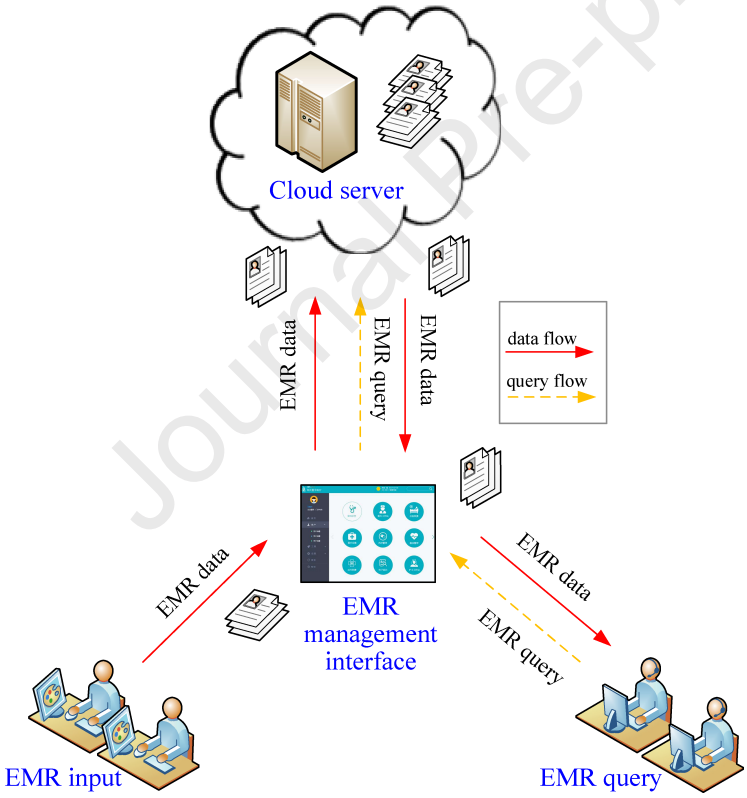

Figure 1: A basic architecture of an EMR management system under a cloud environment

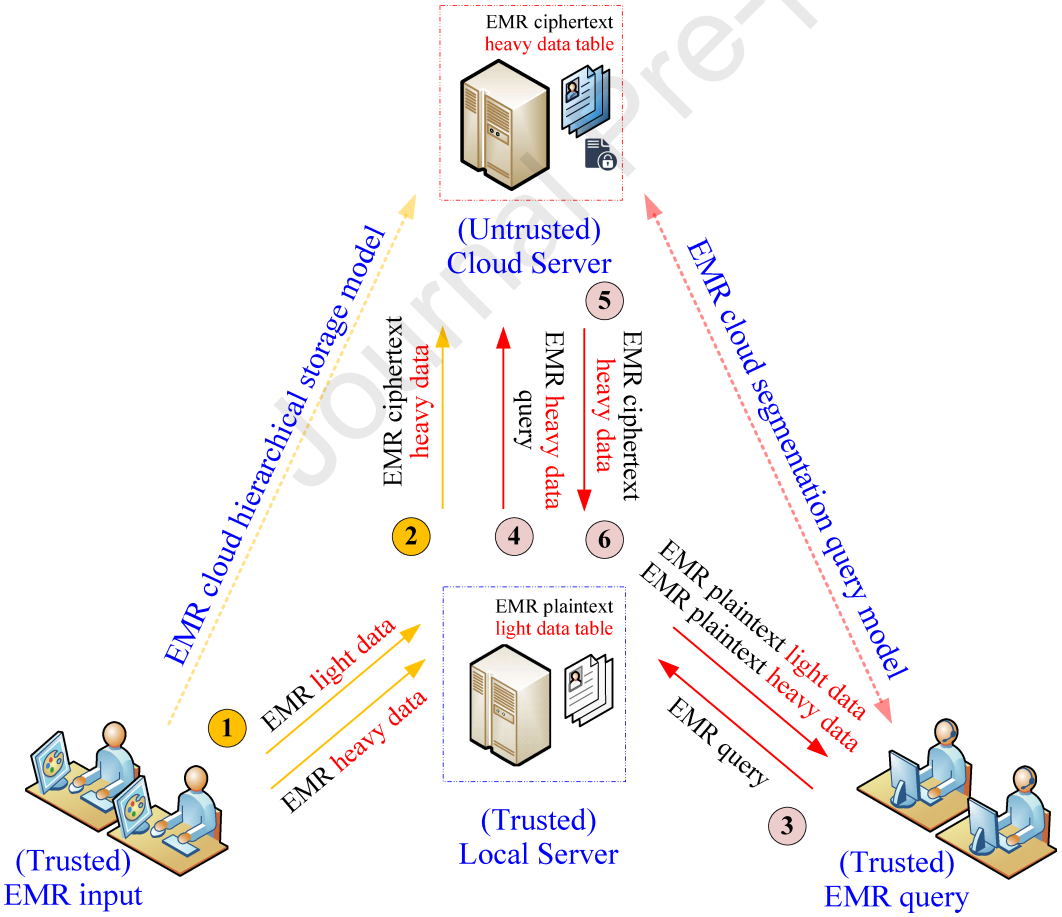

Figure 2: A system framework of confidentiality management of EMRs on the cloud

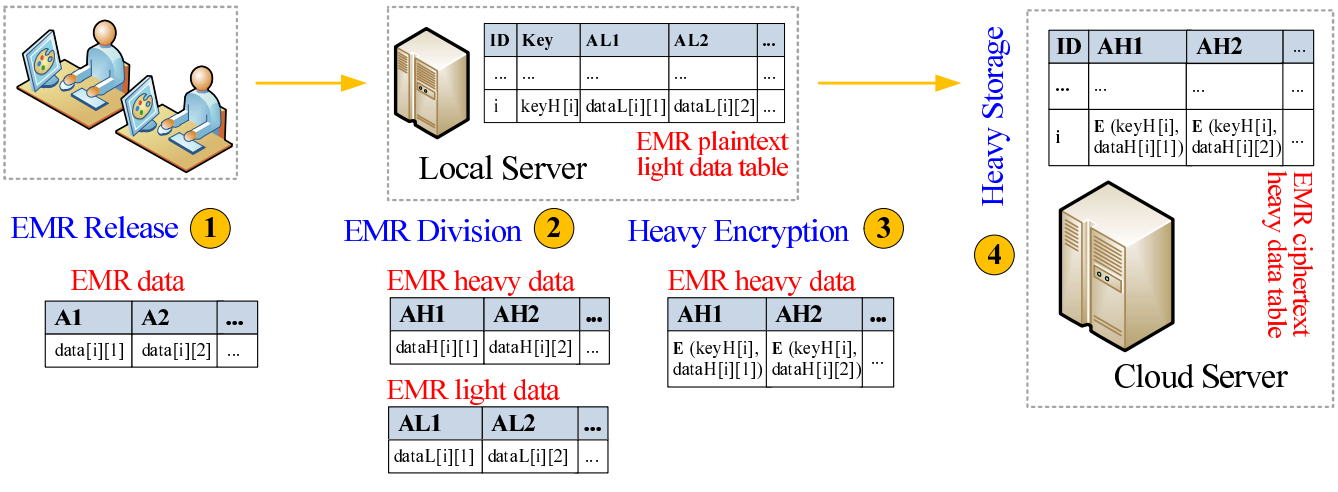

Figure 3: An EMR hierarchical storage model

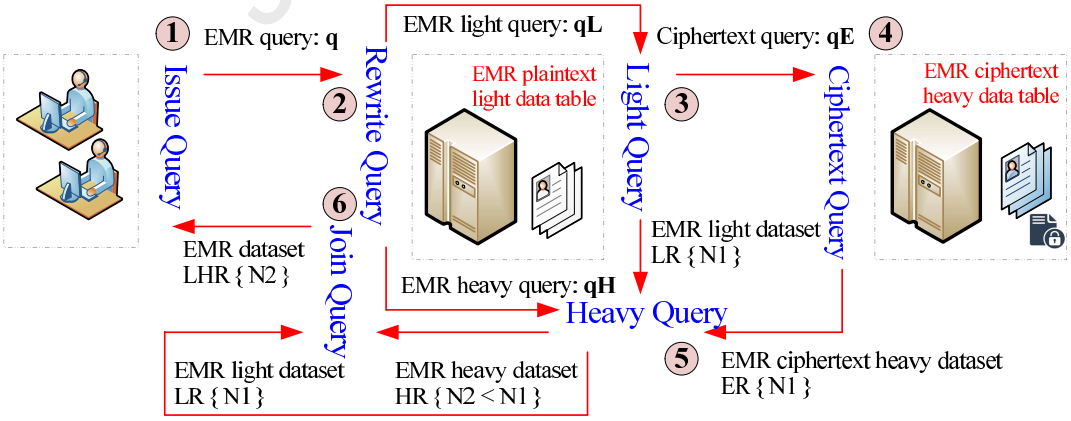

Figure 4: An EMR segmentation query model

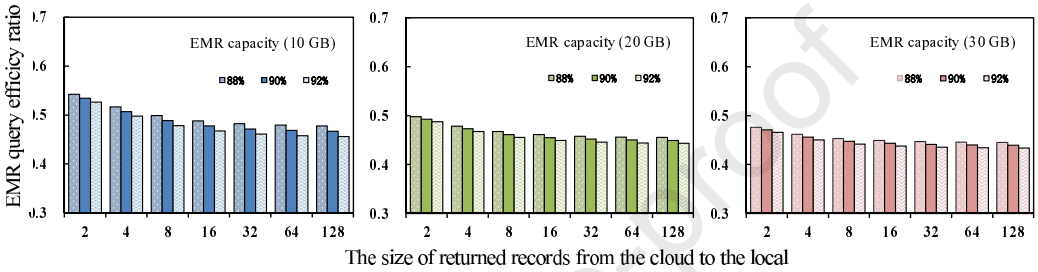

Figure 5: An experimental evaluation result of the efficiency of EMR query operations

-An EMR hierarchical storage model is proposed to ensure the confidentiality of EMRs on the cloud

-An EMR segmentation query model is proposed to ensure the accuracy and efficiency of each EMR query statement

-Both theory analysis and experimental evaluation are performed to demonstrate the performance of the proposed solution.

Conflict of interest

The authors declared that they have no conflicts of interest to this work.