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

Min Chen, *Senior Member, IEEE*, Yongfeng Qian, Jing Chen,
Kai Hwang, *Fellow, IEEE*, Shiwen Mao, *Senior Member*, Long Hu

Abstract—With the popularity of wearable devices, along with the development of clouds and cloudlet technology, there has been increasing need to provide better medical care. The processing chain of medical data mainly includes data collection, data storage and data sharing, etc. Traditional healthcare system often requires the delivery of medical data to the cloud, which involves users' sensitive information and causes communication energy consumption. Practically, medical data sharing is a critical and challenging issue. Thus in this paper, we build up a novel healthcare system by utilizing the flexibility of cloudlet. 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, we present a new trust model to help users to select trustable partners who want to share stored data in the cloudlet. The trust model also helps similar patients to communicate with each other about their diseases. Thirdly, we divide users' medical data stored in remote cloud of hospital into three parts, and give them proper protection. Finally, in order to protect the healthcare system from malicious attacks, we develop a novel collaborative intrusion detection system (IDS) method based on cloudlet mesh, which can effectively prevent the remote healthcare big data cloud from attacks. Our experiments demonstrate the effectiveness of the proposed scheme.

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

1 Introduction

With the development of healthcare big data and wearable technology [1], as well as cloud computing and communication technologies [2], cloud-assisted healthcare big data computing becomes critical to meet users' evergrowing demands on health consultation [3]–[5]. However, it is challenging issue to personalize specific healthcare data for various users in a convenient fashion [6]. Previous work suggested the combination of social networks and healthcare service to facilitate [7] the trace of the disease treatment process for the retrieval of realtime disease information [8]. Healthcare social platform, such as PatientsLikeMe [9], can obtain information from other similar patients through data sharing in terms of user's own findings. Though sharing medical data on the social network is beneficial to both patients and doctors, the sensitive data might be leaked or stolen, which causes privacy and security problems [10] [11] without efficient protection for the shared data [12]. Therefore, how to balance privacy protection with the convenience of medical data sharing becomes a challenging issue.

With the advances in cloud computing, a large amount of data can be stored in various clouds [13], including cloudlets [14] and remote clouds [15], facilitating data sharing and intensive computations [16] [17]. However, cloud-based data sharing entails the following fundamental problems:

- How to protect the security of user's body data during its delivery to a cloudlet?
- How to make sure the data sharing in cloudlet will not cause privacy problem?
- As can be predicted, with the proliferation of electronic medical records (EMR) and cloud-assisted applications, more and more attentions should be paid to the security problems regarding to a remote cloud containing healthcare big data. How to secure the healthcare big data stored in a remote cloud?
- How to effectively protect the whole system from malicious attacks?

In terms of the above problems, this paper proposes a cloudlet based healthcare system. 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. During this stage, data privacy is the main concern. 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 based on cloudlet mesh to protect the cloud ecosystem.

M. Chen, Y. Qian, L. Hu are with School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China. J. Chen is with School of Computer, Wuhan University, Wuhan, China 430072. K. Hwang is with Electrical Engineering and Computer Science, the Univ. of Southern California, USA. S. Mao is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201 USA. E-mail: minchen@ieee.org, chenjing@whu.edu.cn, kaihwang@usc.edu, smao@ieee.org, longhu.cs@gmail.com.

TABLE 1
Feature table according to data style

Data Category	Data Type	Privacy Protect	Data Sharing
Physical Data	Physiological data	Medium	Medium
	Activity level	Low	Low
	Location	Low	Medium
	Environmental	Low	High
Cyber Data	Call logs	High	Low
	SMS logs	High	Low
	Application logs	High	Low
Social Network Data	SNS logs	low	High
Electronic Medical Data	Medical Data	High	Medium

In summary, the main contributions of this paper include:

- A cloudlet based healthcare system is presented, where the privacy of users' physiological data and the efficiency of data transmissions are our main concern. We use NTRU for data protection during data transmissions to the cloudlet.
- In order to share data in the cloudlet, we use users' similarity and reputation to build up trust model. Based on the measured users' trust level, the system determines whether data sharing is performed.
- We divide data in remote cloud into different kinds and utilize encryption mechanism to protect them respectively.
- We propose collaborative IDS based on cloudlet mesh to protect the whole healthcare system against malicious attacks.

The remainder of this article is organized as follows. In Section 2, we introduce the related work. For the healthcare data in the remote cloud and users' private health data, we propose a security system and introduce the framework of the entire system in Section 3. In Section 4, regarding protection of users' private data, we present the process for wearable medical device encryption; meanwhile, we discuss data sharing in the cloudlet, as well as protection and access of user EMR data in the cloud. Section 5 describes the collaborative IDS system based on the cloudlet mesh integrating several IDS's so that it can protect the remote cloud effectively. In Section 6, the performance metrics and evaluation of encryption algorithm are presented. The experimental results of collaborative IDS are given. Final conclusions are provided in Section 7.

2 Related Work

Our work is closely related to cloud-based privacy preserving and cloudlet mesh based collaborative IDS. We will give a brief review of the works in these aspects.

2.1 Cloud-based Privacy Preservation

Despite the development of the cloud technology and emergence of more and more cloud data sharing platforms, the clouds have not been widely utilized for healthcare data sharing due to privacy concerns [18]. There exist various works on conventional privacy protection of healthecare data [11], [19]–[25]. In Lu et al. [19], 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. The article [21] proposed a compound resolution which applies multiple combined technologies for the privacy protection of healthcare data sharing in the cloud environment. In Cao et al. [11], an MRSE (multi-keyword ranked search over encrypted data in cloud computing) privacy protection system was presented, which aims to provide users with a multi-keyword method for the cloud's encrypted data. Although this method can provide result ranking, in which people are interested, the amount of calculation could be cumbersome. In Zhang et al. [24], a priority based health data aggregation (PHDA) scheme was presented to protect and aggregate different types of healthcare date in cloud assisted wireless boby area network (WBANs). The article [25] investigates security and privacy issues in mobile healthcare networks,including the privacy-protection for healthcare data aggregation, the security for data processing and misbehavior. [26] describes a flexible security model especially for data centric applications in cloud computing based scenario to make sure data confidentiality, data integrity and fine grained access control to the application data. [27] give a systematic literature review of privacy-protection in cloud-assisted healthcare system.

2.2 Collaborative IDS based on cloudlet mesh

A number of prior works [28] have studied different intrusion detection systems with quite some advances. For example, [29] proposed a behavior-rule specification-based technique for intrusion detection. The main contribution is the performance outperforms other methods of anomaly-based techniques. [30] proposed a collaborative model for the cloud environment based on distributed IDS and IPS (intrusion prevention system). This model makes use of a hybrid detection technique to detect and take corresponding measures for any types of intrusion which harm the system, especially distributed intrusion. However, collaborative IDS based on the cloudlet mesh structure is a new kind of intrusion detection technique, which was first proposed in Shi et al. [31]. The authors demonstrated that the detection rate of the intrusion detection system established on the basis of a cloudlet mesh is relatively high. [32] describes design space, attacks that evade CIDSs and attacks on the availability of the CIDSs, and introduces comparison of specific CIDS approaches. [33] describes the IDS for privacy cloud. The authors give an overview of intrusion detection of cloud computing and provide a new idea for privacy cloud protection.

3 System Framework

The framework of the proposed cloudlet-based healthcare system is shown in Fig. 1. The client's physiological data are first

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2016.2617382, IEEE Transactions on Cloud Computing

collected by wearable devices such as smart clothing [34]. Then, those data are delivered to cloudlet. The following two important problems for healthcare data protection is considered. The first problem is healthcare data privacy protection and sharing data, as shown in Fig. 1(a). The second problem is to develop effective countermeasures to prevent the healthcare database from being intruded from outside, which is shown in Fig. 1(b).

We address the first problem on healthcare data encryption and sharing as follows.

- **Client data encryption.** We utilize the model presented in [23], and take the advantage of NTRU [35] 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, for example, patients suffer from similar kind of disease exchange information of treatment and share related data. For this purpose, we use users' similarity and reputation as input data. After we obtain users' trust levels, a certain threshold is set for the comparison. Once reaching or exceeding the threshold, it is considered that the trust between the users is enough for data sharing. Otherwise, the data will not shared with low trust level.
- **Remote cloud data privacy protection.** Compared to user's daily data in cloudlet, the data stored in remote contain larger scale medical data, e.g., EMR, which will be stored for a long term. We use the methods presented in [36] [21] to divide EMR into explicit identifier (EID), quasi-identifier (QID) and medical information (MI), which will be discussed in 4.3. After classifying, proper protection is given for the data containing users' sensitive information.
- **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.

4 Content Sharing and Privacy Protection

In this section, we address the problem of protection and data sharing. First, we introduce the encryption process for users' privacy data, which prevents the leakage or malicious use of users' private data during transmissions. Next, we present the identity management of users who want to access to the hospital's healthcare data. Thus, we can assign different users with different levels of permissions for data access, while avoiding data access beyond their permission levels. Finally, we give an application of using users' private data, which is beneficial to both users and doctors. Based on the healthcare big data stored in the remote cloud, a disease prediction model is built based on decision tree. The predictions will be reported to the users and doctors on demand.

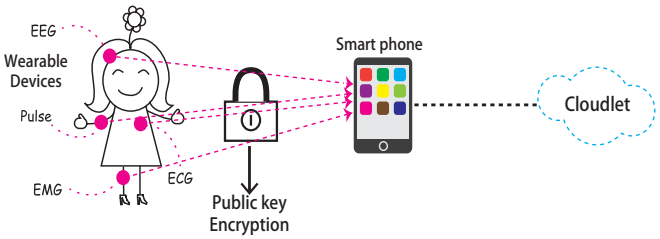


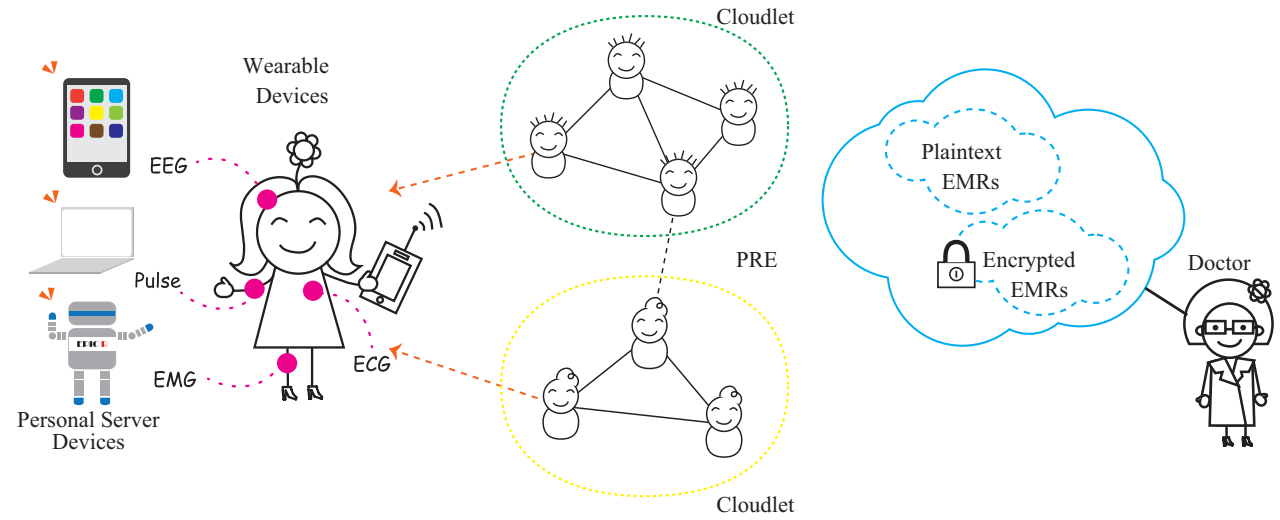
Fig. 2. Collection of encrypted data in the cloudlet.

4.1 Encryption at the User End

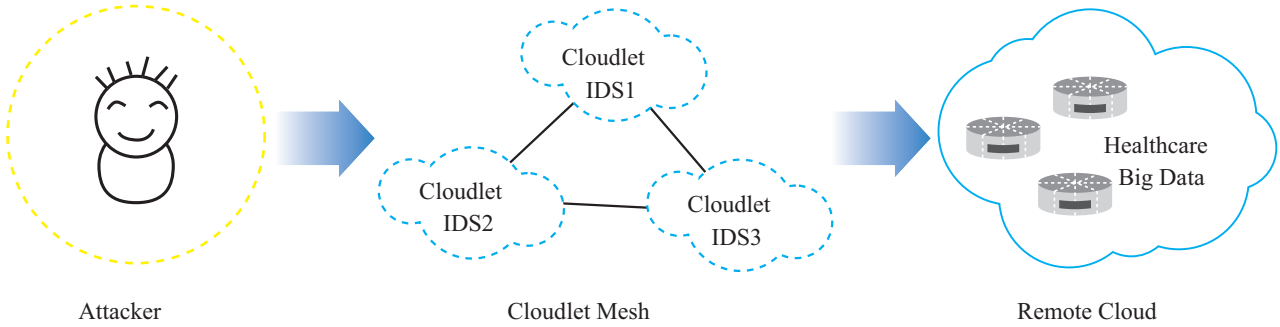
When using wearable devices to collect users' data, the procedure inevitably involves the user's sensitive information. Therefore, how to effectively collect and transmit users' data under efficient privacy protection is a critical problem [19]. In [24] a data collection method, called PHDA, is proposed based on data priority which can give proper cost and delay to different priorities data. In [37], 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 [38], Lu et al., study 3V data privacy protection issue based on big data of healthcare. Based on the model presented in [23], this paper utilizes the advantages of NTRU encryption scheme [35]. NTRU can protect the user's physiological data, such as heart rate, blood pressure and Electrocardiography (ECG), etc. Before transmitted to a smartphone, NTRU encryption scheme executed. The encrypted data will then be stored in the cloudlet through a cellular network or WiFi, as shown in Fig. 2.

Usually, the data collected by smart clothing are all unsigned integer vectors. For example, for heart rate data, the average heart beats detected each minute is denoted by hr and the plain data shall be $[hr, 0, \cdots, 0]$. We need to define clear space and cipher space for the encryption. As the definition of the polynomial ring is $R := \mathbb{Z}[x]/(x^n + 1)$, in the case of an arbitrary positive integer q, the definition of the quotient ring is known as $R_q = R/qR$. We define the clear space as R_p, so that the length is n and the integer vector is modulus p, which is always between 2 and 210. The cipher space is R_q, so the length is n and the integer vector is modulus p. In consideration of bandwidth, we generally make the R_q pass using the Chinese Remainder Transform (CRT) representation. For the sake of initial safety, we have $n = 1024$ and $q = 32$. We hereby describe the processes of encryption and deciphering in the following.

- KeyGen()$\rightarrow (pk, sk)$: 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 $pk = h = g \cdot f^{-1} \bmod q$.
- Enc$(pk = h, \mu \in R_p) \rightarrow c \in R_q$: let $r \in R, m \in R, m = \mu \bmod p$. Both m' and r follow the discrete Gaussian distribution, and we have $m = p \cdot m' + \mu, c = p \cdot r \cdot h + m \bmod q$.
- Dec$(sk = f, c \in R_q) \rightarrow \mu$: calculate $\bar{b} = f \cdot c \bmod q$, and make it an integer polynomial b, with factors within $[-q/2, q/2)$. Thus, we have $\mu = b \bmod p$.



(a) Illustrate of system framework.



(b) Collaborative IDS of remote cloud.

Fig. 1. Illustration of the system architecture: (a) Privacy protection; (b) Collaborative IDS.

The encrypted data will be transmitted to the smartphone with the homomorphic processing. We assume that the clear data of heart beat is $[hr, 0, ..., 0]$ and the array encryption is c_1. In the same way, if the blood pressure is bp, then the clear data is denoted as $[0, bp, 0, ..., 0]$ and the enciphered data shall be c_2. This way, we can get clear data and cipher data of all sensors. Since we use a public key encryption system and homomorphic encryption (HE), the smartphone can receive data $\{c_1, c_2, \ldots, c_n\}$ transmitted to $c_{agg} = c_1 + ... + c_n \bmod q$. Therefore, after we process the data with homomorphic encryption, the bandwidth is reduced effectively before the data are uploaded to the cloudlet, thus achieving energy and bandwidth savings.

4.2 Medical Data Sharing in the Cloudlet

The purpose of medical data sharing is to make better use of data between users. The paper [39] proposed data sharing strategy among several clouds, which used encryption method based on attribute to realize data sharing under semi-trusted cloud environment. However, it didn't consider users' social activities. In [40], Fabian et al., propose big data sharing method based on community cloud, but it didn't aim at medical data particularly. Based on the discussion above, we give the judgement during data sharing as follows.

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 [41] 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.

TABLE 2
Variable and Value for the Trust Model

Variable	Value level	Default scope
Reputation of user u_q	Bad	[0,0.2]
	Average	(0.2,0.6]
	Good	(0.6,1]
Similarity between u_p and u_q	Low	[0,0.2]
	Moderate	(0.2,0.6]
	High	(0.6,1]

- 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.

According to [36] [21], 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 [42]; (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 in [21], we part the EMR data table A into two independent tables, i.e., a ciphertext table T_e and a plaintext table T_p. 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 $\{D_1, D_2, \ldots, D_M\}$. For each disease D_i, there are corresponding characteristics $\{C_{i,1}, C_{i,2}, \ldots C_{i,i_n}\}$, $i = 1, \ldots, M$. In order to quantize disease characteristics, we define a question $Q_{i,j}$ for each characteristic $C_{i,j}, i = 1, \ldots, M, j = 1, \ldots, i_n$. 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 $\{Q_{i,1}, Q_{i,2}, \ldots, Q_{i,i_n}\}$ for each characteristic in $\{C_{i,1}, C_{i,2}, \ldots C_{i,i_n}\}$ of the corresponding diseases D_i, $i = 1, 2, \ldots, M$. For the sake of simplicity, we assume that the answer to each question is 0 or 1. Therefore, each disease D_i can acquire its testing results $\{e_{i,1}, e_{i,2}, \ldots, e_{i,i_n}\}$, $i = 1, \ldots, M$, with each $e_{i,j} = 0$ or $e_{i,j} = 1$.

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 $\{p_i, q_i, w_i\}$ satisfying the following conditions.

$$p_i + q_i = bw_i, \quad \frac{bw_i}{2} < q_i < bw_i, \quad a^2 buw_i < c, \qquad (1)$$

where u is integer.

After the parameters of a, p_i, q_i are obtained, encrypted data can be calculated. Then we have

$$v_i = ae_i + p_i, \quad v_i^{'} = s \cdot q_i \bmod c, \quad v_0^{'} = s \cdot q_0 \bmod c. \quad (2)$$

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.

5 Collaborative Intrusion Detection

In order to protect medical data, we also develop an intrusion detection system in this paper. Once a malicious attack is detected, the system will fire an alarm. This section presents a novel scheme to build a collaborative IDS system to deter intruders. In the following, we first consider what happens if the system is suffering from different attacks, while detection rates for individual IDS vary with the cloudlet servers. We will plot the detection rate and false alarm rate as the receiver operating characteristic (ROC) curves.

Next, we evaluate the collaborative detection rate and estimate the expected cost of implementation in the cloudlet mesh. We apply a decision tree to choose the optimal number of IDS's to be deployed on the mesh. The goal is to achieve a prescribed detection accuracy against the false alarm rate under the premise of minimizing the system cost.

5.1 Collaborative IDS

In this section, collaborative IDS is designed among m IDS, e.t., $S_1, S_2, \cdots, S_m$, 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 $\{S_1, S_2, \ldots, S_m\}$ 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 behavior

TABLE 3
Key Parameters Used in the CIDS

Variable	Explanation
I_i	Intrusion i
A_i	Alarm for IDS i
$1-\beta$	Detection Rate
α	False Alarm Rate
E_c	Expected Cost (Relative degree, no unit attached)
q_1	Probability of Collaborative IDS Reporting an Alarm
q_2	Probability of Collaborative IDS Reporting No-Alarm

in 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 $1-\beta$ to indicate the detection rate and α as the false alarm rate. If there exists K different types of intrusion, denoted as $I_1, I_2, \ldots, I_K$, then we have $I = I_1 \cup I_2 \cdots \cup I_K$. Assume that the probability of I_j is p_j, $j = 1, 2, \ldots, K$. Therefore, the probability of intrusion behavior in this system is $p(I) = \sum_{i=1}^{K} p_i$, while the probability of no intrusion behavior is $P(NI) = 1 - p(I)$. We thus have that $p(A|I) = 1 - \beta$ and $p(A|NI) = \alpha$.

As for each IDS, we use $p(NA_i|I_j) = \beta_{ij}$ to represent the probability of IDS S_i not triggering an alarm when having I_j, and $p(A_i|NI) = \alpha_i$ as the probability of S_i triggering an alarm when not being attacked. It follows that

$$\beta = p(NA|I) = p(NA_1|I) \cdots p(NA_m|I). \tag{3}$$

Since $I_i \cap I_j = \phi$, $i \neq j$, applying the total probability formula, we can obtain the probability that system S_1 does not trigger an alarm when there is an attack to intrude the system, as

$$\begin{aligned} p(NA_1|I) &= \frac{p(NA_1 \cap (I_1 \cup I_2 \cdots \cup I_K))}{P(I)} \\ &= \frac{\sum_{j=1}^{K} \beta_{1j} p_j}{\sum_{j=1}^{K} p_i}. \end{aligned} \tag{4}$$

For system $S_i, i = 2, 3, \ldots, m$, let $p(NA_i|I)$ denote the probability that no alarm is triggered by S_i. We have

$$p(NA_i|I) = \frac{\sum_{j=1}^{K} \beta_{ij} p_j}{\sum_{j=1}^{K} p_i}. \tag{5}$$

We can derive β as follows.

$$\beta = \prod_{i=1}^{m} \frac{\sum_{j=1}^{K} \beta_{ij} p_j}{\sum_{j=1}^{K} p_i}. \tag{6}$$

The false alarm rate $\alpha = p(A|NI) = 1 - p(NA|NI)$ can be obtained in a similarly manner, as

$$p(NA|NI) = \prod_{i=1}^{m} (1 - \alpha_i). \tag{7}$$

The false alarm rate α can be computed as follows.

$$\alpha = 1 - \prod_{i=1}^{m} (1 - \alpha_i). \tag{8}$$

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 C_α;
- when the system suffers from intrusion I_i, $1 \leq$i$\leq K$, 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 $\tilde{C}_i, 1 \leq i \leq K$;
- the cost in other scenarios is marked as 0.

Without loss of generality, we define the cost rate as $C_i = \tilde{C}_i / C_\alpha$. In the following, we adopt the decision tree to model the corresponding expected cost problem. Let $q_1, q_2 = p(NA)$ denote the probability of no alarm in a system. Based on the total probability formula, we have

$$q_1 = (1 - \beta) \sum_{t=1}^{K} p_i + \alpha (1 - \sum_{t=1}^{K} p_i). \tag{9}$$

$$q_2 = \beta \sum_{t=1}^{K} p_i + (1 - \alpha)(1 - \sum_{t=1}^{K} p_i). \tag{10}$$

Let $p_{1,i} = p(I_i|A)$, $i = 1, 2, \ldots, K$, denote the probability of intrusion occurrence under the condition that the system fires an alarm. Thus, $p_{1,i}$ can be calculated as follows.

$$p_{1,i} = \frac{(1 - \prod_{j=1}^{m} \beta_{ji}) p_i}{(1 - \beta) \sum_{t=1}^{K} p_i + \alpha (1 - \sum_{t=1}^{K} p_i)}, \quad i = 1, 2, \ldots, K. \tag{11}$$

Let $p_{1,K+1} = p(NI|A)$ denote the probability of no intrusion under the condition that the system fires an alarm, then:

$$p_{1,K+1} = 1 - \sum_{i=1}^{K} \frac{(1 - \prod_{j=1}^{m} \beta_{ji}) p_i}{(1 - \beta) \sum_{t=1}^{K} p_i + \alpha (1 - \sum_{t=1}^{K} p_i)}. \tag{12}$$

Let $p_{2,i} = p(I_i|NA)$, $i = 1, 2, \cdot, K$, denote the probability of intrusion occurrence when no alarm is given. It follows that

$$p_{2,i} = \frac{\prod_{j=1}^{m} \beta_{ji} p_i}{\beta \sum_{t=1}^{K} p_i + (1 - \alpha)(1 - \sum_{t=1}^{K} p_i)}, \quad i = 1, 2, \ldots, K. \tag{13}$$

Let $p_{2,K+1} = p(NI|NA)$ denote the probability of no intrusion occurrence when no alarm is given. We have

$$p_{2,K+1} = 1 - \sum_{i=1}^{K} \frac{\prod_{j=1}^{m} \beta_{ji} p_i}{\beta \sum_{t=1}^{K} p_i + (1 - \alpha)(1 - \sum_{t=1}^{K} p_i)}. \tag{14}$$

From the above analysis and the assumption on the cost rate, we can derive the expected cost as follows.

$$E_c = q_1 \cdot p_{1,K+1} + q_2 \cdot \sum_{j=1}^{K} p_{2,j} \cdot C_j. \tag{15}$$

Now let's consider how to choose the optimal IDS numbers and IDS combinations when constructing the collaborative IDS

system. Hereby we formulate an optimization problem based on the decision tree model. That is, under the circumstances of guaranteeing a certain detection rate $1-\tilde{\beta}$ 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.

$$\begin{aligned} \text{minimize} \quad & E_c && (16)\\ \text{subject to:} \quad & \alpha < \tilde{\alpha}, \quad \beta < \tilde{\beta} && (17)\\ & 0 \le p_{ij} \le 1, \quad i,j = 1,2,\ldots,K && (18)\\ & 0 \le q_i \le 1, \quad i = 1,2,\ldots,K && (19)\\ & C_j > 0, \quad j = 1,2,\ldots,K. && (20) \end{aligned}$$

The optimization problem is a integer programming problem which can be solved by a conventional solver,such as Matlab. Then, we can select a certain number of IDS system, in order to guarantee: (i) the detection rate ($1-\beta \ge 1-\tilde{\beta}$) is sufficiently large; (ii) the false alarm rate (α) is sufficiently small; and (iii) the expect cost of the entire system is minimized.

6 SIMULATION STUDY

In this chapter, firstly we utilize the delivery ratio to compare client data encryption method with remote cloud encryption mechanism. Then in terms of collaborative IDS based on cloudlet mesh, we describe ROC curve and relationship figure between IDS number and cost and detection rate.

6.1 Performance Discussion about data encryption

As discussed, we shall encrypt the data with the algorithm, which has been introduced previously, to protect private information after the data are collected by the users themselves. However, we also need to evaluate the performance of the proposed algorithm. We describe the changes of delivery ratio of client data encryption method with remote cloud encryption mechanism with the increasement of time. Fig. 3 shows the results. Through this figure, we can see two methods will both achieve a good delivery ratio with the increasement of time, while in general, the encryption method in remote cloud have better performance than encryption method at user end.

In Section 4.2, we have analyzed the timing of data sharing within cloudlet based on trust model. Here, the scope of user's reputation (denoted by r) is set to [0, 1]. As shown in Table 2, three levels (i.e., bad, average and good) are assigned to individual reputation. Specifically, reputations with ranges of [0,0.2], (0.2,0.6], (0.6,1] are marked as "bad", "average" and "good", respectively. Likewise, the similarity (denoted by s) between a pair of users is classified into three categories, i.e., low, moderate and right. For the sake of simplicity, the three categories corresponds to reputations with scopes of [0,0.2], (0.2,0.6] and (0.6,1] as well. Let $R1 \in [0, 0.2]$, $S1 \in [0, 0.2]$, $R2 \in (0.2, 0.6]$, and $S2 \in (0.6, 1]$. R1, R2, S1, S2 are random variables. As shown in Fig. 4, when users suffering from poor reputation while the similarity of users is low, the output of trust model is quite low, typically lower than 0.4. Practically, users would not like to share data under low trust level, since it's unsafe to share with a low reputation and similarity. Based on the observation of the two curves, compared to similarity, user's reputation generates larger impact on the output of trust level. In the other words, given a low reputation, even users are quite similar with each other, the system may make a decision of not sharing and not trust. With the increase of reputation and similarity, trust level of user will be improved. Users enjoy the data sharing with other partners with high reputation and similarity

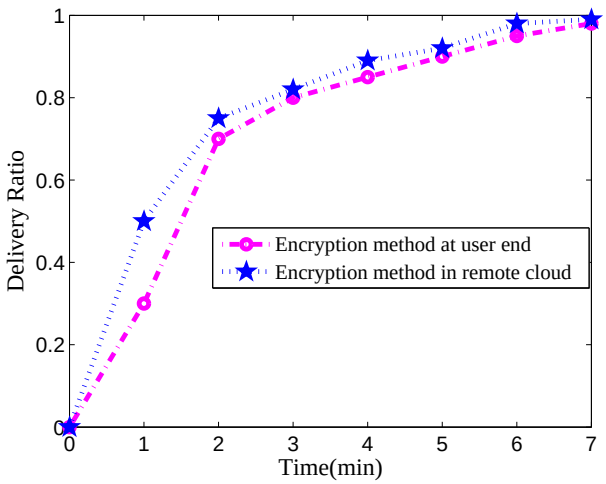


Fig. 3. Comparison of the delivery ratio of the encryption method in the remote cloud and user end.

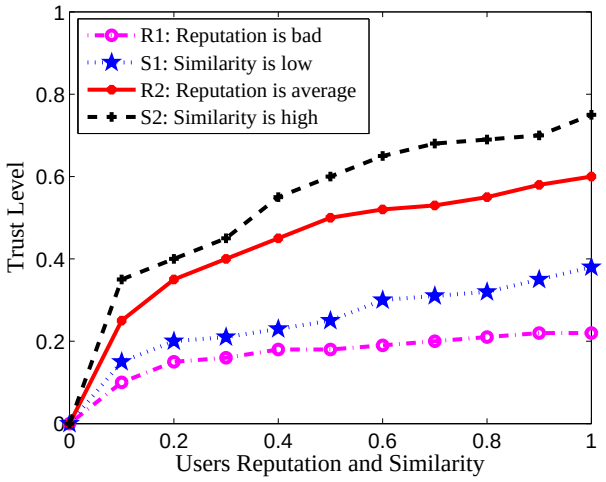


Fig. 4. Comparison of the trust level.

6.2 Collaborative IDS Performance Results

We use the cloudlet mesh simulator [43] to evaluate the effectiveness of the mesh security infrastructure. We develop a collaborative intrusion detection system (IDS) executed by multiple servers in the mesh. We use three independent IDS's and two intrusion types in our experiment. The probabilities of different types of intrusion are $p_1 = 0.001$ and $p_2 = 0.0015$.

Figure 5 plots the detection rate in the ROC curve of various IDS's used in the experiment against the false alarm rate. According to Fig. 5, the detection rate of every single IDS is below 30%. However, the collaborative IDS can achieve a detection rate of 60%, which is a considerable improvement over the single IDS approach.

If the IDS generates no alarm when there is actually an intrusion, the system would suffer heavy loss. Our proposed

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TCC.2016.2617382, IEEE Transactions on Cloud Computing

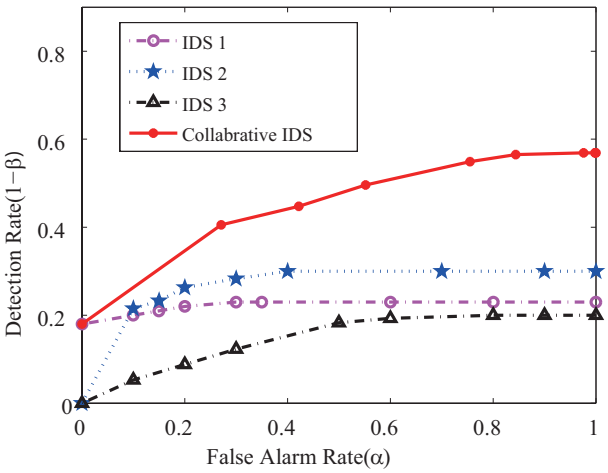


Fig. 5. Comparison of ROC curves for collaborative IDS's.

TABLE 4
Detection Rates of Various IDS Schemes

Cloudlet	α_i	$1-\beta_{i1}$	$1-\beta_{i2}$
Cloudlet 1	0.0090	0.35	0.36
Cloudlet 2	0.0080	0.34	0.35
Cloudlet 3	0.0060	0.32	0.34
Cloudlet 4	0.0050	0.30	0.32
Cloudlet 5	0.0040	0.28	0.30
Cloudlet 6	0.0020	0.26	0.28

collaborative IDS performs well from this regard. Nevertheless, we want to minimize the cost in addition to achieving a high detection rate. We consider two cost measures: $C_1 = 5$ and $C_2 = 6$. The unit of the cost is not shown here, because only relative costs are compared. There are six IDS's in this experiment, whose operational parameters are given in Table 4. We assume baseline values $\tilde{\alpha} < 0.035$ and $\tilde{\beta} < 0.3$.

If $m = j$, $1 \leq j \leq 6$, there are C_6^j choices, and the cost value m is chosen as the smallest cost among those C_6^j costs. We guarantee the detection rate to be above 70% and the false alarm rate to be below 3.5%. At the same time, we search for the lowest cost configuration for the collaborative IDS system. The theoretical derivation in Section 5.1 leads to the optimal solution. Because the detection rate of a single IDS is below 35%, our collaborative system has doubled the detection rate at a minimum cost. It can be seen from Fig. 6 that four IDSs should be chosen to work collectively and cooperatively to yield the optimal performance.

7 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 to

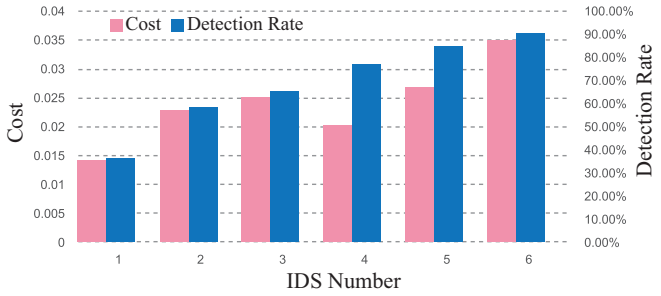


Fig. 6. Cost and detection rate of the entire IDS system. The optimal configuration is shown to use 4 IDS's with a 75% detection rate under a minimum system cost of 0.02. Only relative costs are shown here.

make 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.

8 Acknowledgments

This work is supported by the National Natural Science Foundation of China under Grant No. 61272451 and 61572380.

References

[1] K. Hung, Y. Zhang, and B. Tai, "Wearable medical devices for tele-home healthcare," in *Engineering in Medicine and Biology Society, 2004. IEMBS'04. 26th Annual International Conference of the IEEE*, vol. 2. IEEE, 2004, pp. 5384–5387.

[2] M. S. Hossain, "Cloud-supported cyber–physical localization framework for patients monitoring," 2015.

[3] J. Zhao, L. Wang, J. Tao, J. Chen, W. Sun, R. Ranjan, J. Kołodziej, A. Streit, and D. Georgakopoulos, "A security framework in g-hadoop for big data computing across distributed cloud data centres," *Journal of Computer and System Sciences*, vol. 80, no. 5, pp. 994–1007, 2014.

[4] M. S. Hossain and G. Muhammad, "Cloud-assisted industrial internet of things (iiot)–enabled framework for health monitoring," *Computer Networks*, vol. 101, pp. 192–202, 2016.

[5] R. Zhang and L. Liu, "Security models and requirements for healthcare application clouds," in *Cloud Computing (CLOUD), 2010 IEEE 3rd International Conference on*. IEEE, 2010, pp. 268–275.

[6] K. He, J. Chen, R. Du, Q. Wu, G. Xue, and X. Zhang, "Deypos: Deduplicatable dynamic proof of storage for multi-user environments," 2016.

[7] L. Griffin and E. De Leastar, "Social networking healthcare," in *Wearable Micro and Nano Technologies for Personalized Health (pHealth), 2009 6th International Workshop on*. IEEE, 2009, pp. 75–78.

[8] W. Xiang, G. Wang, M. Pickering, and Y. Zhang, "Big video data for light-field-based 3d telemedicine," *IEEE Network*, vol. 30, no. 3, pp. 30–38, 2016.

[9] "https://www.patientslikeme.com/."

[10] C. Zhang, J. Sun, X. Zhu, and Y. Fang, "Privacy and security for online social networks: challenges and opportunities," *Network, IEEE*, vol. 24, no. 4, pp. 13–18, 2010.

[11] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," *Parallel and Distributed Systems, IEEE Transactions on*, vol. 25, no. 1, pp. 222–233, 2014.

[12] K. T. Pickard and M. Swan, "Big desire to share big health data: A shift in consumer attitudes toward personal health information," in *2014 AAAI Spring Symposium Series*, 2014.

[13] T. Xu, W. Xiang, Q. Guo, and L. Mo, "Mining cloud 3d video data for interactive video services," *Mobile Networks and Applications*, vol. 20, no. 3, pp. 320–327, 2015.

[14] M. Quwaider and Y. Jararweh, "Cloudlet-based efficient data collection in wireless body area networks," *Simulation Modelling Practice and Theory*, vol. 50, pp. 57–71, 2015.

[15] K. Dongre, R. S. Thakur, A. Abraham *et al.*, "Secure cloud storage of data," in *Computer Communication and Informatics (ICCCI), 2014 International Conference on*. IEEE, 2014, pp. 1–5.

[16] M. S. Hossain, G. Muhammad, M. F. Alhamid, B. Song, and K. Al-Mutib, "Audio-visual emotion recognition using big data towards 5g," *Mobile Networks and Applications*, pp. 1–11, 2016.

[17] J. Chen, K. He, R. Du, M. Zheng, Y. Xiang, and Q. Yuan, "Dominating set and network coding-based routing in wireless mesh networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 26, no. 2, pp. 423–433, 2015.

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

[19] 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.

[20] J.-J. Yang, J. Li, J. Mulder, Y. Wang, S. Chen, H. Wu, Q. Wang, and H. Pan, "Emerging information technologies for enhanced healthcare," *Computers in Industry*, vol. 69, pp. 3–11, 2015.

[21] J.-J. Yang, J.-Q. Li, and Y. Niu, "A hybrid solution for privacy preserving medical data sharing in the cloud environment," *Future Generation Computer Systems*, vol. 43, pp. 74–86, 2015.

[22] A. Andersen, K. Y. Yigzaw, and R. Karlsen, "Privacy preserving health data processing," in *e-Health Networking, Applications and Services (Healthcom), 2014 IEEE 16th International Conference on*. IEEE, 2014, pp. 225–230.

[23] K. Rohloff and D. B. Cousins, "A scalable implementation of fully homomorphic encryption built on ntru," in *Financial Cryptography and Data Security*. Springer, 2014, pp. 221–234.

[24] K. Zhang, X. Liang, M. Baura, R. Lu, and X. S. Shen, "Phda: A priority based health data aggregation with privacy preservation for cloud assisted wbans," *Information Sciences*, vol. 284, pp. 130–141, 2014.

[25] K. Zhang, K. Yang, X. Liang, Z. Su, X. Shen, and H. H. Luo, "Security and privacy for mobile healthcare networks: from a quality of protection perspective," *Wireless Communications, IEEE*, vol. 22, no. 4, pp. 104–112, 2015.

[26] S. Saha, R. Das, S. Datta, and S. Neogy, "A cloud security framework for a data centric wsn application," in *Proceedings of the 17th International Conference on Distributed Computing and Networking*. ACM, 2016, p. 39.

[27] A. Sajid and H. Abbas, "Data privacy in cloud-assisted healthcare systems: State of the art and future challenges," *Journal of Medical Systems*, vol. 40, no. 6, pp. 1–16, 2016.

[28] S. M. Sajjad, S. H. Bouk, and M. Yousaf, "Neighbor node trust based intrusion detection system for wsn," *Procedia Computer Science*, vol. 63, pp. 183–188, 2015.

[29] R. Mitchell and I.-R. Chen, "Behavior rule specification-based intrusion detection for safety critical medical cyber physical systems," *Dependable and Secure Computing, IEEE Transactions on*, vol. 12, no. 1, pp. 16–30, 2015.

[30] H. Mohamed, L. Adil, T. Saida, and M. Hicham, "A collaborative intrusion detection and prevention system in cloud computing," in *AFRICON, 2013*. IEEE, 2013, pp. 1–5.

[31] Y. Shi, S. Abhilash, and K. Hwang, "Cloudlet mesh for securing mobile clouds from intrusions and network attacks," in *The Third IEEE International Conference on Mobile Cloud Computing, Services, and Engineering,(Mobile Cloud 2015)*. IEEE, 2015.

[32] E. Vasilomanolakis, S. Karuppayah, M. Mühlhäuser, and M. Fischer, "Taxonomy and survey of collaborative intrusion detection," *ACM Computing Surveys (CSUR)*, vol. 47, no. 4, p. 55, 2015.

[33] P. K. Rajendran, B. Muthukumar, and G. Nagarajan, "Hybrid intrusion detection system for private cloud: a systematic approach," *Procedia Computer Science*, vol. 48, pp. 325–329, 2015.

[34] M. Chen, Y. Ma, J. Song, C.-F. Lai, and B. Hu, "Smart clothing: Connecting human with clouds and big data for sustainable health monitoring," *ACM/Springer Mobile Networks and Applications*.

[35] D. Nuñez, I. Agudo, and J. Lopez, "Ntrureencrypt: An efficient proxy re-encryption scheme based on ntru," in *Proceedings of the 10th ACM Symposium on Information, Computer and Communications Security*. ACM, 2015, pp. 179–189.

[36] J. Li, J.-J. Yang, Y. Zhao, and B. Liu, "A top-down approach for approximate data anonymisation," *Enterprise Information Systems*, vol. 7, no. 3, pp. 272–302, 2013.

[37] Q. Li, G. Cao, and T. La Porta, "Efficient and privacy-aware data aggregation in mobile sensing," *Dependable and Secure Computing, IEEE Transactions on*, vol. 11, no. 2, pp. 115–129, 2014.

[38] R. Lu, H. Zhu, X. Liu, J. K. Liu, and J. Shao, "Toward efficient and privacy-preserving computing in big data era," *Network, IEEE*, vol. 28, no. 4, pp. 46–50, 2014.

[39] B. Fabian, T. Ermakova, and P. Junghanns, "Collaborative and secure sharing of healthcare data in multi-clouds," *Information Systems*, vol. 48, pp. 132–150, 2015.

[40] Y. Wu, M. Su, W. Zheng, K. Hwang, and A. Y. Zomaya, "Associative big data sharing in community clouds: The meepo approach," *IEEE Cloud Computing*, vol. 2, no. 6, pp. 64–73, 2015.

[41] J. Sun, F. Wang, J. Hu, and S. Edabollahi, "Supervised patient similarity measure of heterogeneous patient records," *ACM SIGKDD Explorations Newsletter*, vol. 14, no. 1, pp. 16–24, 2012.

[42] 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.

[43] K. A. Khan, Q. Wang, C. Luo, X. Wang, and C. Grecos, "Comparative study of internet cloud and cloudlet over wireless mesh networks for real-time applications," in *SPIE Photonics Europe*. International Society for Optics and Photonics, 2014, pp. 91 390K–91 390K.