A Privacy-Preserving Health Data Aggregation Scheme

Yining Liu[1], Gao Liu[2], Chi Cheng[3], Zhe Xia[4], and Jian Shen[5]
[1] Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin, Guangxi 541004 - China
[e-mail: ynliu@guet.edu.cn]
[2] School of Mathematics and Computational Science, Guilin University of Electronic Technology
Guilin, Guangxi 541004 - China
[e-mail: gaoliu9865@gmail.com]
[3] School of Computer Science, China University of Geoscience, Wuhan, Hubei 430074 - China
[e-mail: chengchizz@gmail.com]
[4] School of Computer Science and Technology, Wuhan University of Technology, Wuhan, Hubei 430070 - China
[e-mail: xiazhe@whut.edu.cn]
[4] School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing, Jiangsu 210044 -China
[e-mail: s_shenjian@126.com]
*Corresponding author: Yining Liu

Received May 16, 2016; accepted June 27, 2016; published August 31, 2016

Abstract

Patients' health data is very sensitive and the access to individual's health data should be strictly restricted. However, many data consumers may need to use the aggregated health data. For example, the insurance companies needs to use this data to setup the premium level for health insurances. Therefore, privacy-preserving data aggregation solutions for health data have both theoretical importance and application potentials. In this paper, we propose a privacy-preserving health data aggregation scheme using differential privacy. In our scheme, patients' health data are aggregated by the local healthcare center before it is used by data comsumers, and this prevents individual's data from being leaked. Moreover, compared with the existing schemes in the literature, our work enjoys two additional benefits: 1) it not only resists many well known attacks in the open wireless networks, but also achieves the resilience against the human-factor-aware differential aggregation attack; 2) no trusted third party is employed in our proposed scheme, hence it achieves the robustness property and it does not suffer the single point failure problem.

Keywords: Health Data Aggregation, Privacy-Preservation, Differential Privacy, Robustness

This work was partly supported by National Natural Science Foundation of China under grant Nos. 61363069 and 61301166, partly supported by Guangxi Natural Science Foundation (No. 2014GXNSFAA118364) and Innovation Project of Guangxi Graduate Education (No. YCSZ2015149).

http://dx.doi.org/10.3837/tiis.2016.08.023 **ISSN : 1976-7277**

1. Introduction

In the wireless body area network [1], the implanted or wearable biosensor can be used to measure the patients' health data, such as the temperature, the blood pressure, etc. In the authenticated manner [2-7], after the health data is collected, it will be transmitted to the doctor in the local healthcare center (*LHC*) in the authenticated manner. Therefore, the doctor can give precise diagnosis and treatment. Moreover, the aggregated health data has many real world applications. For example, the insurance company can analyze the aggregated result of the health data in a specific area, and then make a decision. However, if the health data of the patient is transmitted directly, the privacy will be violated, and this might have serious consequences, such as financial fines or even law prosecutions. For instance, with the knowledge of some people's poor body condition, the insurance company might refuse to provide the insurance service for them. Therefore, it is necessary to design a privacy-preserving health data aggregation scheme, which allows *LHC* to aggregate the health data in a designated region without knowing an individual one.

In order to ensure the privacy property, the individual health data should be encrypted or processed anonymously. As shown in **Fig. 1**, the patient transmits the processed health data to *LHC*, and the doctor in *LHC* can make the diagnosis and give the treatment due to the patient's data. Furthermore, *LHC* aggregates the received data, and sends the aggregated result to the healthcare cloud. Moreover, the data consumers can utilize the aggregated result which is stored in the healthcare cloud.

Although there are many existing works on data aggregation in the literature, the majority of them may suffer the human-factor-aware differential aggregation (HDA) attack [8], which aims to break the privacy. Moreover, many data aggregation schemes rely on a trusted entity to ensure confidentiality for the sensitive data, so that the robustness requirement is not satisfied in a high level because of the potential single point failure problem. In [9-11], using trusted gateway and operating center, the single data is protected by the homomorphic encryption technique. However, the privacy will be violated if the gateway and the operating center are not trusted. In [12], a one-way virtual ring is used for the aggregation. However, the aggregation operation will fail if any smart device of the ring breaks down. In 2014, Fan et al. proposed a data aggregation scheme [13] based on the subgroup decision assumption. However, each user's private key can be extracted from the public information in the registration phase, and this flaw has been resolved later [14]. Moreover, the privacy is preserved by the blind factor, which is distributed by an off-line trusted third party, and thus there exists the trust bottleneck in the proposed scheme. Therefore, many of the existing schemes need further improvement in order to suit the practical environment [15,16].

In this paper, we propose a health data aggregation scheme, which also allows *LHC* to aggregate the health data in a specific area without knowing a single one. The security of the proposed scheme is mainly based on the differential privacy [8] and the subgroup decision assumption [13]. Compared with other data aggregation schemes, the proposed scheme has two contributions: 1) The proposed scheme not only resists many well know attacks, such as external attack, internal attack, replay attack, impersonation attack and modification attack, but also it is robust against the new HDA attack. Therefore, our proposed scheme achieves a higher level of privacy. 2) The proposed scheme does not employ a trusted third party. Hence it achieves the robustness property and it does not suffer the single point failure problem.

The remainder of the paper is organized as follows: The necessary preliminaries are introduced in Section 2. Afterwards, the health data aggregation scheme is presented in Section 3, and its security and efficiency are analyzed in Section 4. Finally, the paper is concluded in Section 5.

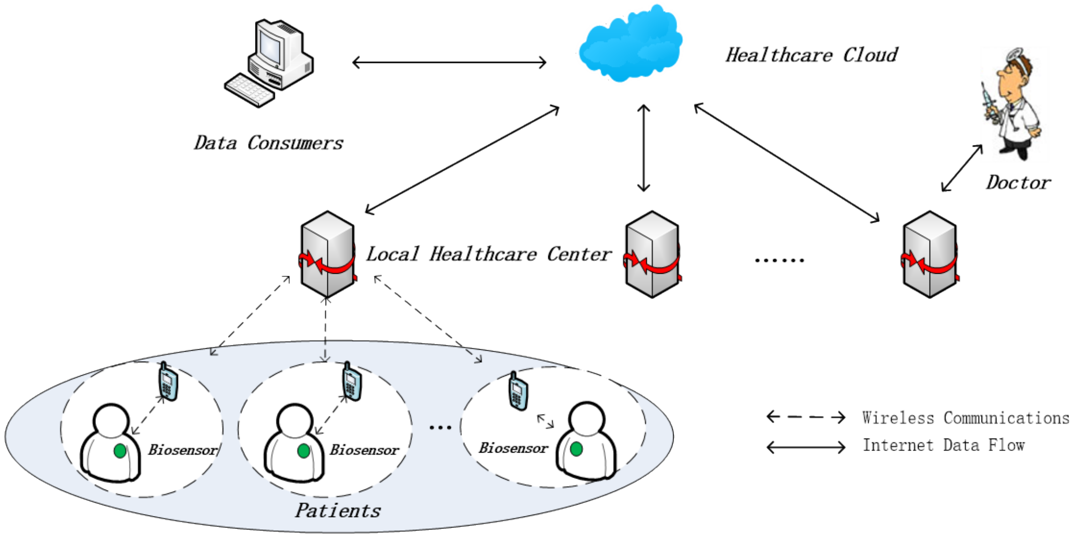

Fig. 1. Network model

2. Preliminaries

In this section, we describe the related assumptions and techniques.

- **Secure Hash Function**

Assume $h(x)$ is a secure hash function. It is computationally infeasible to extract a from a given value $h(a)$ or to find a pair of values (a, b) such that $h(a) = h(b)$ where $a \neq b$ [17].

- **Subgroup Decision Assumption**

Given an element x that belongs to a group G_0 with a composite order $N = q_1 q_2$, where q_1, q_2 are large prime numbers, it is computationally infeasible to decide if $x \in G_0$ is in a subgroup with order q_1 [18].

- **Discrete Logarithm Assumption**

Suppose g_2 is the generator of a cyclic multiplicative group G_1 with order q, it is computationally infeasible to compute $x = log_{g_2} y$ given $y = {g_2}^x$ [19].

- **Bilinear Pairing**

Suppose G_1 and G_2 are two cyclic multiplicative groups with order q, and g_2 is a generator of G_1. Furthermore, the discrete logarithm assumption holds both in G_1 and G_2. A bilinear map e: $G_1 \times G_1 \rightarrow G_2$ satisfies the following properties [20]:

Bilinear: For any $P, Q \in G_1$, $a, b \in {Z_q}^*$, $e(P^a, Q^b) = e(P, Q)^{ab}$ and $e(P, P) \neq 1_{G_2}$.

Non-degenerate: There exist $P, Q \in G_1$ such that $e(P, Q) \neq 1_{G_2}$.

Computable: For any $P, Q \in G_1$, there exists an efficient algorithm to compute $e(P, Q)$.

- **Gap Diffie-Hellman Group**

Assume that g_2 is the generator of a cyclic multiplicative group G_1 with the order q.

Computational Diffie-Hellman (CDH) problem: For any $a, b \in {Z_q}^*$, the CDH problem asks to derive ${g_2}^{ab}$ from the given $({g_2}^a, {g_2}^b)$.

Decision Diffie-Hellman (DDH) problem: For any $a, b, c \in {Z_q}^*$, given $({g_2}^a, {g_2}^b, {g_2}^c)$, the DDH problem asks to determine whether ${g_2}^{ab} = {g_2}^c$.

If the computational Diffie-Hellman problem is hard but the decision Diffie-Hellman problem is easy to solve in a cyclic multiplicative group G_1, G_1 is referred to as the gap Diffie-Hellman (GDH) group [21].

- **HDA attack**

Suppose that the health data of P_1, P_2, P_3, P_4, P_5 are aggregated, and P_5 is the target member. In addition, assume P_5 does not use the device in the time slot T_1 but uses it in the adjacent time slot T_2, and the health data of P_1, P_2, P_3, P_4 are relatively stable in these two time slots. Therefore, LHC can derive the health data of P_5 in the time slot T_2 by comparing the two aggregated results [8].

- **Assumption for Byzantine Agreement**

The classical assumption of the Byzantine literature (The classical assumption for Byzantine agreement) [22] is employed to resist against the collusion attack. In the assumption, the attacker might corrupt LHC, and compromise no more than $1/3$ patients. Finally, the attacker colludes with the compromised LHC and patients, and launches the collusion attack (i.e., HDA attack).

- **Differential Privacy**

In the query access, the differential privacy [22] is usually employed to achieve the privacy. By adding the proper Gaussian or exponentially distributed random noise, the administrator can obscure the true answer slightly before the query result is sent to the user. Furthermore, the similar inputs, which differ on a tiny entry, generate the indistinguishable outputs.

A randomized algorithm $\mathcal{K}$ is *ε-indistinguishability δ-approximation*: Given two data sets D_1 and D_2, which differ on at most one element, and all $S \subseteq Range(\mathcal{K})$, where $Range(\mathcal{K})$ consists of all possible values of $\mathcal{K}$.

$$Pr[\mathcal{K}(D_1) \in S] \leq e^{\varepsilon} Pr[\mathcal{K}(D_2) \in S] + \delta \tag{1}$$

If all computations are performed over a finite field, the unbiased binomial distribution $B(w, 1/2)$ [8] is employed to replace the Gaussian distribution. Afterwards, the following facts take the important roles in the proposed scheme.

Fact 1. Given the global sensitivity Δ (i.e., the interval of each patient's health data), and in order to make $B(w, 1/2)$ *ε-indistinguishability δ-approximation*, w should be at least $64\Delta^2 log(2/\delta)/\varepsilon^2$ [8].

Fact 2. If $V_i \sim B(w_i, pr)$ and $V_i, i = 1, 2, \cdots, n$ are independent and identically distributed, $\sum_{i=1}^{n} V_i \sim B(\sum_{i=1}^{n} w_i, pr)$.

3. Our Proposal

In this section, we present a novel aggregation scheme, where there only exists n patients and LHC in the specific area, and LHC can derive the summation of the patients' health data without the knowledge of the individual one. Some notations for the relavant parameters are defined in **Table 1**.

3.1 Initialization Phase

1. Given the pre-set security parameters ε, δ, which are determined by LHC due to the tradeoff between the security and the usability, LHC computes $w_n = \lceil 3w/2n \rceil$, where $w = 64\Delta^2 log(2/\delta)/\varepsilon^2$.

2. LHC chooses three large prime numbers q, q_1, q_2, and computes $N = q_1 q_2$.

3. From a cyclic multiplicative group G_0 of order N, LHC determines a generator g_0 and a random number $u \in G_0$, and computes $h = u^{q_2}$, $g_1 = {g_0}^{q_1}$. Then LHC chooses a generator g_2 of a cyclic multiplicative group G_1 with order q. Moreover, the subgroup decision assumption holds in G_0, and the discrete logarithm assumption holds in the GDH group G_1.

4. LHC keeps q_1, q_2 secretly, chooses a secure hash function $H(x)$ and a bilinear map $e(G_1, G_1) \rightarrow G_2$, and publishes

$$\{N, q, g_0, g_2, h, w_n, H(x), e\} \tag{2}$$

5. Each patient P_i registers at LHC using the public key $y_i = {g_2}^{x_i} \in G_1$ with the identifier ID_i. Finally, LHC stores $\{ID_i, y_i\}$ in its database for the verification in the Aggregation Phase.

Table 1. Notation for related parameters

Notation	Definition
P_i	The patients in the specific area, where $i = 1, 2, \cdots, n$.
ID_i	The identifier of P_i.
x_i	The private key of P_i.
y_i	The public key of P_i.
H	The secure hash function, $H: \{0,1\}^* \rightarrow G_1$.
t	The time for the aggregation.
m_i	The health data collected by P_i at time t.
Δ	The interval of m_i.
$B(w_n, 1/2)$	The unbiased binomial distribution.

3.2 Aggregation Phase

1. P_i collects the health data $m_i \in [0, 1, \cdots, \Delta]$ at time t, then chooses $v_i \sim B(w_n, 1/2)$ and ${r_i}' \in {Z_N}^*$ randomly. P_i computes the ciphertext $CT_i = {g_0}^{m_i + v_i} h^{{r_i}'}$ and the corresponding signature $\sigma_i = H(t||CT_i)^{x_i}$, and sends $\{ID_i, CT_i, \sigma_i\}$ to LHC.

2. With the received $\{ID_i, CT_i, \sigma_i\}$, LHC extracts P_i's public key y_i with ID_i in the database, and verifies them by checking $e(\sigma_i, g_2) = e(H(t||CT_i), y_i)$, $i = 1, 2, \cdots, n$. With the selected n random numbers $k_i \in {Z_q}^*, i = 1, 2, \cdots, n$, LHC checks the equation $\prod_{i=1}^{n} e\left({\sigma_i}^{k_i}, g_2\right) = e\left(\prod_{i=1}^{n} H(t||CT_i)^{k_i}, y_i\right)$ to speed up the verification.

3. If all the verifications hold, LHC computes $V = (\prod_{i=1}^{n} CT_i)^{q_1} = {g_1}^{\sum_{i=1}^{n} m_i + v_i}$. Furthermore, LHC derives $\sum_{i=1}^{n} m_i + v_i$ from V with the base g_1 using the Pollard's lambda method, which costs the expected polynomial time $\tilde{O}\left(\sqrt{n(\Delta + w_n)}\right)$ [16, 23] due to the non-cryptographic interval $0 < \sum_{i=1}^{n} m_i + v_i < n(\Delta + w_n)$. As a consequence, LHC outputs the approximate aggregated result $\sum_{i=1}^{n} m_i + v_i - \lceil n w_n/2 \rceil$, where $n w_n/2$ is the expectation of the added noise summation $\sum_{i=1}^{n} v_i$. Each step is depicted in **Fig. 2**.

3.3 Correctness of Health Data Aggregation

The parameter g_0 is the generator of the cyclic multiplicative group G_0 with order N, and thus ${g_0}^N = 1$. Furthermore, u belongs to G_0, and there thus exists a number $\alpha \in {Z_N}^*$ satisfying that $u = {g_0}^\alpha$. Therefore, $u^N = ({g_0}^\alpha)^N = ({g_0}^N)^\alpha = 1^\alpha = 1$. The correctness of the health data aggregation is shown as follows:

$$
\begin{aligned}
V &= (\prod_{i=1}^{n} CT_i)^{q_1} = {g_0}^{q_1 \sum_{i=1}^{n} m_i + v_i} h^{q_1 \sum_{i=1}^{n} r_i'} \\
&= {g_1}^{\sum_{i=1}^{n} m_i + v_i} u^{q_1 q_2 \sum_{i=1}^{n} r_i'} \\
&= {g_1}^{\sum_{i=1}^{n} m_i + v_i} u^{N \sum_{i=1}^{n} r_i'} \\
&= {g_1}^{\sum_{i=1}^{n} m_i + v_i}.
\end{aligned}
$$

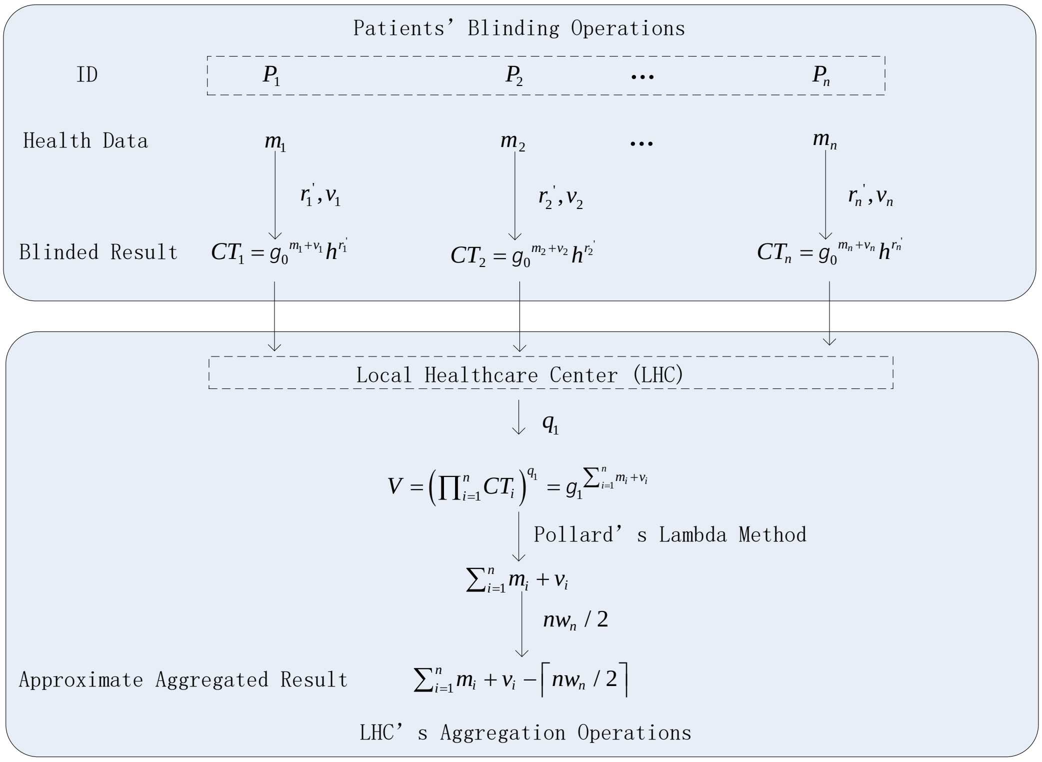

Fig. 2. Aggregation

4. Analysis

In this section, we provide security and efficiency analysis of our proposed scheme. Moreover, we briefly discuss its usability in real world applications.

4.1 Security Analysis

In this subsection, we demonstrate that the proposed scheme resists against not only the well known attacks (i.e., the external attack, the internal attack, the impersonation attack, the modification attack, and the replay attack), but also the new HDA attack. Moreover, it is shown that the robustness is achieved in the proposed scheme.

- **Privacy-preservation**

Generally speaking, the attackers can be divided into two categories: the inside attacker and the outside attacker. The inside attacker includes *LHC* and the patients who attempt to violate the privacy of other patients, and the outside attacker is an illegal party, who does not involve in the proposed scheme.

Scenario 1. The proposed scheme can resist against the external attack, i.e., it is computationally infeasible for an outside adversary to obtain m_i from CT_i.

Proof The ciphertext $CT_i = {g_0}^{m_i+v_i}h^{{r_i}'}$ can be eavesdropped by the outsider. If the adversary manages to derive m_i from CT_i, he should know $v_i, {r_i}'$ or v_i, q_1. Unfortunately, $v_i, {r_i}'$ are secretly hold by the patient P_i, and q_1 is privately hold by *LHC*.

Scenario 2. The proposed scheme can resist against the internal attack, i.e., it is computationally infeasible for an internal adversary to extract m_i from CT_i.

Proof The inside adversary (other patient $P_j, j \neq i$) cannot extract m_i from CT_i successfully, since he has no idea about $v_i, {r_i}'$ or v_i, q_1. Furthermore, if *LHC* succeeds in deriving m_i, he should at least learn v_i which is randomly selected by the patient P_i. Therefore, the proposed scheme can resist against the internal attack.

Scenario 3. The proposed scheme can resist against the HDA attack.

Suppose there exist 3 patients P_1, P_2, P_3 in a specific area, and the health data m_1, m_2 of P_1, P_2 are relatively stable at two adjacent time slots T_1 and T_2. However, P_3 uses the medical device at time slot T_1, but does not use it at time slot T_2. By comparing the aggregated results at the two time slots, it is impossible for the adversary to derive the health data m_3 of P_3 at time slot T_1.

Proof The noise aggregated result at the time slots T_1 and T_2 are $M_1 = \sum_{i=1}^{3} m_i + V_1$ and $M_2 = \sum_{i=1}^{2} m_i + V_2$ respectively, where $V_1, V_2 \sim B(3w_3, 1/2)$. It is infeasible for the adversary to derive m_3 by computing $M_1 - M_2$, since $B(3w_3, 1/2)$ is ε *-indistinguishability* δ*-approximation*.

Therefore, the proposed scheme resists against not only the external attack and the internal attack, but also the new HDA attack. As a consequence, the privacy property has been enhanced to a higher level compared with existing schemes.

- **Resilience against impersonation attack**

Scenario 4. The proposed scheme can resist against the impersonation attack, i.e., it is infeasible for the adversary to impersonate the legal patient P_i to provide *LHC* with the valid message.

Proof To impersonate P_i, the adversary should have knowledge about the private key x_i of P_i. Given the public key $y_i = {g_2}^{x_i}$ and signature $\sigma_i = H(t||CT_i)^{x_i}$, it is infeasible in polynomial time to extract x_i due to the discrete logarithm assumption in G_1. As a result, the adversary cannot launch the impersonation attack.

- **Resilience against modification attack**

Scenario 5. The proposed scheme can resist against the modification attack, i.e., if the adversary modifies a message being sent to *LHC*, and transmits the modified result to *LHC*, it can be detected by *LHC*.

Table 2. Security features comparision of related works

		Our scheme	Li et al.'s scheme [10]	Fan et al.'s scheme [13]	He et al.'s scheme [14]
PPR	REX	Yes	Yes	Yes	Yes
	RIN	Yes	No	Yes	Yes
	RHD	Yes	No	No	No
RIM		Yes	Yes	No	Yes
RMO		Yes	Yes	Yes	Yes
RRE		Yes	Yes	Yes	Yes
ROU		Yes	No†	No††	No††

PPR: Privacy-Preservation
REX: Resilience against External Attack
RIN: Resilience against Internal Attack
RHD: Resilience against HDA Attack
RIM: Resilience against Impersonation Attack
RMO: Resilience against Modification Attack
RRE: Resilience against Replay attack
ROU: Robustness
†: Relying on On-line Trusted Third Party
††: Relying on Off-line Trusted Third Party

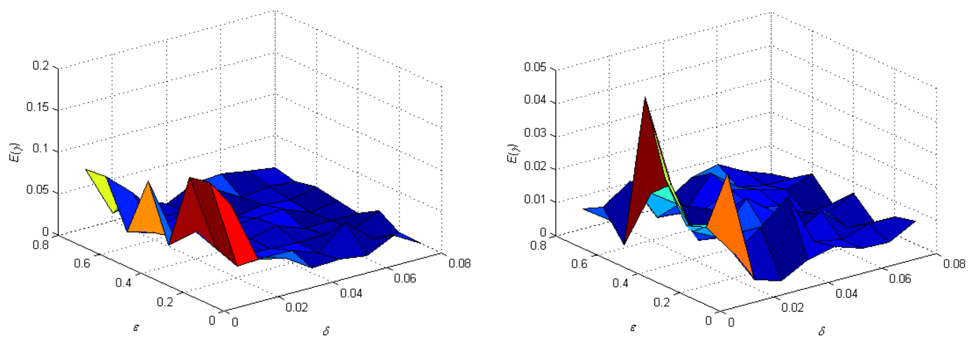

Fig. 3. Relative error. **(a)** When $n = 3000$, $\Delta = 5$ and $\sum_{i=1}^{3000} m_i = 7500$. **(b)** When $n = 6000$, $\Delta = 5$ and $\sum_{i=1}^{6000} m_i = 15000$.

Proof Suppose the adversary modifies $\{ID_i, CT_i, \sigma_i\}$ into $\{ID_i, CT_i{}', \sigma_i{}'\}$, and tries to enable the modified result to pass the verification $e(\sigma_i{}', g_2) = e(H(t||CT_i{}'),\ y_i)$.

Except for guessing the correct σ_i', it is impossible for the adversary to determine σ_i' from $e(\sigma_i', g_2) = e(H(t||CT_i{}'),\ y_i)$ for the given CT_i', since G_1 is a GDH group [13]. Similarly, for the given σ_i', it is also infeasible to obtain $CT_i{}'$ from $e(\sigma_i', g_2) = e(H(t||CT_i{}'),\ y_i)$ due to the GDH group G_1 and the feature of the secure hash function.

As a consequence, if the adversary transmits a modified result, it can be detected by *LHC*. Therefore, the proposed scheme can resist against the modification attack.

- **Resilience against replay attack**

Scenario 6. The proposed scheme can resist against the replay attack, i.e., at time t_2, the adversary sends a message $\{ID_i, CT_i{}^1, \sigma_i{}^1\}$ which has been used at time t_1 $(t_1 < t_2)$, and this can be detected by *LHC*.

Proof To launch the replay attack, the adversary provides *LHC* with the used

$\{ID_i, CT_i{}^1, \sigma_i{}^1\}$ at t_2. It can be detected by *LHC*, since $e(\sigma_i{}^1, g_2) \neq e(H(t_2||CT_i{}^1), y_i)$.

- **Robustness**

Scenario 7. The proposed scheme achieves the robustness.

Proof The proposed scheme does not rely on any trusted third party, and the duty of *LHC* is only to verify the patient's message and aggregate the health data in a specific area. Therefore, anyone, who has the knowledge of q_1, can verify the message from the patients, and extract the aggregated result. As a result, the trust bottleneck is eliminated, so that the robustness is achieved in the proposed scheme.

Moreover, the security features of the proposed scheme are compared with several works [10, 13, 14], and the comparison is demonstrated in **Table 2**.

4.2 Performance Evaluation

We mainly compare the aggregation performance of the proposed scheme with the related works in [10, 13, 14]. Assume there exists n patients in the specific area. We only count the expensive computation, such as modular multiplication, modular exponentiation, Pollard's lambda method, Paillier cryptosystem decryption, and pairing operation. In addition, the time cost for the related computations is listed in **Table 3**, and $T_e \approx T_{pc} \approx 1.5T_{pl}$ [13].

As for the aggregation efficiency, the comparison result is shown in **Table 4**. Obviously, the aggregation efficiency of the proposed scheme is comparable to that of Li et al.'s scheme [10] and He et al.'s scheme [14], and it is higher than that of Fan et al.'s scheme [13].

Table 3. Notation for time cost

Notation	Definition
T_e	Modular exponentiation computation time cost.
T_m	Modular multiplication computation time cost.
T_{pl}	Pollard's lambda method time cost.
T_{pc}	Paillier cryptosystem decryption time cost.
T_{po}	Pairing operation time cost.

Table 4. Time cost comparison of aggregation

	n users (patients)	*LHC*(Aggregator)	Total
Our scheme	$2nT_e + nT_m$	$T_e + (n-1)T_m + T_{pl}$	$(2n+1)T_e + (2n-1)T_m + T_{pl}$
Li et al.'s scheme [10]	$2nT_e + nT_m$	$(n-1)T_m + T_{pc}$	$2nT_e + (2n-1)T_m + T_{pc}$
Fan et al.'s scheme [13]	$3nT_e + 2nT_m$	$3T_e + nT_m + T_{pl}$	$(3n+3)T_e + 3nT_m + T_{pl}$
He et al.'s scheme [14]	$2nT_e + nT_m$	$T_e + nT_m + T_{pl}$	$(2n+1)T_e + 2nT_m + T_{pl}$

4.3 Utility Analysis

Suppose the aggregation operation involves n patients in the designated area, the approximate aggregated result is $\sum_{i=1}^{n} m_i + v_i - \lceil nw_n/2 \rceil$, and the overall relative error is denoted as $E(\gamma) = |\sum_{i=1}^{n} v_i - \lceil nw_n/2 \rceil| / \sum_{i=1}^{n} m_i$. When the interval of the added noise is smaller, the relative error thus is also smaller. Moreover, $E(\gamma)$ is regarded as a binary function of the security parameters ε and δ, and $E(\gamma)$ is roughly reduced if ε and δ increase simultaneously. Therefore, we can choose the proper ε, δ and $E(\gamma)$ to balance the security and the usability.

In order to achieve the tradeoff between security and usability, we can roughly determine ε and δ due to a given relative error $E(\gamma)$. For simplicity, when $n = 3000,\ \Delta = 5,\ \sum_{i=1}^{3000} m_i = 7500$ and $n = 6000,\ \Delta = 5,\ \sum_{i=1}^{6000} m_i = 15000$, the binary function $E(\gamma)$ with respect to ε and δ are shown in **Fig. 3 (a)** and **Fig. 3 (b)**, respectively. In **Fig. 3 (a)**, if $E(\gamma) = 0.05$, the rough parameters are determined, i.e., $\varepsilon = 0.3,\ \delta = 0.03$. Meanwhile, $\varepsilon = 0.5,\ \delta = 0.05$ can also be determined when $E(\gamma) = 0.01$ in **Fig. 3 (b)**. In **Fig. 4**, 200 experiments show that almost all the relative errors fall in the pre-determined interval $[0, 0.05]$ with $n = 3000, \varepsilon = 0.3,\ \delta = 0.03, \Delta = 5,\ \sum_{i=1}^{3000} m_i = 7500$, and $[0, 0.01]$ with $n = 6000,\ \varepsilon = 0.5,\ \delta = 0.05,\ \Delta = 5,\ \sum_{i=1}^{6000} m_i = 15000$. It suggests that the interval of $|\sum_{i=1}^{n} v_i - \lceil nw_n/2 \rceil|$ is relatively stable and small for the aggregated expectation $\sum_{i=1}^{n} m_i$ with the proper security parameters. As a result, before implementing the proposed scheme, we can determine the proper parameters ε, δ and $E(\gamma)$ to balance the security and the utility.

5. Conclusion

Based on the differential privacy and the subgroup decision assumption, we propose a privacy-preserving health data aggregation scheme. In the proposed scheme, the local healthcare center can aggregate the health data of the patients in a specific area without leaking the individual one. Moreover, the proposed scheme not only resists against the well known attacks, such as external attack, internal attack, impersonation attack, modification attack, and replay attack, but also overcomes the new HDA attack. Therefore, the privacy is preserved. Notably, no trusted third party is needed in the proposed scheme, such that there exists no trust bottleneck, and thus the robustness is achieved. Hence, the proposed scheme is more practical.

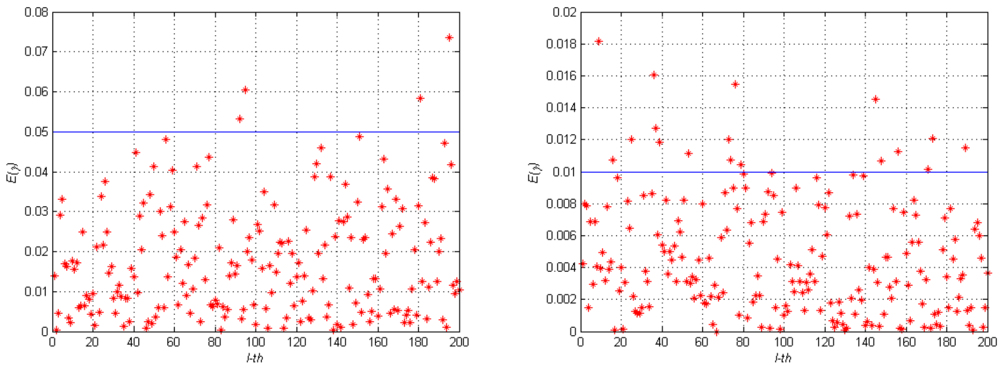

Fig. 4. Relative error. **(a)** 200 experiments when $n = 3000, \varepsilon = 0.3, \delta = 0.03, \Delta = 5$, and $\sum_{i=1}^{3000} m_i = 7500$. **(b)** 200 experiments when $n = 6000, \varepsilon = 0.5, \delta = 0.05, \Delta = 5$, and $\sum_{i=1}^{6000} m_i = 15000$.

References

[1] C. Hu, X. Liao and D. Chen, "Securing communications between external users and wireless body area networks," in *Proc. of the 2nd ACM Workshop on Hot Topics on Wireless Network Security and Privacy*, pp. 31-36, 2013. Article (CrossRef Link)

[2] D. He, C. Chen, S. Chan, J. Bu and P. Zhang, “Secure and lightweight network admission and transmission protocol for body sensor networks,” *IEEE Journal of Biomedical and Health Informatics*, vol. 17, no. 3, pp. 664-674, 2013. Article (CrossRef Link)
[3] T. Cao and J. Zhai, “Improved dynamic ID-based authentication scheme for telecare medical information systems,” *Journal of Medical Systmes*, vol. 37, no. 2, pp. 1-7, 2013. Article (CrossRef Link)
[4] H.Y. Lin, “On the security of a dynamic ID-based authentication scheme for telecare medical information systems,” *Journal of Medical Systems*, vol. 37, no. 2, pp. 1-5, 2013. Article (CrossRef Link)
[5] M. Raghavendra and K.B. Amit, “A privacy preserving secure and efficient authentication scheme for telecare medical information systems,” *Journal of Medical System*, vol. 39, 2015. Article (CrossRef Link)
[6] Zhangjie Fu, Xingming Sun, Qi Liu, Lu Zhou and Jiangang Shu, "Achieving Efficient Cloud Search Services: Multi-keyword Ranked Search over Encrypted Cloud Data Supporting Parallel Computing," *IEICE Transactions on Communications*, vol. E98-B, no. 1, pp.190-200, 2015. Article (CrossRef Link)
[7] Zheng Yuhui, Jeon Byeungwoo, Xu Danhua, Wu Q.M. Jonathan and Zhang Hui, "Image segmentation by generalized hierarchical fuzzy C-means algorithm," *Journal of Intelligent and Fuzzy Systems*, vol .28, no. 2, pp. 961-973, 2015. Article (CrossRef Link)
[8] W. Jia, H. Zhu, Z. Cao, X. Dong and C. Xiao, “Human-factor-aware privacy-preserving aggregation in smart grid,” *IEEE Systems Journal*, vol. 8, no. 2, pp. 598-607, 2014. Article (CrossRef Link)
[9] R. Lu, X. Liang, X. Li, X. Lin and X. Shen, “EPPA: An efficient and privacy-preserving aggregation scheme for secure smart grid communications,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1621-1632, 2012. Article (CrossRef Link)
[10] H. Li, X. Lin, H. Yang, X. Liang, R. Lu and X. Shen, “EPPDR: An efficient privacy-preserving demand response scheme with adaptive key evolution in smart grid,” *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 8, pp. 2053-2064, 2014. Article (CrossRef Link)
[11] K. Zhang, X. Liang, M. Baura, R. Lu and X. Shen, “PHDA: A priority based health data aggregation with privacy preservation for cloud assisted WBANs,” *Information Sciences*, vol. 284, pp. 130-141, 2014. Article (CrossRef Link)
[12] M. Badra and S. Zeadally, “Design and performance analysis of a virtual ring architecture for smart grid privacy,” *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 2, pp. 321-329, 2014. Article (CrossRef Link)
[13] C.I. Fan, S.Y. Huang and Y.L. Lai, “Privacy-enhanced data aggregation scheme against internal attackers in smart grid,” *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 666-675, 2014. Article (CrossRef Link)
[14] D. He, N. Kumar and J.H. Lee, “Privacy-preserving data aggregation scheme against internal attackers in smart grids,” *Wireless Networks*, vol. 22, no. 2, pp. 491-502, 2016. Article (CrossRef Link)
[15] Ping Guo, Jin Wang, Bing Li and Sungyoung Lee, “A Variable Threshold-value Authentication Architecture for Wireless Mesh Networks," *Journal of Internet Technology*, vol. 15, no. 6, pp. 929-936, 2014. Article (CrossRef Link)
[16] J. Shen, H. Tan, S. Moh, I. Chung, Q. Liu and X. Sun, “Enhanced Secure Sensor Association and Key Management in Wireless Body Area Networks,” *Journal of Communications and Networks*, vol. 17, no. 5, pp. 453-462, 2015. Article (CrossRef Link)
[17] J. Shao, “Efficient verifiable multi-secret sharing scheme based on hash function,” *Information Sciences*, vol. 278, pp. 104-109, 2014. Article (CrossRef Link)
[18] D. Boneh, E. Goh and K. Nissim, “Evaluating 2-DNF formulas on ciphertexts,” in *Proc. of Theory of Cryptography (LNCS)*, pp. 325-341, 2005. Article (CrossRef Link)
[19] C. Meshram, “An efficient ID-based cryptographic encryption based on discrete logarithm problem and integer factorization problem,” *Information Processing Letters*, vol. 115, pp. 351-358, 2015. Article (CrossRef Link)

[20] K.A. Shim, "An efficient ring signature scheme from pairings," *Information Sciences*, vol. 300, pp.63-69, 2015. Article (CrossRef Link)

[21] D. Boneh, B. Lynn and H. Shacham, "Short signatures from the Weil pairing," *Advances in Cryptology - ASIACRYPT*, pp. 514-532, 2001. Article (CrossRef Link)

[22] D. Cynthia, K. Kenthapadi, F. McSherry, I. Mironov and M. Naor, "Our data, ourselves: privacy via distributed noise generation," *Advances in Cryptology - EUROCRYPT*, pp. 486-503, 2006. Article (CrossRef Link)

[23] J.M. Pollard, "Monte carlo methods for index computation (mod *p*)," *Mathematics of Computation*, vol. 32, no. 143, pp. 918-924, 1978. Article (CrossRef Link)

Yining Liu is currently a professor in School of Computer and Information Security, Guilin University of Electronic Technology(GUET), China. He received the B.S. degree in Applied Mathematics from Information Engineering University, Zhengzhou, China, in 1995, the M.S. in Computer Software and Theory from Huazhong University of Science and Technology, Wuhan, China, in 2003, and the Ph.D. degree in Mathematics from Hubei University, Wuhan, China, in 2007. His research interests include the analysis of information security protocol, the smart grid, e-voting, and privacy-preserving data aggregation.

Gao Liu is now pursuing his M.S. degree in School of Mathematics and Computational Sciences, Guilin University of Electronic Technology, China, from 2013. He received the B.S. degree in Applied Mathematics from Yibin University, Sichuan, China, in 2013. His research interests focus on e-voting, micropayment, and data aggregation.

Chi Cheng is an Associate Professor in School of Computer Science, China University of Geosciences, China, and is also an International Research Fellow of the Japan Society for the Promotion of Science (JSPS), Institute of Mathematics for Industry, Kyushu University, Japan. He received the B.S. and M.S. degrees in Mathematics from Hubei University, Wuhan, P. R. China, in 2003 and 2006, respectively, and the Ph.D. degree in information and communication engineering from Huazhong University of Science and Technology, China, in December 2013. His research interests focus on network and information security.

Zhe Xia is an Associate Professor in the Department of Computer Science at Wuhan University of Technology (WHUT), China. He received the PhD degree from University of Surrey (UK) in 2009, supervised by Prof. Steve Schneider. He has worked as a Research Fellow in the Department of Computing at University of Surrey between 2009 and 2013 before joining WHUT. His research interests include secure e-voting protocols, secret sharing and secure multiparty computation. Dr. Xia serves as the Associate Editor for Journal of Information Security Applications (JISA). He also serves on the program committees for many international conferences, such as NSS, EVT, VOTE-ID, DCIT.

Jian Shen received the BE degree from Nanjing University of Information Science and Technology, Nanjing, China, in 2007 and the ME and PhD degrees in Computer Science from Chosun University, Gwangju, Korea, in 2009 and 2012, respectively. Since late 2012, he has been a full professor in the School of Computer and Software at Nanjing University of Information Science and technology, Nanjing, China. His research interests include information and network security, security systems, public-key cryptography, cloud computing and security, wireless networking and mobile computing.