ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Vol 4, Issue 8, August 2017

Avoiding Intrusion and Privacy Protection for Cloudlet-based Medical Data Sharing

[1] Karthika.J, [2] Lavanya.S, [3] JayaSankari.J, [4] Mariya sushmitha.T.F, [5] kanimozhi.R

Abstract— **The popularity of wearable devices with the development of clouds and cloudlet technology, increased the need to provide better medical care. The process of medical data includes data collection, data storage and data sharing,. In the Traditional healthcare system ,it requires the delivery of medical data to the cloud, which involves users' integrated information and causes communication energy consumption. Practically, medical data sharing is a critical and most challenging issue. The functions of cloudlet include privacy protection, data sharing and intrusion detection. In the stage of data collection, we first utilize Number Theory Research Unit (NTRU) method to encrypt user's body data collected by wearable devices. Those data will be transmitted to nearby cloudlet in an energy efficient fashion. Secondly, a new trust model to help users to select trustable partners who want to share stored data in the cloudlet. Thirdly, we divide users' medical data stored in remote cloud of hospital into three parts, and give them proper protection. Lastly , to protect the healthcare system from malicious attacks, we develop a novel collaborative intrusion detection system method based on cloudlet mesh.**

***Index Terms*—data sharing, privacy protection, health care, collaborative intrusion detection system (IDS).**

I. INTRODUCTION

The development of healthcare big data and wearable technology , as well as cloud computing and communication technologies , cloud-assisted healthcare big data computing becomes critical to meet users' evergrowing demands on health consultation. However, it is challenging issue to personalize specific healthcare data for various users in a convenient fashion.. With the advances in cloud computing, a large amount of data can be stored in various clouds, including cloudlets and remote clouds, facilitating data sharing and intensive computations . The body data collected by wearable devices are transmitted to the nearby cloudlet. Those data are further delivered to the remote cloud where doctors can access for disease diagnosis. According to data delivery chain, we separate the privacy protection into three stages. In the first stage, user's vital signs collected by wearable devices are delivered to a closet gateway of cloudlet. In the second stage, user's data will be further delivered toward remote cloud through cloudlets. A cloudlet is formed by a certain number of mobile devices whose owners may require and/or share some specific data contents. Thus, both privacy protection and data sharing are considered in this stage. Especially, we use trust model to evaluate trust level between users to determine sharing data or not. Considering the users' medical data are stored in remote cloud, we classify these medical data into different kinds and take the corresponding security policy. In addition to above three stages based data privacy protection, we also consider collaborative IDS .

II. WORK RELATED

In Lu et al. [4] a system called SPOC, which stands for the secure and privacy-preserving opportunistic computing framework, was proposed to treat the storage problem of healthcare data in a cloud environment and addressed the problem of security and privacy protection under such an environment.

III. SYSTEM FRAMEWORK

The framework of the proposed cloudlet-based healthcare system .• Client data encryption. We utilize the model presented in, and take the advantage of NTRU to protect the client's physiological data from being leaked or abused. This scheme is to protect the user's privacy when transmitting the data from the smartphone to the cloudlet. • Cloudlet based data sharing. Typically, users geographically close to each other connect to the same cloudlet. It's likely for them to share common aspects, • Remote cloud data privacy protection. Compared to user's daily data in cloudlet, the data stored in remote contain larger scale medical data,

• Collaborative IDS based on cloudlet mesh.

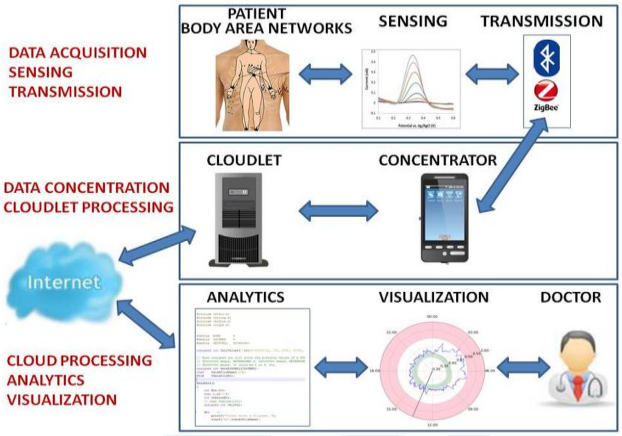
There is a vast volume of medical data stored in the remote cloud, it is critical to apply security mechanism to protect the database from malicious intrusions. In this paper, we develop specific countermeasures to establish a defense system for the large medical database in the remote cloud storage. Specifically, collaborative IDS based on the cloudlet mesh structure is used to screen any visit to the database as a protection border. If the detection shows a malicious intrusion in advance, the collaborative IDS will fire an alarm and block the visit, and vice-versa. The collaborative IDS, as a guard of the cloud database, can protect a vast number of medical data and make sure of the security of the database.

IV.CONTENT SHARING AND PRIVACY PROTECTION

4.1 Encryption at the User End

When using wearable devices to collect users' data, the procedure inevitably involves the user's sensitive information. Therefore, how to effectively collect and transmit users' data under effi- cient privacy protection is a critical problem .A data collection method, called PHDA, is proposed based on data priority which can give proper cost and delay to different priorities data. , Li et al. discuss the process of data collection and utilizes sum aggregation to obtain data to make sure the security of users' privacy in the presence of unreliable sensors. In , Lu et al., study 3V data privacy protection issue based on big data of healthcare. Based on the model presented in , herewe utilizes the advantages of NTRU encryption scheme . NTRU can protect the user's physiological data, such as heart rate, blood pressure and Electrocardiography (ECG), etc, the data collected by smart clothing are all unsigned integer vectors. We need to define clear space and cipher space for the encryption. We hereby describe the processes of encryption and deciphering in the following.

• KeyGen()→ (pk, sk): let $f \in R$, $g \in R$, while f, g follows the discrete Gaussian distribution, $f = 1 \bmod q$, and f is reversible. Thus, the secret key is denoted by sk = f; the public key is denoted by pk = h = g • f −1 mod q. • Enc(pk = h, μ ∈Rp)→ c ∈Rq: let $r \in R$, $m \in R$, $m = \mu \bmod p$. Both m′ and r follow the discrete Gaussian distribution, and we have m = p • m′ + μ, c = p • r • h + m mod q. • Dec(sk = f, c ∈Rq) → μ: calculate $\bar{b}$ = f • c mod q, and make it an integer polynomial b, with factors within [−q/2, q/2). Thus, we have μ = b mod p



4.2 Medical Data Sharing in the Cloudlet

The purpose of medical data sharing is to make better use of data between users. We set the hospital for trusted authority (TA). Assume the user p asks TA to check the data of user q, i.e., user p wants to share data with user q. Then the TA work is divided into the following two steps: Step 1: Compare the similarity of user p and user q. For example, we can utilize the model similar as and use users' data stored in TA, such as EMR, to measure the similarity of user p and user q. Similarity can be divided into three levels, namely Low, moderate and high. Step 2: Describe the trust level between user p and user q. We use the reputation of user p which includes bad, average and good, and the similarity of user p and user q which obtained through step 1, as input data. We can utilize trust model to obtain trust level as follows. • Determine the input and output. The input consists of reputation and similarity and output consists of the corresponding trust level. In order to represent these variables, we quantify each of them as a scalar between 0 and 1. Select a Gaussian function as the corresponding function, which will map the value in the collection into a trust level. • Formulate the relevant guidelines and have the experts set up the trust-related guidelines with the related knowledge and experience. • Build a model that can determine the creditability according to the character, credit, and similarity. After obtaining users' trust level, we can judge whether to trust user p based on threshold value set by user q. If the trust level is equal to or greater than the threshold value, then the user p can be trusted, so TA will share user q information to user p. If the trust level is less than the threshold value, then the user p can not be trust, so TA will refuse the request of the user p. 4.3 Medical Data Privacy Protection in the Cloud Data in remote cloud are generated from the patients treated in the hospital. As the records of diagnosis and payments will be kept in many personal files belonging to a vast number of patients, saving such data in the cloud can reduce costs and be convenient for doctors to diagnose and analyze diseases. Therefore, we shall create a safe environment to ensure that the medical data sharing occurs without risk of leakage. Thus, we shall pay attention to protection of privacy in such data sharing. we can divide the EMR table into the following three types: (i) EID: the properties which can identify the user apparently, e.g., name, phone number, email, home address, and so on; (ii) QID: the property which can identify the user approximately, e.g., a user may be identified according to values such as zip code, date of birth, and gender ; (iii) MI, or some clinical manifestation and disease types. In order to protect the privacy of data and make it convenient for doctors or other patients with a similar disease to access the data, we shall encrypt EID and QID but share MI. Refer to the way of expression , we part the EMR data table A into two independent tables, i.e., a ciphertext table Te and a plaintext table Tp. The ciphertext table contains mainly structural data including the encryption table of EID and QID property; while the plaintext table contains mainly structural and semi-structural data including a clear text table of MI property. We need to protect the shared data and some physiological indexes collected by monitoring the specific diseases. Suppose there are M types of diseases, marked as {D1, D2, . . . , DM}. For each disease Di , there are corresponding characteristics {Ci,1, Ci,2, . . . Ci,in }, i = 1, . . . , M. In order to quantize disease characteristics, we define a question Qi,j for each characteristic Ci,j , i = 1, . . . , M, j = 1, . . . , in. For example, heart Disease exhibits characteristics of dyspnea, palpitation, pectoralgia, etc. For the characteristic of palpitation, we can design the question such as “Do you have palpitation?”. If the query result is ‘1’, then it means yes, otherwise, it means no with the mark of ‘0’. That is to say, there are corresponding test questions {Qi,1, Qi,2, . . . , Qi,in } for each characteristic in {Ci,1, Ci,2, . . . Ci,in } of the corresponding diseases Di , i = 1, 2, . . . , M. For the sake of simplicity, we assume that the answer to each question is 0 or 1. Therefore, each disease Di can acquire its testing results {ei,1, ei,2, . . . , ei,in }, i = 1, . . . , M, with each ei,j = 0 or ei,j = 1. The initial privacy data of users are acquired by completing a survey. In order to be convenient for encryption, we adopt the methods as discussed above to convert these characteristics into numerical data, namely the combination of 0’s and 1’s. We choose a three-tuple {a, b, c} satisfying |a 2 | < |b| < |c| . Then we choose three random numbers {pi ,qi , wi} satisfying the following conditions. pi + qi = bwi , bwi 2 <qi<bwi , a2 buwi< c, (1) where u is integer. After the parameters of a, pi ,qi are obtained, encrypted data can be calculated. Then we have vi = aei + pi , v ′ i = s • qi mod c, v ′ 0 = s • q0 mod c. (2) Therefore, we obtain (a, c, v, v ′) as the encrypted data, which is hard to be decrypted without the secret keys (because of the unknown value of α. Thus, the encryption process of users’ private data is completed.

V. COLLABORATIVE INTRUSION DETECTION

5.1 Collaborative IDS

In this section, collaborative IDS is designed among m IDS, e.t., S1, S2, • •• ,Sm, in order to get higher detection rate and lower false alarm rate. The m IDS are assumed to detect independently. There exists K different types of intrusion. So according to deduce in the following, we can get the detection rate and false alarm rate of collaborative IDS. In order to evaluate it , we give the ROC curve. Before transmitting data to the remote cloud, we establish the collaborative IDS based on the cloudlet mesh to complete the intrusion detection task. We use {S1, S2, . . . ,Sm} to represent the set of IDS’s in the collaborative IDS (CIDS) system. Suppose that each IDS is able to detect intrusion independently. For the sake of simplicity, we use I to indicate that there is intrusion behaviorin this system and NI to indicatethat there is no intrusion. Furthermore, A means that IDS raises an alarm while NA means no alarm. We use 1−β to indicate the detection rate and α as the false alarm rate. If there exists K different types of intrusion, denoted as I1, I2, . . . , IK, then we have I = I1∪I2 • • •∪IK. Assume that the probability of Ij is pj , j = 1, 2, . . . , K. Therefore, the probability of intrusion behavior in this system is p(I) = ∑K i=1 pi , while the probability of no intrusion behavior is P(NI) = 1 − p(I). We thus have that p(A|I) = 1 − β and p(A|NI) = α. As for each IDS, we use p(NAi |Ij) = βij to represent the probability of IDS Si not triggering an alarm when having Ij , and p(Ai |NI) = αi as the probability of Si triggering an alarm when not being attacked. It follows that β = p(NA|I) = p(NA1|I)• • • p(NAm|I). (3) Since Ii ∩ Ij = ϕ, i≠ j, applying the total probability formula, we can obtain the probability that system S1 does not trigger an alarm when there is an attack to intrude the system, as p(NA1|I) = p(NA1 ∩ (I1 ∪ I2 • ••∪ IK)) P(I) = ∑K j=1 β1jpj ∑K j=1 pi . (4) For system Si ,i = 2, 3, . . . , m, let p(NAi |I) denote the probability that no alarm is triggered by Si . We have p(NAi |I) = ∑K j=1 βijpj ∑K j=1 pi . (5) We can derive β as follows. β = ∏m i=1 ∑K j=1 βijpj ∑K j=1 pi . (6) The false alarm rate α = p(A|NI) = 1 − p(NA|NI) can be obtained in a similarly manner, as p(NA|NI) = ∏m i=1 (1 − αi). (7) The false alarm rate α can be computed as follows. α = 1 − ∏m i=1 (1 − αi). (8) We thus obtain the detection rate α and false alarm rate β of the collaborative IDS system. The corresponding ROC curve can be obtained

5.2 Evaluation of collaborative IDS:

We next consider the cost problem of collaborative IDS, with its cost being divided into three parts: • when the intrusion behavior is not detected by the system, but IDS generates an alarm, the system will prevent the transmission of this user's data, which will affect the normal use of the healthcare system by the user, and may lead to decrease of the system's reliability. The cost at this moment is denoted as Cα; • when the system suffers from intrusion Ii , 1 ≤i≤ K, but the IDS does not generate an alarm, the system will allow this intrusive behavior, which will break the healthcare big data; the healthcare data in the remote cloud is attacked and may probably cause leakage of patients' data. The cost of this scenario is denoted as C˜ i , 1 ≤ i ≤ K; • the cost in other scenarios is marked as 0. Without loss of generality, we define the cost rate as Ci = C˜ i/Cα. In the following, we adopt the decision tree to model the corresponding expected cost problem. Let q1, q2 = p(NA) denote the probability of no alarm in a system. Based on the total probability formula, we have q1 = (1 − β) ∑ K t=1 pi + α(1 − ∑ K t=1 pi). (9) q2 = β ∑ K t=1 pi + (1 − α)(1 − ∑ K t=1 pi).

Hereby we formulate an optimization problem based on the decision tree model. That is, under the circumstances of guaranteeing a certain detection rate 1 − β˜ and false alarm rate α, we shall choose the optimal number m, so that we can achieve the minimum expected cost. The formulated problem is given below.

minimizeEc (16)
subject to: α < α, β < ˜ β˜ (10)
0 ≤ pij ≤ 1, i, j = 1, 2, . . . , K (11)
0 ≤qi ≤ 1, i = 1, 2, . . . , K (12)
Cj> 0, j = 1, 2, . . . , K. (13)

VI. CONCLUSIONS

In this paper, we investigated the problem of privacy protection and sharing large medical data in cloudlets and the remote cloud. We developed a system which does not allow users to transmit data to the remote cloud in consideration of secure collection of data, as well as low communication cost. However, it does allow users to transmit data to a cloudlet, which triggers the data sharing problem in the cloudlet. Firstly, we can utilize wearable devices to collect users' data, and in order to protect users privacy, we use NTRU mechanism tomake sure the transmission of users' data to cloudlet in security. Secondly, for the purpose of sharing data in the cloudlet, we use trust model to measure users' trust level to judge whether to share data or not. Thirdly, for privacy-preserving of remote cloud data, we partition the data stored in the remote cloud and encrypt the data in different ways, so as to not just ensure data protection but also accelerate the efficacy of transmission. Finally, we propose collaborative IDS based on cloudlet mesh to protect the whole system. The proposed schemes are validated with simulations and experiments.

REFERENCES

[1] https://www.patientslikeme.com/.

[2] C. Zhang, J. Sun, X. Zhu, and Y. Fang, "Privacy and security for online social networks: challeng

[3] L. Sweeney, "k-anonymity: A model for protecting privacy," International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, vol. 10, no. 05, pp. 557–570, 2002.

[4] L. M. Kaufman, "Data security in the world of cloud computing," Security & Privacy, IEEE, vol. 7, no. 4, pp. 61–64, 2009.

[4] R. Lu, X. Lin, and X. Shen, "Spoc: A secure and privacy-preserving opportunistic computing framework for mobile-healthcare emergency," Parallel and Distributed Systems, IEEE Transactions on, vol. 24, no. 3, pp. 614–624, 2013.