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

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

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. Second, 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. Third, 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], [4], [5]. However, it is challenging issue to personalize specific healthcare data for various users in a convenient fashion. Previous work suggested the combination of social networks and healthcare service to facilitate [6] the trace of the disease treatment process for the retrieval of realtime disease information. Healthcare social platform, such as PatientsLikeMe [7], 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 [8], [9] without efficient protection for the shared data [10]. Therefore, how to balance privacy protection with the convenience of medical data sharing becomes a challenging issue.

- *M. Chen, Y. Qian, and L. Hu are with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China. E-mail: minchen@ieee.org, yongfengqian.epic@qq.com, longhu.cs@gmail.com.*
- *J. Chen is with the School of Computer, Wuhan University, Wuhan 430072, China. E-mail: chenjing@whu.edu.cn.*
- *K. Hwang is with Electrical Engineering and Computer Science, University of Southern California, Los Angeles, CA 90089-0097 USA. E-mail: kaihwang@usc.edu.*
- *S. Mao is with the Department of Electrical and Computer Engineering, Auburn University, Auburn, AL 36849-5201 USA. E-mail: smao@ieee.org.*

With the advances in cloud computing, a large amount of data can be stored in various clouds, including cloudlets [11] and remote clouds [12], facilitating data sharing and intensive computations [13]. 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

TABLE 1
Feature Table According to Data Style

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

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.

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 [14]. There exist various works on conventional privacy protection of healthecare data [9], [15], [16], [17], [18], [19], [20], [21]. In Lu et al. [15], 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 [17] 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. [9], 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. [20], 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 [21] investigates security and privacy issues in mobile healthcare networks,including the privacy-protection for healthcare data aggregation, the security for data processing and misbehavior. Ref. [22] 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. [23] 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 [24] have studied different intrusion detection systems with quite some advances. For example, [25] proposed a behavior-rule specification-based technique for intrusion detection. The main contribution is the performance outperforms other methods of anomaly-based techniques. [26] 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. [27]. The authors demonstrated that the detection rate of the intrusion detection system established on the basis of a cloudlet mesh is relatively high. Ref. [28] describes design space, attacks that evade CIDSs and attacks on the availability of the CIDSs, and introduces comparison of specific CIDS approaches. [29] describes the IDS
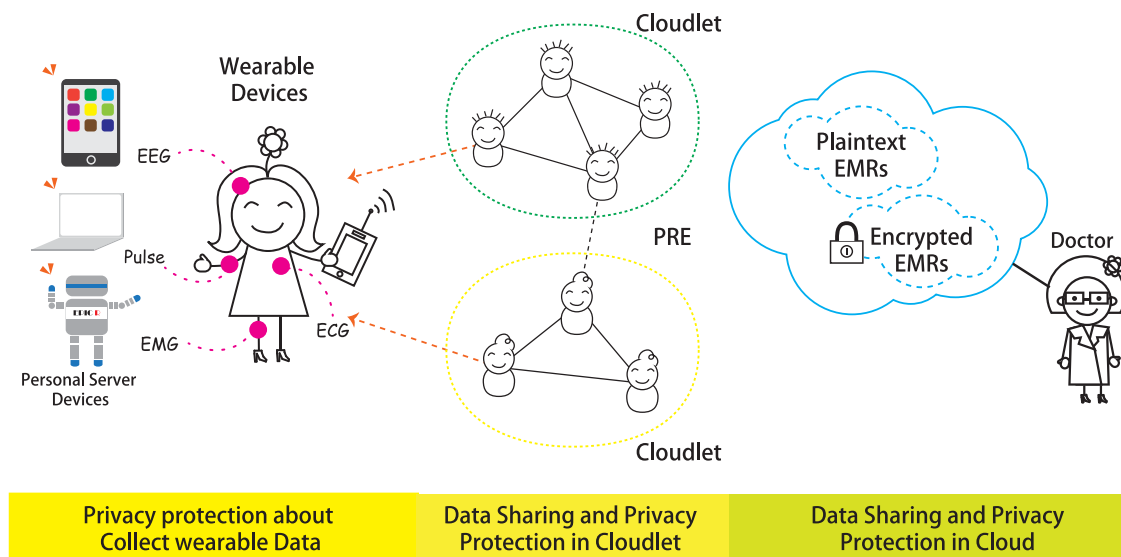
(a) Illustrate of system framework.
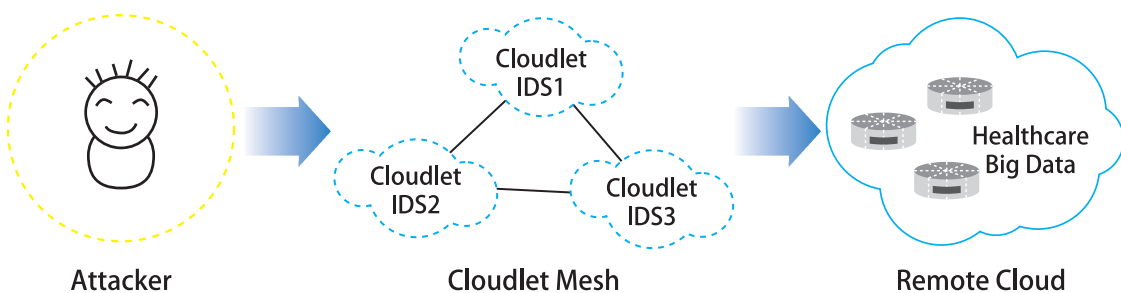
(b) Collaborative IDS of remote cloud.

Fig. 1. Illustration of the system architecture: (a) Privacy protection; (b) Collaborative 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 collected by wearable devices such as smart clothing [30]. 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. 1a. The second problem is to develop effective counter-measures to prevent the healthcare database from being intruded from outside, which is shown in Fig. 1b.

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

- *Client Data Encryption.* We utilize the model presented in [19], and take the advantage of NTRU [31] 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 [17], [32] 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
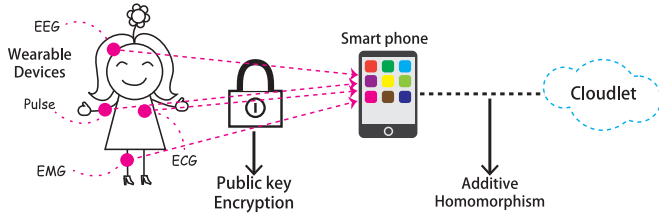
Fig. 2. Collection of encrypted data in the cloudlet.

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.

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 [15]. In [20] a data collection method, called PHDA, is proposed based on data priority which can give proper cost and delay to different priorities data. In [33], 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 [34], Lu et al., study 3V data privacy protection issue based on big data of healthcare. Based on the model presented in [19], this paper utilizes the advantages of NTRU encryption scheme [31]. 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 $[hr, 0, \ldots, 0]$. 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 $n = 1,024$ and $q = 32$. We hereby describe the processes of encryption and deciphering in the following.

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

The encrypted data will be transmitted to the smartphone with the homomorphic processing. We assume that the clear data of heart beat is $[hr, 0, \ldots, 0]$ and the array encryption is c_1. In the same way, if the blood pressure is bp, then the clear data is denoted as $[0, bp, 0, \ldots, 0]$ 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 $c_{agg} = c_1 + \cdots + c_n \mod q$. 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 [35] 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 [36], 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 [37] 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.

According to [17], [32], 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 [38]; (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 [17], 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

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

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., $S_1, S_2, \ldots, S_m$, 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 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 $I = I_1 \cup I_2 \cup I_K$. 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

$$\beta = p(NA|I) = p(NA_1|I) \cdots p(NA_m|I). \qquad (3)$$

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

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

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

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

We can derive β as follows.

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

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

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

The false alarm rate α can be computed as follows

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

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

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

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

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:

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

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

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

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

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

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

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

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

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

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:
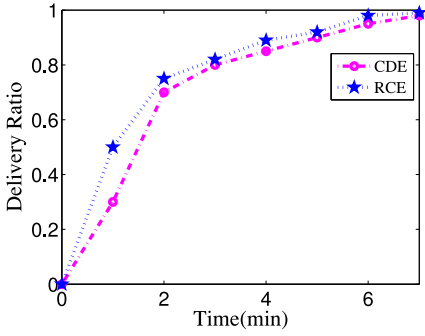
Fig. 3. Comparison of the delivery ratio of CDE and RCE.

$$\text{minimize} \quad E_c \quad\quad\quad (16)$$

$$\text{subject to:} \quad \alpha < \tilde{\alpha}, \quad \beta < \tilde{\beta} \quad\quad (17)$$

$$0 \le p_{ij} \le 1, \quad i,j = 1,2,\ldots,K \quad (18)$$

$$0 \le q_i \le 1, \quad i = 1,2,\ldots,K \quad\quad (19)$$

$$C_j > 0, \quad j = 1,2,\ldots,K. \quad\quad (20)$$

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, first 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. For the sake of simplicity, we use CDE to denote client data encryption and RCE to denote remote cloud encrytion. Fig. 3 shows the results. Through this figure, we can see

TABLE 2
Variable and Value for the Trust Model

Variable	Value
Reputation of user u_q	Bad
	Average
	Good
Similarity between u_p and u_q	Low
	Moderate
	High

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

CDE and RCE will both achieve a good delivery ratio with the increasement of time, while in general, RCE have better performance than CDE.

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

6.2   Collaborative IDS Performance Results

We use the cloudlet mesh simulator [39] 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 $p_2 = 0.0015$.
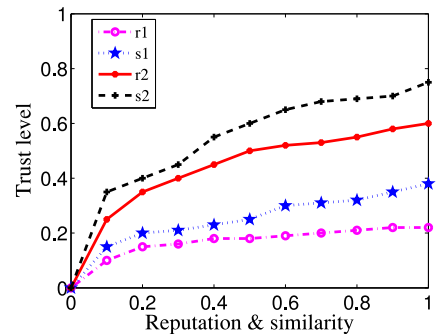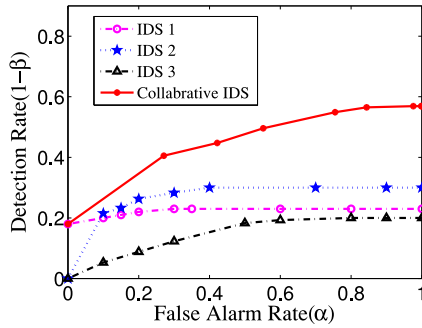

Fig. 4. Comparison of the trust level.

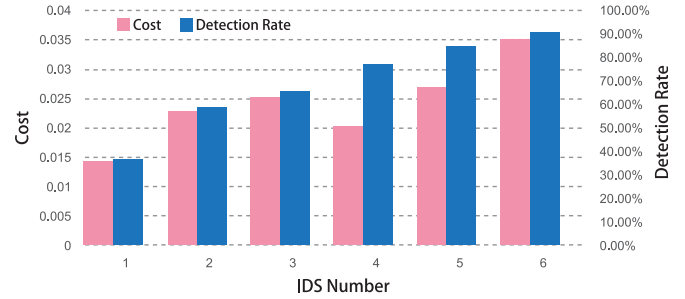Fig. 5. Comparison of ROC curves for collaborative IDS's.

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

Fig. 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 percent. However, the collaborative IDS can achieve a detection rate of 60 percent, 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 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 percent and the false alarm rate to be below 3.5 percent. 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 percent, 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.

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

First, we can utilize wearable devices to collect users' data, and in order to protect users privacy, we use NTRU mechanism to make sure the transmission of users' data to cloudlet in security. Second, 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. Third, 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] K. Hung, Y. Zhang, and B. Tai, "Wearable medical devices for tele-home healthcare," in *Proc. 26th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, 2004, vol. 2, pp. 5384–5387.

[2] M. S. Hossain, "Cloud-supported cyber-physical localization framework for patients monitoring," *IEEE Syst. J.*, vol. pp, no. 99, pp. 1–10, 2015.

[3] J. Zhao, et al., "A security framework in G-hadoop for big data computing across distributed cloud data centres," *J. Comput. Syst. Sci.*, 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," *Comput. Netw.*, vol. 101, pp. 192–202, 2016.

[5] R. Zhang and L. Liu, "Security models and requirements for healthcare application clouds," in *Proc. IEEE 3rd Int. Conf. Cloud Comput.*, 2010, pp. 268–275.

[6] L. Griffin and E. De Leastar, "Social networking healthcare," in *Proc. 6th Int. Workshop Wearable Micro Nano Technol. Personalized Health (pHealth)*, 2009, pp. 75–78.

[7] [Online]. Avaialble: https://www.patientslikeme.com/

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

[9] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 1, pp. 222–233, Jan. 2014.

[10] K. Hwang and M. Chen, "Big data analytics for cloud, IoT and congnitive learning," John Wiley, 2017 (in press).

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

[12] K. Dongre, R. S. Thakur, and A. Abraham, "Secure cloud storage of data," in *Proc. Int. Conf. Comput. Commun. Informat.*, 2014, pp. 1–5.

[13] M. S. Hossain, G. Muhammad, M. F. Alhamid, B. Song, and K. Al-Mutib, "Audio-visual emotion recognition using big data towards 5g," *Mobile Netw. Appl.*, vol. 21, pp. 753–763, 2016.

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

[15] R. Lu, X. Lin, and X. Shen, "Spoc: A secure and privacy-preserving opportunistic computing framework for mobile-healthcare emergency," *IEEE Trans. Parallel Distrib. Syst.*, vol. 24, no. 3, pp. 614–624, Mar. 2013.

[16] J.-J. Yang, et al., "Emerging information technologies for enhanced healthcare," *Comput. Industry*, vol. 69, pp. 3–11, 2015.

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

[18] A. Andersen, K. Y. Yigzaw, and R. Karlsen, "Privacy preserving health data processing," in *Proc. IEEE 16th Int. Conf. e-Health Netw. Appl. Services*, 2014, pp. 225–230.

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

[20] 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," *Inf. Sci.*, vol. 284, pp. 130–141, 2014.

[21] 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," *IEEE Wireless Commun.*, vol. 22, no. 4, pp. 104–112, May 2015.

[22] S. Saha, R. Das, S. Datta, and S. Neogy, "A cloud security framework for a data centric WSN application," in *Proc. 17th Int. Conf. Distrib. Comput. Netw.*, 2016, Art. no. 39.

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

[24] S. M. Sajjad, S. H. Bouk, and M. Yousaf, "Neighbor node trust based intrusion detection system for WSN," *Procedia Comput. Sci.*, vol. 63, pp. 183–188, 2015.

[25] R. Mitchell and I.-R. Chen, "Behavior rule specification-based intrusion detection for safety critical medical cyber physical systems," *IEEE Trans. Dependable Secure Comput.*, vol. 12, no. 1, pp. 16–30, Jan./Feb. 2015.

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

[27] Y. Shi, S. Abhilash, and K. Hwang, "Cloudlet mesh for securing mobile clouds from intrusions and network attacks," in *Proc. 3rd IEEE Int. Conf. Mobile Cloud Comput. Services Eng.*, 2015, pp. 109–118.

[28] E. Vasilomanolakis, S. Karuppayah, M. Mühlhäuser, and M. Fischer, "Taxonomy and survey of collaborative intrusion detection," *ACM Comput. Surveys*, vol. 47, no. 4, 2015, Art. no. 55.

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

[30] 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 Netw. Appl.*, vol. 21, pp. 825–845, 2016.

[31] D. Nuñez, I. Agudo, and J. Lopez, "NTRUReEncrypt: An efficient proxy re-encryption scheme based on NTRU," in *Proc. 10th ACM Symp. Inf. Comput. Commun. Secur.*, 2015, pp. 179–189.

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

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

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

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

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

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

[38] L. Sweeney, "k-anonymity: A model for protecting privacy," *Int. J. Uncertainty Fuzziness Knowl.-Based Syst.*, vol. 10, no. 5, pp. 557–570, 2002.

[39] K. Hwang, "Cloud and cognitive computing: an machine learning approach," MIT Press, 2017, (in press).

Min Chen is a professor in the School of Computer Science and Technology, Huazhong University of Science and Technology (HUST). He is a chair of the IEEE Computer Society (CS) and Special Technical Communities (STC) on Big Data. He was an assistant professor in the School of Computer Science and Engineering, Seoul National University (SNU) from Sep. 2009 to Feb. 2012. He worked as a post-doctoral fellow in the Department of Electrical and Computer Engineering, University of British Columbia (UBC) for three years. Before joining UBC, he was a post-doctoral fellow with SNU for one and half years. He received Best Paper Award from QShine 2008, IEEE ICC 2012, ICST IndustrialIoT 2016, and IEEE IWCMC 2016. He serves as an editor or an associate editor of Information Sciences, Information Fusion, and the *IEEE Access*, etc. He is a guest editor of the *IEEE Network*, the *IEEE Wireless Communications*, and the *IEEE Trans. Service Computing*, etc. He is co-chair of the IEEE ICC 2012-Communications Theory Symposium, and co-chair of the IEEE ICC 2013-Wireless Networks Symposium. He is general co-chair of the IEEE CIT-2012, Tridentcom 2014, Mobimedia 2015, and Tridentcom 2017. He is Keynote speaker for CyberC 2012, Mobiquitous 2012, Cloudcomp 2015, IndustrialIoT 2016, and 7th Brainstorming Workshop on 5G Wireless. He has more than 280 paper publications, including 150+ SCI papers, 60+ IEEE Trans./Journal papers, 8 ISI highly cited papers and 1 hot paper. He has published two books: *OPNET IoT Simulation* (2015) and *Big Data Inspiration* (2015) with HUST Press, and a book on big data: *Big Data Related Technologies* (2014) with Springer Series in computer science. His Google scholars Citations reached 7,250+ with an h-index of 42. His top paper was cited 815+ times. He is a senior member of the IEEE since 2009. His research focuses on cyber physical systems, IoT sensing, 5G networks, mobile cloud computing, SDN, healthcare big data, medica cloud privacy and security, body area networks, emotion communications and robotics, etc.

Yongfeng Qian received the MS degree from Huazhong University of Science and Technology, China, in 2015. She is currently working toward the PhD degree in Embedded and Pervasive Computing lab led by Prof. Min Chen in the School of Computer Science and Technology, Huazhong University of Science and Technology. Her research includes Internet of Things, big data analytics, deep learning, mobile cloud computing, network security, and cryptography.

Jing Chen received the PhD degree in computer science from Huazhong University of Science and Technology, Wuhan. He worked as an associate professor from 2010. His research interests in computer science include the areas of network security, cloud security. He is the chief investigator of several projects in network and system security, funded by the National Natural Science Foundation of China (NSFC). He has published more than 60 research papers in many international journals and conferences, such as the *IEEE Transactions on Parallel and Distributed System*, the *International Journal of Parallel and Distributed System*, the *IEEE INFOCOM*, SECON, the *Trust, Security and Privacy in Computing and Communications*, the NSS Journal. He acts as a reviewer for many Journals and conferences, such as the *IEEE Transactions on Wireless Communication*, the *IEEE Transactions on Industrial Informatics*, the *Computer Communications*, and the *IEEE GLOBCOM*.

Kai Hwang received the PhD degree from the University of California, Berkeley, in 1972. He is a professor of electrical engineering and computer science, with the University of Southern California (USC). Prior to joining USC in 1986, he has taught with Purdue University for 11 years. He has served as the founding editor-in-chief of the *Journal of Parallel and Distributed Computing* from 1983 to 2011. He has published 8 books and 250 scientific papers. According to Google scholars, his work was cited more than 15,000 times with an h-index of 52. His most cited book on *Computer Architecture and Parallel Processing* was cited more than 2,300 times and his PowerTrust (*IEEE-TPDS*, April 2007) paper was cited more than 540 times. He received the Lifetime Achievement Award from *IEEE Cloudcom*-2012 for his pioneering contributions in the field of computer architecture, parallel, distributed and cloud computing, and cyber security. He is a life fellow of the IEEE.

Shiwen Mao (S'99–M'04–SM'09) received the PhD degree in electrical and computer engineering from the Polytechnic University (now New York University Polytechnic School of Engineering), Brooklyn, New York, in 2004. He is the Samuel Ginn Endowed professor and director of the Wireless Engineering Research and Education Center (WEREC), Auburn University, Auburn, Alabama. His research interests include wireless networks and multimedia communications. He is on the editorial board of the *IEEE Transactions on Multimedia*, the *IEEE Internet of Things* journal, the *IEEE Communications Surveys and Tutorials*, the *IEEE Multimedia*, among others. He serves as steering committee member for the IEEE ICME and Adhoc-Nets, area TPC chair of the IEEE INFOCOM 2016, technical program vice chair of the Information Systems of IEEE INFOCOM 2015, and symposium/track co-chair for many conferences, including the IEEE ICC, the IEEE GLOBECOM, and the ICCCN. He is a distinguished lecturer of the the *IEEE Vehicular Technology Society*, and the vice chair—Letters and member communications of the *IEEE Communications Society Multimedia Communications Technical Committee*. He received the 2013 IEEE ComSoc MMTC Outstanding Leadership Award and the NSF CAREER Award in 2010, and also a co-recipient of the IEEE WCNC 2015 Best Paper Award, the IEEE ICC 2013 Best Paper Award, and the 2004 IEEE Communications Society Leonard G. Abraham Prize in the Field of Communications Systems. He is a senior member of the IEEE.

Long Hu received the BS and master's degrees from Huazhong University of Science and Technology (HUST) He is working toward the PhD degree in the School of Computer Science and Technology, HUST. He is the publication chair of 4th International Conference on Cloud Computing (CloudComp 2013). Currently, his research includes 5G mobile communication system, big data mining, marine-ship communication, Internet of Things, and multimedia transmission over wireless network, etc. He is a member of IEEE.

▷ **For more information on this or any other computing topic, please visit our Digital Library at** www.computer.org/csdl.