PMRSS: Privacy-Preserving Medical Record Searching Scheme for Intelligent Diagnosis in IoT Healthcare

Yi Sun, *Member, IEEE*, Jie Liu, Keping Yu, *Member, IEEE*, Mamoun Alazab, *Senior Member, IEEE*, and Kaixiang Lin

***Abstract*—In medical field, previous patients' cases are extremely private as well as intensely valuable to current disease diagnosis. Therefore, how to make full use of precious cases while not leaking out patients' privacy is a leading and promising work especially in future privacy-preserving intelligent medical period. In this article, we investigate how to securely invoke patients' records from past case-database while protecting the privacy of both current diagnosed patient and the case-database and construct a privacy-preserving medical record searching scheme based on ElGamal Blind Signature. In our scheme, by blinded the healthy data of the patient and the database of the iDoctor, respectively, the patient can securely make self-helped medical diagnosis by invoking past case-database and securely comparing the blinded abstracts of current data and previous records. Moreover, the patient can obtain target searching information intelligently at the same time he knows whether the abstracts match or not instead of obtaining it after matching. It greatly increases the timeliness of information acquisition and meets high-speed information sharing requirements, especially in 5G era. What's more, our proposed scheme achieves bilateral security, that is, whether the abstracts match or not, both of the privacy of the case-database and the private information of the current patient are well protected. Besides, it resists different levels of violent ergodic attacks by adjusting the number of zeros in a bit string according to different security requirements.**

***Index Terms*—ElGamal blind signature, intelligent medical diagnosis, privacy-preserving medical record searching system.**

Manuscript received December 29, 2020; revised March 3, 2021; accepted March 29, 2021. Date of publication April 2, 2021; date of current version December 6, 2021. This work was supported in part by the National Key R&D Program of China under Grant 2020YFB1006002 and in part by Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) under Grant JP18K18044 and Grant JP21K17736. Paper no. TII-20-5750. *(Corresponding author: Keping Yu.)*

Yi Sun, Jie Liu, and Kaixiang Lin are with the School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications, Beijing 100876, China, and also with the National Engineering Laboratory for Mobile Network Security (e-mail: sybupt@bupt.edu.cn; ljbupt@bupt.edu.cn; xiang_0112@bupt.edu.cn).

Keping Yu is with Global Information and Telecommunication Institute, Waseda University, Tokyo 169-8050, Japan (e-mail: yukeping@fuji.waseda.jp).

Mamoun Alazab is with the with the College of Engineering, IT, and Environment, Charles Darwin University, Casuarina, NT 0811, Australia (e-mail: mamoun.alazab@cdu.edu.au).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TII.2021.3070544.

Digital Object Identifier 10.1109/TII.2021.3070544

I. Introduction

WITH the popularization of sensor technology and Internet of Things healthcare, multiple home medical equipment such as infrared thermometer, blood pressure monitor, and heart rate monitor are already quite common in people's daily life and have been used to measure essential body parameters such as heart rate, body temperature, etc. Therefore, in IoT healthcare scenario, patients can make self-helped medical diagnosis by uploading physical healthy data obtained from IoT medical devices to iDoctor, which is a kind of self-helped service medical system, to obtain professional healthcare advice [1]. Furthermore, these kind of self-helped service medical devices are becoming more portable, accurate, and individuate with the rapid development of information technology [2]–[4]. In this case, intelligent medical diagnosis is an irresistible and promising trend in future medical area. Like E-commerce, it will be extremely convenient for patients to get personalized and professionalized diagnostic report anytime and anywhere, especially with the popular of IoT healthcare.

However, the high requirement in security of patients' data hinders iDoctor from providing credible and accurate medical services. In recent years, medical information leakage occurred at any time owing to the security vulnerability of healthcare information system, since malware is more and more difficult to detect and resist [5]–[7]. As a result, related healthcare data are in jeopardy [8],[9]. For example, Anthem company, the second largest U.S. health insurance provider, was once attacked by hackers and lead to 78 million pieces of customer information disclosure including patients' individual information, healthy data, and some other sensitive data. Therefore, how to protect the security of both current diagnosed patients' information and the data base of the iDoctor besides obtaining credible and accurate medical result intelligently is the most difficult problem in the application and development of intelligent medical diagnosis, which makes securely searching related diagnosis report from the case-database of the iDoctor be a promising trend in future intelligent medical diagnosis.

TABLE I
COMPARISON IN TERMS OF DESIRABLE PROPERTIES

SCHEME	DESIRABLE PROPERTY							
	DP1	**DP2**	**DP3**	**DP4**	**DP5**	**DP6**	**DP7**	**DP8**
Medisn-2010 [10]	-	-	-	-	-	-	-	-
Healthedge-2017 [11]	-	-	-	-	-	-	-	-
PHY-2017 [12]	√	-	-	-	√	-	-	-
SSAC-2019 [13]	-	-	-	-	√	√	-	-
LSAS-2019 [14]	√	-	-	√	√	-	-	-
AGE-2019 [15]	√	-	-	√	√	√	-	-
SMAP-2020 [16]	√	-	-	√	√	-	-	-
ECC-2020 [17]	√	-	-	√	√	-	-	-
SHS-2020 [18]	√	-	√	√	√	√	√	-
OUR PMRSS	√	√	√	√	√	√	√	√

A. Related Works on Privacy-Preserving Healthcare

Many previous research have introduced some cryptography tools [19], [20] to solve this problem such as full homomorphic encryption (FHE) [21], [22]. Due to the shortcomings of FHE in cost, these solutions inevitably bring immense computing cost and are not applicable to the large-scale application scenarios. Besides, some researchers also have proposed other solutions based on RSA [23]. However, RSA signature scheme is feasible depending on a higher precondition. For example, the RSA signature scheme is secure on condition that the RSA known target inversion problem (RSA-KTI) is hard. Moreover, the privacy-preserving searching technology based on differential privacy [24] and secret sharing [25] divides the privacy data into several pieces called "share," controls the number of acquired shares to achieve the required accuracy according to the actual privacy requirements; and maps the privacy data to n copies through threshold policy, and randomly selects K copies from n subdata sources when the user requests data retrieval. It brings many redundant interactions and not applicable to instant communication, which requires both of efficiency and security. Searchable encryption [26], as a new cryptographic primitive, enables users to search keywords in ciphertext domain. In these schemes, data are always stored on the cloud server in ciphertext and the powerful cloud server is used to retrieve keywords without disclosing any user's privacy. By this way, it indeed protects the user's privacy effectively as well as improving the retrieval efficiency greatly with the help of the cloud server [27]. But it is difficult to avoid the cloud service provider from participating in partial decipherment and the cloud will be fatal once being corrupted. In brief, these solutions, which introduce a third party or other participants, are not suitable for secure one-to-one information searching scene in our real life. Herein, the solution that can search target information while protecting the privacy of the owners depends on no third party is our aim. In this respect, even though many new solutions have been put forward continuously with the development of new technology in semiconductor technology, sensors, and microcontrollers in IoT healthcare. Only a few of them have considered security. For convenience, we have made a comparison in Table I , where DP is the abbreviation of desirable property; DP1: Privacy of current patient; DP2: Privacy of healthcare system; DP3: Identity privacy; DP4: Localization privacy; DP5: Resistance to replay and forgery attacks; DP6: Data security; DP7: Physical security of the nodes; DP8: Bilateral security.

TABLE II
NOTATIONS

Notation	Description
PLH	patient's local host
iDD	iDoctor's disease database
M^*	the vector of parameter items
M_w	the w-th parameter item
M_{iw}	the i-th patient's parameter value of the w-th item
M_i^*	query vector of the i-th patient
t	capacity of iDD
d_i	the i-th disease
m_i	trait vector of d_i
c_i	disease diagnostic report
l	the number of zeros in the bit string 0^l

The work in [10] and [11] were not able to provide any concise solution to meet the desirable properties. In [12], [13] can meet only two desirable properties. Some important desirable security features such as the privacy of current patient, the privacy of healthcare system, identity privacy, localization privacy, physical security of the nodes, and the bilateral security are not satisfied. Among the recent works, most of them [14]–[18] considered the desirable security features in the privacy of current patient, localization privacy, and physical security of the nodes. Moreover, some of them [18] can ensure the desirable security features in the privacy of healthcare system, identity privacy, data security, and physical security of the nodes. However, all of these solutions cannot provide bilateral security to protect the security of both current diagnosed patients' information and the database of the iDoctor. With that in mind, our proposed PMRSS aims to achieve the eight desirable properties, including bilateral security.

B. Problem Statement and Motivation

In term of information searching, we first describe two typical scenarios of information searching as shown in Fig. 1, and then summarize corresponding requirements and our motivation in designing a secure information searching scheme in IoT Healthcare.

1) Direct Information Searching: In real life, information requirements can be satisfied by directly searching. It is desirable that the target can be discovered over the information pool by searching directly related information such as the target information itself or the keyword. For instance, a reader who wants to find a book named "Jane Eyre" can directly search it from a library with the book name "Jane Eyre." In daily life, we also directly search Alice's phone number over phone directory with the name "Alice."

2) Indirect Information Searching: If the target is not determined, information requirements can be satisfied by indirectly searching. It is desirable that the target can be discovered over the information pool by searching indirectly related information such as the descriptive information or requirements. For instance, a reader who wants to find a medium-length novel book can search it from a library with the description "medium-length novel." We can also find out the people contacted two days before by date information over the call record.

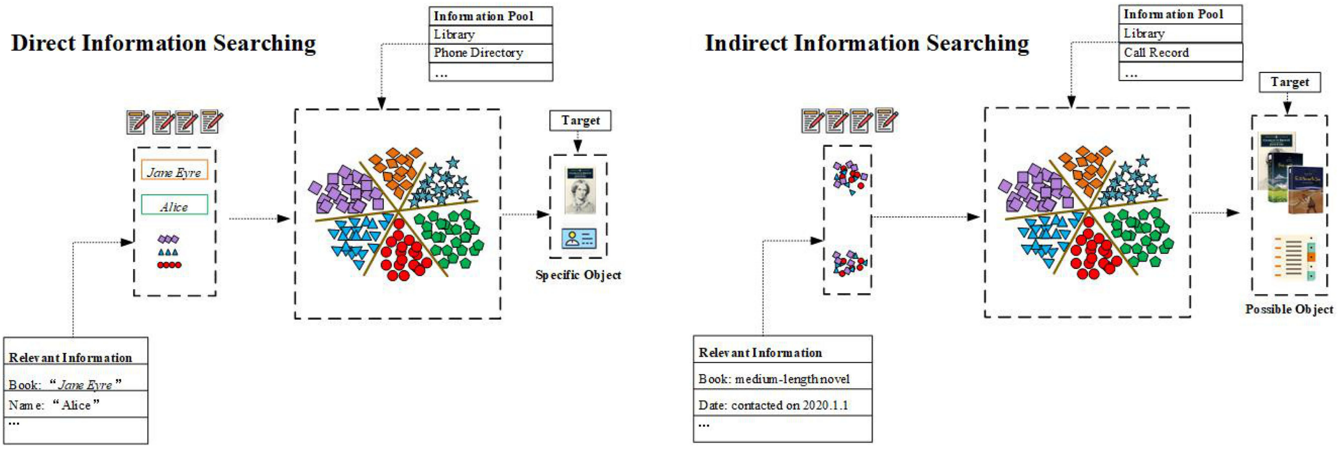


Fig. 1. Typical scenarios of information searching.

Nowadays, people deeply realize the importance of data security. The consciousness that no data is not private is agreeing by more and more people. No matter the directly related data such as book name, "Jane Eyre," "Alice," or the indirectly related information, "medium-length novel" or seemly irrelated information, "2020.1.1," become the key data in target information searching. People thus do not want to leak out any kind of information in target searching and no one wants to easily share his information, even the nonconfidential data. People no longer blindly enjoy the convenience of the Internet and become cautious and conservative. Especially in IoT era, how to securely share data [28], [29] and obtain target information is the main bottleneck in intelligent medical service. Therefore, in order to securely invoke and share past medical records to make diagnosis, privacy-preserving medical record searching scheme (PMRSS) in IoT healthcare, where the input used for searching must be protected as well as the result, is a challenging and an urgent issue in intelligent medical service. Herein, we propose two basic requirements in PMRSS in terms of secure information searching.

1) Correctness: The scheme should extract the target from the information pool, which is large volume and consisted of multiple kinds of data. The result should be the target information from the information pool the user searched instead of any irrelated information from the pool or related and irrelated information from elsewhere. In intelligent medical service, that is, the patient should extract the diagnosis report from the case-database of the iDoctor, which is derived from previous cases of the database instead of any irrelated report of the database or related and irrelated disease from elsewhere.

2) Bilateral Security: The searcher should not get any other information of the information pool except of the target information he searched and the private information of the searcher should not be leaked out. In intelligent medical service, that is, no matter whether the searching task is successful or not, the data privacy of both current patient and the iDoctor will not be disclosed. If the patient successfully searches the diagnosis report, the patient should not be able to obtain any other information of the iDoctor except the diagnosis report. If the patient cannot search any associated diagnosis report, then both parties can not obtain any information of the other one.

Thus, a practical intelligent medical diagnosis scheme, which can protect the security of both current diagnosed patients' information and the data base of the iDoctor while obtaining credible and accurate medical result intelligently is the most desirable and difficult problem in the application and development of intelligent medical diagnosis.

C. Main Contribution

In this article, we focus on this problem and design a secure and practical PMRSS based on ELGamal blind signature. Compared with previous solutions, our solution has the following four advantages.

1) PMRSS realizes intelligent self-helped medical diagnosis by IoMT data privately. There is no need for the participation of real doctors or any centers.
2) PMRSS has low latency. The patient obtains the diagnosis report at the same time he knows whether the encrypted abstracts match or not instead of obtaining after matching. Compared with previous information searching solutions, the proposed scheme does not need the two extra steps, Feedback and Resend, to get the target information after matching, which increases the timeliness of information acquisition and meets high-speed information sharing requirements, especially in future D2D communication of 5G.
3) Bilateral security can be achieved. We can protect the security of both current diagnosed patients' information and the database of the iDoctor no matter whether the abstracts match or not.
4) The number of zeros in a bit string denoted as the parameter l can be adjusted to resist different levels of violent ergodic attacks according to different security requirements.

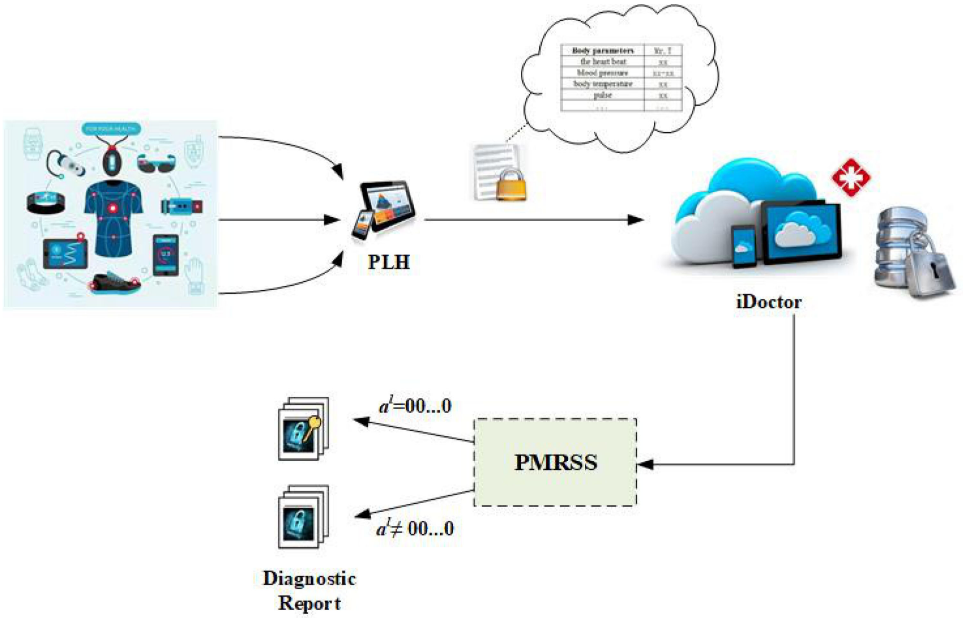


Fig. 2. Framework of PMRSS.

II. PMRSS

In PMRSS, the nations used in this paper are illustrated in Table II. There are two participants: the patient's local host (PLH) and the intelligent doctor (iDoctor) who owns a disease case-database called iDoctor-database (iDD). In this scheme, iDoctor is a data provider, who possesses iDD containing a set of confirmed cases solutions and corresponding body parameters, and the patient is a client who needs to invoke useful confirmed case solutions associated with his body parameters from the iDoctor. For protecting the security of the patient and the iDoctor, the proposed system needs to achieve bilateral security.

A. Proposed Model

In intelligent diagnosis of IoT healthcare, each patient collects his own body parameters like blood pressure, body temperature, pulse and so forth, and then these parameters can be encrypted and uploaded to iDoctor by PLH. The iDD stores sundry parameters about their health date, which are collected by some home medical equipment. In this system, patients can make private self-diagnosis by uploading essential body parameters to iDoctor. It introduces ELGamal blind signature algorithm to encrypt the information between patient and iDoctor so that iDoctor gets no idea about the patient's body parameters and the patient has no information about iDD. In short, the proposed scheme can not only prevent iDD from disclosing, but also protect the patient's health date, personal privacy, and their queries efficiently. The framework of the whole process is illustrated in Fig. 2.

Each patient can measure essential body parameters such as heart rate, body temperature, etc. by multiple home medical equipment by themselves by IoMT. Afterward, patients can store these health data in the local host. Herein, we call the blood pressure, pulse, body temperature, etc, as parameter items and denote these health data as $M^* = \{M_1, M_2, \ldots, M_n\}$, while the value corresponding to the items, blood pressure, pulse, body temperature, and so forth, are the parameter values. For ease of understanding, we can assume that PLH deals with these parameters in the form of vector and we also call them as query vector. For example

$$M_i^* = \{M_{i1}, M_{i2}, \ldots, M_{in}\}.$$

M_i^* is the query vector of the ith patient, where all $\{M_{iw}\}_{w=1,\ldots,n}$ are n necessary parameters uploading to iDoctor, and M_{iw} is the ith patient's parameter value of the wth parameter item.

In the initial stage, PLH only needs to collect the health data, encrypts them, and then uploads the cipher text to iDoctor. In the final stage, PHL receives an answer encrypted by ELGamal blind signature algorithm and then obtains a credible and accurate diagnostic report corresponding to the patient's physical conditions without decrypting the information from iDD.

iDD is a disease database as follows:

$$\text{iDD} = \{d_1, d_2, \ldots, d_t\}$$
$$d_i = (i, m_i, c_i), i = 1, \ldots, t$$

where t is the dimension of iDD; i is the index of a disease; m_i, called as trait vector of the disease d_i, is a vector that contains all standard parameters corresponding to M^*; and c_i is the diagnostic report.

Concerning the above parameters, we have some supplementary explanations as follows.

1) M^*: it includes all common parameter items and some professional parameter items the iDoctor needs for diagnosis. For most diseases, it is necessary to measure blood pressure, pulse, body temperature, and so forth in the early diagnosis. We call these indexes as the common parameter items. However, each category of diseases may show abnormalities in some

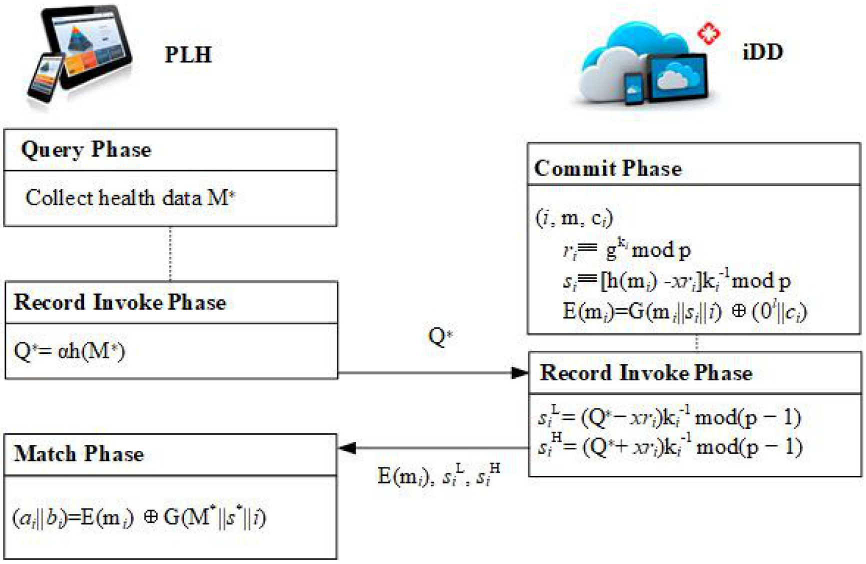

PLH
iDD
Query Phase
Collect health data M*
Commit Phase
(i, m, c_i)
$r_i \equiv g^{k_i} \bmod p$
$s_i \equiv [h(m_i) - xr_i]k_i^{-1} \bmod p$
$E(m_i) = G(m_i \| s_i \| i) \oplus (0^l \| c_i)$
Record Invoke Phase
$Q^* = \alpha h(M^*)$
Q^*
Record Invoke Phase
$s_i^L = (Q^* - xr_i)k_i^{-1} \bmod (p-1)$
$s_i^H = (Q^* + xr_i)k_i^{-1} \bmod (p-1)$
Match Phase
$E(m_i),\ s_i^L,\ s_i^H$
$(a_i \| b_i) = E(m_i) \oplus G(M^* \| s^* \| i)$

Fig. 3. Workflow of PMRSS.

professional parameter items. For example, respiratory diseases need to observe the value of vital capacity, total lung capacity, and eosinophilic granulocyte. If the patient feels upset stomach, his professional parameter items should include pepsin secretion factor, gastrin, and so on. After selecting uncomfortable body parts and entering the physical symptoms, iDoctor can tell the patient the needed parameter items in current diagnosis.

2) iDD $= \{d_1, d_2, \ldots, d_t\}$: it concludes all diseases the iDoctor can diagnose. The iDoctor collects various diseases' diagnostic parameters and solutions from existing advanced hospital as much as possible and updates the database from time to time. Thence, the disease database is the core asset to the iDoctor and its security is the central to the whole scheme.

3) m_i: it includes all common parameter items and various diseases' professional parameter items the iDoctor needs for diagnosis. Different kinds of diseases have different trait vectors, moreover, the trait vectors of the same disease at different time can be different, since the iDD collects and refines some additional professional parameter items. The dimension t and $\{m_i\}_{i=1,\ldots,t}$ of the trait vector are determined by iDoctor's medical technology level. So keeping improving iDD means that patients can get a more accurate and professional diagnostic report from the iDoctor.

4) c_i: it includes the description of the disease, doctors' advices, prescriptions corresponding to the ith disease d_i and recommended hospitals and doctors in reality. The report may conclude many doctors' advices and prescriptions. If the patient is not satisfied with the diagnostic report, he can consult the doctor who is recommended by iDoctor face to face.

B. Proposed Scheme

For visualizing the privacy-preserving medical record searching process, we illustrate the workflow in Fig. 3.

In general, our scheme consists of four phases called *Query Phase*, *Commit Phase*, *Record Invoke Phase* and *Match Phase*.

In query phase, the patient sends uncomfortable body symptoms to iDoctor and then iDoctor tells the patient the parameter items needed in this diagnosis. In commit phase, the iDoctor encrypts all trait vectors and stores them in the fixed form. The record invoke phase is actually composed of PLH's initial stage and iDD's deliver phase. After that, the match phase is the final stage of PLH in practice, and in this phase PLH uses private keyword matching algorithm to obtain a credible and accurate diagnostic report corresponding to the patient's physical conditions without decrypting the cipher text from iDD.

- *Query Phase*

  Step 1: The patient selects uncomfortable body parts and enters the physical symptoms with brief words in the PLH. Then PLH sends them to iDoctor.

  Step 2: According to the information from PLH, iDoctor distinguishes what kind of disease is queried by the patient from all kinds of diseases. Then, it tells the patient what parameter items are needed in this diagnosis, especially the professional parameter items.

  Step 3: The patient uses multiple home medical equipment to optionally measure the value of required parameter items and builds the query vector denoted as follows:

$$M^* = \{M_1, M_2, \ldots, M_n\}.$$

- *Commit Phase*

  Step 1: iDoctor extracts all diseases' parameter items $m_i, (i = 1, \ldots, t)$ and diagnostic report $c_i (i = 1, \ldots, t)$ from the disease database, where t is the number of diseases. These diseases are the specific diseases about the category of disease queried by the patient.

  Step 2: Build a temporary disease database

$$\text{iDD}^* = \{d_1, d_2, \ldots, d_t\}$$

Algorithm 1

```
Input: $iDD^*, p, g \in Z_p^*, x \in_R Z_{p-1}^*, k_i, i = 1, \ldots, t,$
for $i = 1$ to $t$, do
    $r_i = g^{k_i} mod p$
    $s_i = [h(m_i) - xr_i]k_i^{-1} mod(p-1)$
    $E(m_i) = G(m_i \| s_i \| i) \oplus (0^l \| c_i)$
end for
Output: $E(m_i)$
```

where $d_i = (i, m_i, c_i), i = 1, \ldots, t$. Herein, iDD* $\subset$ iDD is diagnostic basis.

Step 3: Select a large prime number p, which guarantees that the discrete logarithm problem is hard in Z_p, a generator $g \in Z_p^*$ and a random number $x \in_R Z_{p-1}^*$. The random numbers p, g, and x are the private information of iDoctor and they remain unchanged throughout the whole process.

Step 4: Let G be a pseudorandom generator. For $i = 1, \ldots, t$, iDoctor computes

$$r_i = g^{k_i} \text{mod} p$$

$$s_i = [h(m_i) - xr_i]k_i^{-1} \text{mod}(p-1)$$

$$E(m_i) = G(m_i \| s_i \| i) \oplus (0^l \| c_i)$$

where $\|$ denotes concatenation, $k_i, k_1 \neq k_2 \neq \cdots \neq k_t, i = 1, \ldots, t,$ is a new random number selected by iDoctor in every cycle calculation and the value of k_i cannot be reused. The pseudocode is as the following Algorithm 1.

- *Record Invoke Phase*

  Step 1: PLH chooses a random element α and then computes Q^* and sends Q^* to iDcotor.

  $$Q^* = \alpha h(M^*)$$

  where $h(.)$ is a Hash function.

  For $j = 1, \ldots, t$, iDcotor computes

  $$s_i^L = (Q^* - xr_i)k_i^{-1} \text{mod}(p-1)$$

  $$s_i^H = (Q^* + xr_i)k_i^{-1} \text{mod}(p-1).$$

  Then, iDcotor sends $E(m_i), s_i^L, s_i^H, i = 1, \ldots, t$, to PLH. The pseudocode is as the following algorithm 2.

  Step 3: After PLH receives $E(m_i), s_i^L, s_i^H, i = 1, \ldots, t$, it computes

  $$q_i = \frac{s_i^L + s_i^H}{2a}$$

  $$p_i = \frac{s_i^L - s_i^H}{2}$$

  $$s^* = q_i + p_i.$$

  The pseudocode is as the following algorithm 3.

Algorithm 2

```
Input: $Q^*, p, g \in Z_p^*, x \in_R Z_{p-1}^*, k_i, i = 1, \ldots, t,$
for $i = 1$ to $t$, do
    $r_i = g^{k_i} mod p$
    $s_i^L = (Q^* - xr_i)k_i^{-1} mod(p-1)$
    $s_i^H = (Q^* + xr_i)k_i^{-1} mod(p-1)$
end for
Output: $E(m_i)$, $s_i^L$, $s_i^H$, $i = 1, \ldots, t.$
```

Algorithm 3

```
Input: $\alpha$, $E(m_i), s_i^L, s_i^H, i = 1, \ldots, t,$
for $i = 1$ to $t$, do
    $q_i = \frac{s_i^L + s_i^H}{2a}$
    $p_i = \frac{s_i^L - s_i^H}{2}$
    $s^* = q_i + p_i$
end for
Output: $s^*$
```

- *Match Phase*

  For $i = 1, \ldots, t$, PLH computes

  $$(a_i \| b_i) = E(m_i) \oplus G(M^* \| s^* \| i)$$

  where the zero-label l, the number of zeros in the bit string 0^l, is the security parameter, and the length of l can be adjusted according to the security requirements to resist different levels of violent ergodic attacks.

  If $a_i = 0^l$, $b_i = c_i$. PLH outputs b_i; else PLH outputs nothing.

  From the match phase, it can find that the patient and the iDctor encrypt their own private data, respectively, by blind signature, and then PLH judges whether or not the trait encrypted vectors match with each other. If the trait encrypted vectors match successfully, PLH can obtain the diagnosis report immediately. Moreover, even if the trait encrypted vectors match successfully, PLH can only obtain the diagnosis report, which matches the current input, but not any other more information of iDD. Otherwise, if the the trait encrypted vectors cannot match, neither one can get any information of the other one. What is more, compared with previous solutions, the proposed PMRSS does not need the two extra steps, feedback and resend, to get the target information after matching, which increases the timeliness of result acquisition and can meet high-speed information sharing requirements in future 5G era. Besides, the length of the zero-label l can be adjusted to resist different levels of violent ergodic attacks according to different security requirements.

III. Security Analysis and Evaluation

A. Security Proof

In order to illustrate that the proposed system satisfies desirable security properties, we give the following security proof.

1) Correctness: From the scheme, whether the adversary exists or not, PLH succeeds to obtain the target result c_i if and only if

$$a_i = 0^l.$$

Herein, we define $\beta[1-l,\phi]$ as the front l bit of $\phi, \phi \in_R Z^*_{p-1}$.

Therefore, $a_i = 0^l$ means that the first l bits of $G(m_i \| s_i \| i) \oplus (0^l \| c_i) \oplus G(M^* \| s^* \| i)$ are 0, that is,

$$\beta[1-l, G(m_i\|s_i\|i) \oplus (0^l\|c_i) \oplus G(M^*\|s^*\|i)] = 0.$$

From the formula, it implies that

$$\beta[1-l, G(m_i\|s_i\|i) \oplus G(M^*\|s^*\|i)] = 0$$

which means the trait vectors m_i of c_i and M^* successfully match. Therefore, it is obvious to find that if $a_i = 0^l$, PLH can extract the correct diagnosis report related to the trait M^* from iDD, which he searched instead of any irrelated information.

2) Bilateral Security: In order to show that the proposed PMRSS achieves bilateral secure, it should show that the data and identity confidentiality of both PLH and iDoctor should be protected. Herein, we avow two basic assumptions.

1) Larger integer factorization problem is hard.

2) Computing the discrete logarithm on the finite field is hard.

Then, we analyze the whole process from PLH's view and iDoctor's view, respectively. Concretely, we consider a forger to make queries to iDoctor for some feedback and use the feedback data to recover encryption algorithm. The proposed scheme can overcome this difficulty perfectly by introducing the encryption scheme based on ELGamal digital signature. When a patient makes a query to iDoctor, our scheme immediately uses ELGamal blind signature algorithm to encrypt the patient's parameters and generates a matching trait vector at the same time.

Since all information from PLH to iDoctor are blinded, the adversary only can grab cipher text from insecure transmission channel and iDoctor cannot get any private information of PLH. The messages from iDoctor to PLH are all encrypted by ELGamal blind signature and thus the forger cannot collect any useful information about the iDD by sending queries to iDoctor.

In terms of PLH, the main security threat is that the adversary may steal cipher text from insecure transmission channel and iDoctor. If PLH's private data are transferred in an insecure transmission channel, the adversary can collect the cipher text of PLH to decipher them. Because the cipher text Q^* of PLH is the product of α and $h(M^*)$. The parameter α and the properties of hash function $h(\cdot)$ affect the security of the cipher text. In the scheme, for each inquiry, PLH chooses a random element α by using a random number generator and each one is independent. Therefore, if an adversary wants to decrypt Q^*, the large integer factorization problem is forced to be solved. Besides, for any specify hash value z, seeking out the message x to meet $h(x) = z$ is computationally infeasible. In summary, even the adversary has collected the cipher text of PLH, it is difficult to decrypt it.

Another threat to PLH is from iDoctor. If PLH directly sends the original data or hash coding parameters to iDoctor, iDoctor will obtain patient's parameters by using hash value to compare every record in database one by one. In the proposed scheme, we use the processed Hash value with a random number α. Besides, in order to prevent iDoctor from obtaining the requirement of the patient, PLH performs the match phase in our scheme. Therefore, iDoctor only provides a temporary encryption database iDD* to PLH and stays out of matching operation.

In terms of iDoctor, it applies ElGamal blind signature. Hence, the security of iDoctor is based on the security of ElGamal blind signature. Moreover, computing the discrete logarithm on the finite field is hard and ElGamal blind signature is secure [30]. It is obvious to conclude that the security of iDoctor can be achieved.

Concretely, we select the same k to compute the signature of two different message,(r, s_1) is the signature of m_1 and (r, s_2) is the signature of m_2, we have

$$s_1 = [h(m_1) - xr]k^{-1} \bmod (p-1)$$

$$s_2 = [h(m_2) - xr]k^{-1} \bmod (p-1)$$

therefore

$$(s_1 - s_2)k = [h(m_1) - h(m_2)] \bmod (p-1).$$

Because m_1 is different from m_2, $s_1 - s_2$ is not equal to $0 \bmod (p-1)$.

Finally, we have

$$k = [h(m_1) - h(m_2)](s_1 - s_2)^{-1} \bmod (p-1).$$

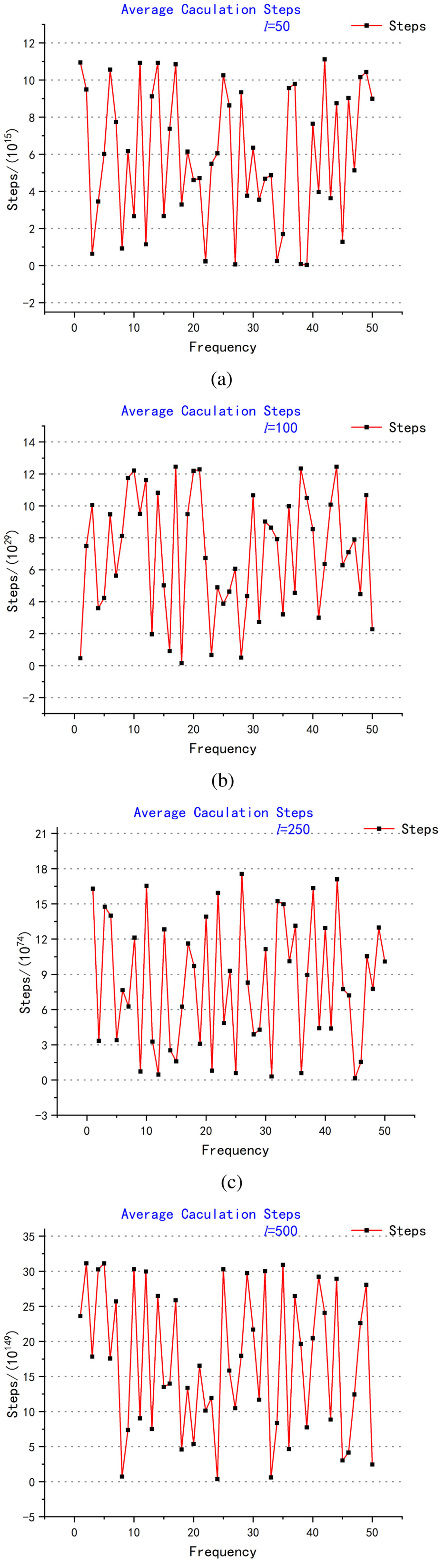
From the above process, it is easy to find that if using the same random number k when signing m_1 and m_2, k can be easily obtained by the adversary, which then will leak out other private information such as r_i and s_i. Therefore, herein, $k_i, k_1 \neq k_2 \neq \cdots \neq k_t, i = 1, \ldots, t$.

Besides, in a medical searching task, there are only two rounds of interactions between PLH and iDD. Thus, to the searching device, it can obtain the result once sending out the encrypted trait vector. Therefore, PMRSS is efficient for intelligent medical service in reality.

B. Experiment

We run the experiment on Lenovo r720 with a single core of a 2.80 GHz Intel Core i7-7700HQ and 16 GB of RAM. As we known, an adversary Λ who is aiming to obtain patient or iDoctor's information need to steal messages from the interactions of the scheme. There are two rounds of interactions in a searching task and the exposed messages in the whole processes are Q^*, s_i^L, s_i^H, $E(m_i)$.

First, from the first round of interaction, it is impossible to get anything from Q^*, which is protected by the secure hash function $f(\cdot)$. In the second interaction, s_i^L and s_i^H are protected by the secure blind signature algorithm. In terms of $E(m_i)$, the adversary Λ only can guess it a bit by a bit with 1/2 accuracy per bit.



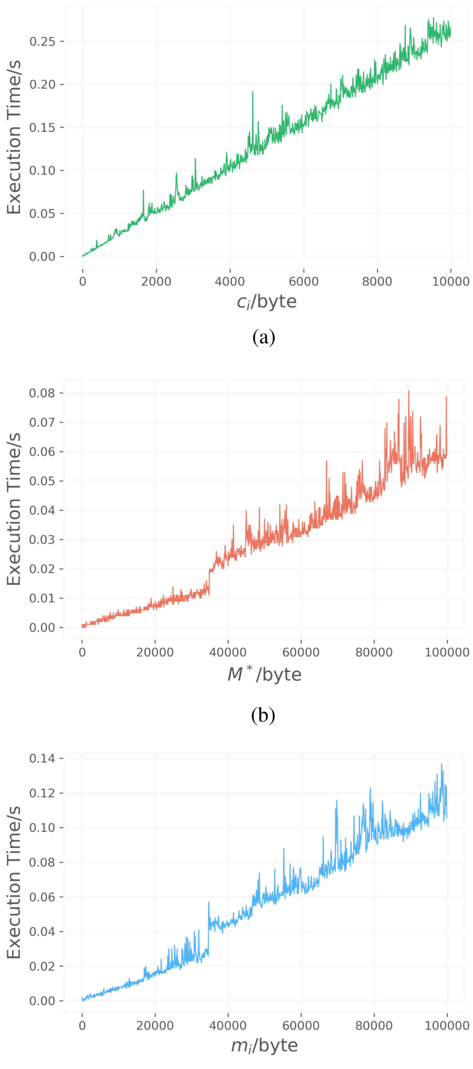
(a)

(b)

(c)

(d)

Fig. 4. Average Calculation Steps. (a) Average calculation steps in length 50. (b) Average calculation steps in length 100. (c) Average calculation steps in length 250. (d) Average calculation steps in length 500.



(a)

(b)

(c)

Fig. 5. Execution time. (a) Execution time with c_i. (b) Execution time with M^*. (c) Execution time with m_i.

In our experiment, after obtaining $E(m_i) = G(m_i \| s_i \| i) \oplus (0^l \| c_i)$, the adversary Λ applies XOR operation to $E(m_i)$ and a binary string r, expecting to output $(0^l \| c_i)$. Due to the randomness of $G(\cdot)$, there is no better way for Λ to get c_i than brute ergodic attack. If Λ can successfully determine the first l bits of the binary string r, that is, $\beta[1-l, r]$, which satisfies that

$$\beta[1-l, r \oplus E(m_i)] = 0.$$

Then, the adversary Λ succeeds to attack the scheme. Herein, l is called the zero-label. In the experiment, we assume that it needs y steps to construct such string r starting from 0^l. Obviously, the maximal steps in the worst situation for successfully guessing r is 2^l.

We set up four groups of experiments for different length l. In every single group, we carry out 50 random tests to show the average level steps for Λ to successfully guess r. The four figures Fig. 4(a)–(d) are average calculation steps for the adversary Λ in length l of 50, 100, 250, 500, respectively. Abscissa represents the number of experiments and ordinate represents the number of average calculation steps.

From the experiments, the upper bound of average calculation step for $l = 50$ is 1.20E+15; the upper bound of average calculation step for $l = 100$ is 1.40E+30; the upper bound of average calculation step for $l = 250$ is 1.80E+75; the upper bound of average calculation step for $l = 500$ is 3.50E+150. It is reasonable to deduce that when l is large, the average calculation step is approaching infinity. Besides, even if the adversary succeeds in obtaining a target c_v on the first try if he is very lucky just like winning the lottery, it is meaningless to other attacks. Therefore, the results show that the cost is extremely too high for an adversary to carry out brute ergodic attack. Users can also adjust the length of the zero-label l according to the security requirements to resist different levels of brute ergodic attack.

Besides, we also test the execution time of the proposed scheme. Generally speaking, the execution time of the ciphertext matching is very short, which can be completed at millisecond level. From the perspective of the relationships between the execution time and the matching content, according to Fig. 5(a)–(c), it can be seen that the execution time is directly proportional to the length of the content, which will increase along with the bit length of c_i, M^*, and m_i. Specifically, Fig. 5(a) shows that the execution time is proportional to the length of c_i, and when the length of c_i changes from 0 to 10 000, the execution time changes from 0 to 0.25 s. Fig. 5(b) shows that the execution time is proportional to the length of M^*, and when the length of M^* changes from 0 to 100 000, the execution time changes from 0 to 0.08 s. Fig. 5(c) shows that the execution time is proportional to the length of m_i, and when the length of m_i changes from 0 to 100 000, the execution time changes from 0 to 0.14 s.

IV. Conclusion

In this article, we considered the problem of how to securely search a medical diagnosis report from iDoctor in IoT healthcare while protecting the security of both current patient's privacy and iDD that consists of previous patients' cases. By applying ELGamal Digital Signature, we proposed a PMRSS to securely search the diagnosis report by only two rounds of interactions without leaking out any other information of the two parties. Moreover, we also have a detailed analysis about the security to show that PMRSS meets the security goals. In the future, how to compress and optimize the iDoctor's database and standardize patient's requirements are the promising issues to work on.

References

[1] Y. Zhang, R. Gravina, H. Lu, M. Villari, and G. Fortino, "Pea: Parallel electrocardiogram-based authentication for smart healthcare systems," *J. Netw. Comput. Appl.*, vol. 117, pp. 10–16, 2018.

[2] Y. Zhang, M. Chen, D. Huang, D. Wu, and Y. Li, "idoctor: Personalized and professionalized medical recommendations based on hybrid matrix factorization," *Future Gener. Comput. Syst.*, vol. 66, pp. 30–35, 2017.

[3] K. Yu, L. Tan, X. Shang, J. Huang, G. Srivastava, and P. Chatterjee, "Efficient and privacy-preserving medical research support platform against covid-19: A blockchain-based approach," *IEEE Consum. Electron. Mag.*, vol. 10, no. 2, pp. 111–120, 1 Mar. 2021.

[4] K. P. Yu, L. Tan, M. Aloqaily, H. Yang, and Y. Jararweh, "Blockchain-enhanced data sharing with traceable and direct revocation in IIoT," *IEEE Trans. Ind. Informat.*, to be published, doi: 10.1109/TII.2021.3049141.

[5] D. Vasan, M. Alazab, S. Venkatraman, J. Akram, and Z. Qin, "Mthael: Cross-architecture IoT Malware detection based on neural network advanced ensemble learning," *IEEE Trans. Comput.*, vol. 69, no. 11, pp. 1654–1667, Nov. 2020.

[6] S. Sriram, R. Vinayakumar, V. Sowmya, M. Alazab, and K. Soman, "Multi-scale learning based malware variant detection using spatial pyramid pooling network," in *Proc. IEEE INFOCOM IEEE Conf. Comput. Commun. Workshops*, 2020, pp. 740–745.

[7] S. Sriram, R. Vinayakumar, M. Alazab, and K. Soman, "Network flow based IoT Botnet attack detection using deep learning," in *Proc. IEEE INFOCOM Conf. Comput. Commun. Workshops*, 2020, pp. 189–194.

[8] R. Gellman, "Data privacy law: A study of United States data protection: By Paul M. Schwartz and Joel R. Reidenberg. Charlottesville, VA: Michie, 1996," *J. Government Inf. Quart.*, vol. 14, no. 2, pp. 215–217, 1997.

[9] H. Nissenbaum, *Privacy in Context: Technology, Policy, and the Integrity of Social Life*. Stanford, CA, USA: Stanford Univ. Press, 2009.

[10] J. Ko *et al.*, "Medisn: Medical emergency detection in sensor networks," *ACM Trans. Embedded Comput. Syst.*, vol. 10, no. 1, pp. 1–29, 2010.

[11] H. Wang, J. Gong, Y. Zhuang, H. Shen, and J. Lach, "Healthedge: Task scheduling for edge computing with health emergency and human behavior consideration in smart homes," in *Proc. IEEE Int. Conf. Big Data*, 2017, pp. 1213–1222.

[12] P. Hao and X. Wang, "A phy-aided secure IoT healthcare system with collaboration of social networks," in *Proc. IEEE 86th Veh. Technol. Conf.*, 2017, pp. 1–6.

[13] Y. Yang, X. Zheng, W. Guo, X. Liu, and V. Chang, "Privacy-preserving smart IoT-based healthcare big data storage and self-adaptive access control system," *Inf. Sci.*, vol. 479, pp. 567–592, 2019.

[14] M. Shuai, B. Liu, N. Yu, and L. Xiong, "Lightweight and secure three-factor authentication scheme for remote patient monitoring using on-body wireless networks," *Secur. Commun. Netw.*, vol. 2019, 2019, Art. no. 814508, doi: 10.1155/2019/8145087.

[15] T. Kumar, A. Braeken, A. D. Jurcut, M. Liyanage, and M. Ylianttila, "Age: Authentication in gadget-free healthcare environments," *Inf. Technol. Manage.*, vol. 21, pp. 95–114, 2020, doi: 10.1007/s10799-019-00306-z.

[16] S. Binu, M. Misbahuddin, and J. Paulose, "A signature-based mutual authentication protocol for remote health monitoring," *SN Comput. Sci.*, vol. 1, no. 1, pp. 1–14, 2020.

[17] K. Sowjanya, M. Dasgupta, and S. Ray, "An elliptic curve cryptography based enhanced anonymous authentication protocol for wearable health monitoring systems," *Int. J. Inf. Secur.*, vol. 19, no. 1, pp. 129–146, 2020.

[18] P. Gope, Y. Gheraibia, S. Kabir, and B. Sikdar, "A secure IoT-based modern healthcare system with fault-tolerant decision making process," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 3, pp. 862–873, Mar. 2021.

[19] C. Feng, K. Yu, M. Aloqaily, M. Alazab, Z. Lv, and S. Mumtaz, "Attribute-based encryption with parallel outsourced decryption for edge intelligent IoV," *IEEE Trans. Veh. Technol.*, vol. 69, no. 11, pp. 13784–13795, Nov. 2020.

[20] K. Yu, M. Arifuzzaman, Z. Wen, D. Zhang, and T. Sato, "A key management scheme for secure communications of information centric advanced metering infrastructure in smart grid," *IEEE Trans. Instrum. Meas.*, vol. 64, no. 8, pp. 2072–2085, Aug. 2015.

[21] C. Gentry, *A Fully Homomorphic Encryption Scheme*. Stanford, CA, USA: Stanford Univ., 2009.

[22] D. P. Rajan, S. J. Alexis, and S. Gunasekaran, "Dynamic multi-keyword based search algorithm using modified based fully homomorphic encryption and prim's algorithm," *Cluster Comput. J. Netw. Softw. Tools Appl.*, vol. 22, no. 5, pp. 11411–11424, Sep. 2019.

[23] W. Ogata and K. Kurosawa, "Oblivious keyword search," *J. Complexity*, vol. 20, no. 2/3, pp. 356–371, Apr.–Jun. 2004.

[24] C. Dwork, "Differential privacy: A survey of results," in *Proc. Int. Conf. Theory Appl. Models Comput.*, 2008, pp. 1–19.

[25] E. Karnin, J. Greene, and M. Hellman, "On secret sharing systems," *IEEE Trans. Inf. Theory*, vol. 29, no. 1, pp. 35–41, Jan. 1983.

[26] A. O'Neill, "Deterministic and efficiently searchable encryption," in *Proc. Int. Cryptol. Conf.*, 2006, pp. 535–552.

[27] W. He, G. Yan, and L. Da Xu, "Developing vehicular data cloud services in the IoT environment," *IEEE Trans. Ind. Informat.*, vol. 10, no. 2, pp. 1587–1595, May 2014.

[28] L. Qi *et al.*, "Privacy-aware data fusion and prediction with spatial-temporal context for smart city industrial environment," *IEEE Trans. Ind. Informat.*, vol. 17, no. 6, pp. 4159–4167, Jun. 2021.

[29] M. Hao, H. Li, X. Luo, G. Xu, H. Yang, and S. Liu, "Efficient and privacy-enhanced federated learning for industrial artificial intelligence," *IEEE Trans. Ind. Informat.*, vol. 16, no. 10, pp. 6532–6542, Oct. 2020.

[30] T. ElGamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," *IEEE Trans. Inf. Theory*, vol. 31, no. 4, pp. 469–472, Jul. 1985.

Yi Sun (Member, IEEE) received the Ph.D. degree in computer science and technology from the Beijing University of Posts and Telecommunications, Beijing, China, in 2015.

She is currently a Lecturer with the School of Computer Science (National Pilot Software Engineering School), Beijing University of Posts and Telecommunications. Her research interests include information security, secure multiparty computation, blockchain, artificial intelligence, and Internet of Things.

Dr. Sun was the Associate Editor for *Wireless Communications and Mobile Computing Journal*. She is a Guest Editor for Electronic Markets, ACM/Springer Mobile Networks and Applications, etc. She is a PC of 2020&2021 IEEE International Conference on Multimedia and Expo, 2019 DSC.

Jie Liu received the B.E. degree in Internet of Things engineering from Nanchang Hangkong University, Nanchang, China, in 2019, and the master's degree in cyberspace security from the Beijing University of Posts and Telecommunications, Beijing, China.

His research interests include IoT device identification, privacy-preserving data mining, artificial intelligence, and information security.

Keping Yu (Member, IEEE) received the M.E. and Ph.D. degrees in global information and telecommunication studies from Waseda University, Tokyo, Japan, in 2012 and 2016, respectively.

He was a Research Associate and a Junior Researcher with the Global Information and Telecommunication Institute, Waseda University, from 2015 to 2019 and 2019 to 2020, respectively, where he is currently a Researcher. His research interests include smart grids, information-centric networking, artificial intelligence, blockchain, and information security.

Dr. Yu has hosted and participated in more than ten projects, is involved in many standardization activities organized by ITU-T and ICNRG of IRTF, and has contributed to ITU-T Standards Y.3071 and Supplement 35. He was a recipient of the Best Paper Award from ITU Kaleidoscope 2020, the Student Presentation Award from JSST 2014. He has authored 100+ publications including papers in prestigious journal/conferences such as the IEEE WirelComMag, NetMag, TFS, IoTJ, TII, T-ITS, TVT, TNSE, TGCN, CEMag, IoTMag, ICC, GLOBECOM etc. He is an Editor for the IEEE OPEN JOURNAL OF VEHICULAR TECHNOLOGY (OJVT).

Mamoun Alazab (Senior Member, IEEE) received the Ph.D. degree in computer science from the Federation University of Australia, School of Science, Information Technology and Engineering, Ballarat, Australia, in 2012.

He is currently an Associate Professor with the College of Engineering, IT and Environment, Charles Darwin University, Australia. He is a cyber security Researcher and Practitioner with industry and academic experience. His research is multidisciplinary that focuses on cyber security and digital forensics of computer systems with a focus on cybercrime detection and prevention. He has more than 150 research papers in many international journals and conferences.

Dr. Alazab is the founding chair of the IEEE Northern Territory (NT) Subsection.

Kaixiang LIN received the B.E. degree in Intelligence Science and Technology from Beijing University of Posts and Telecommunications, Beijing, China, in 2020. He is currently a master of cyberspace security in Beijing University of Posts and Telecommunications. His interests include network security, privacy-preserving intelligent healthcare.

His research interests include network security, privacy-preserving intelligent healthcare.