PAID: Privacy-Preserving Incentive Mechanism Based on Truth Discovery for Mobile Crowdsensing

Tao Wan[1(✉)], Shixin Yue[1], and Weichuan Liao[2]

[1] School of Information Engineer, East China Jiaotong University, Nanchang 330013, China
[2] School of Science, East China Jiaotong University, Nanchang 330013, China

Abstract. Incentive mechanisms are an essential method to encourage users to participate in the mobile crowdsensing task. However, most incentive mechanisms based on data quality do not consider users' security and privacy protection. To overcome the above problems, we propose a privacy protection incentive mechanism based on truth discovery, named PAID. Specifically, we use the secure truth discovery scheme to calculate ground truth and the weight of users' data while protecting their privacy. Besides, to ensure the accuracy of the MCS results, a data eligibility assessment is proposed to remove unreliable user data before the truth discovery scheme. Finally, we distribute rewards to users based on their data quality. The analysis and evaluation demonstrate the security and effectiveness of our PAID.

Keywords: Incentive mechanism · Truth discovery · Mobile crowdsensing · Privacy-preserving

1 Introduction

As more and more sensors are integrated into human-carried mobile devices, such as GPS locators, gyroscopes, environmental sensors, and accelerometers, they can collect various types of data. Therefore, the MCS system [1,2] can utilize the sensors equipped in mobile devices to collect sensing data and complete various sensing tasks [3], such as navigation service, traffic monitoring, indoor positioning, and environmental monitoring. In general, the MCS system consists of three entities: a task requester, a sensing server, and participating users. The task requester publishes sensing tasks and pays awards for sensing results. The server recruits users according to the sensing task, processes the data from users, and sends the results to the task publisher. Users collect sensing data based on the requirements of the sensing task and get rewards.

In the practical MCS system, the sensing data collected by users are not always unreliable [4] due to various factors (such as poor sensor quality, lack of

J. Xiong et al. (Eds.): MobiMedia 2021, LNICST 394, pp. 264–277, 2021.
https://doi.org/10.1007/978-3-030-89814-4_20

effort, background noise). Therefore, the final result may be inaccurate if we treat the data provided by each user equally (e.g., averaging). To solve this problem, truth discovery [5,6] has been widely concerned by industry and academia. But one problem with these methods is that users have to be online to interact with the server. Therefore, if we design a truth discovery scheme that allows users to exit, the MCS system can get stronger robustness.

The proper functioning of the truth discovery requires enough users and high-quality sensing data. Generally, the MCS system utilizes an incentive mechanism [7] to motivate sufficient users to participate in sensing tasks. However, because of monetary incentives, malicious users attempt to earn rewards with little or no effort. Consequently, the evaluation of data quality is critical to the MCS system. To improve data quality, users who provide incorrect data can be removed before sensing data aggregated [8]. And the MCS system can output more accurate aggregation results.

Although the incentive mechanism has been improved a lot, users' privacy protection remains inadequate. When users submit sensing data, their sensitive or private information [9] may be leaked, including identity privacy, location privacy, and data privacy. And privacy disclosure [10] will reduce users' willingness to participate in sensing tasks. Recently, some researchers have designed the incentive mechanism scheme of privacy protection [11,12]. In [8], an incentive method is proposed to protect the user's identity and data privacy. Still, the user's sensing data will be submitted to the task publisher regardless of the privacy of the sensing data.

To address these issues, we propose a privacy-preserving incentive mechanism based on truth discovery, called PAID. In our PAID, the task publisher set data constraints, such as time, location, budget, and sensing data. If the user does not collect the sensing data at the required time and location or sensing data is not in the qualified range, we believe that the user's sensing data is not credible (i.e., unqualified). After removing the unqualified user's data, the qualified user's sensing data will be submitted to the server to calculate the ground truth and weight. We also design a secure truth discovery scheme, which can still work when some users drop out. Moreover, our truth discovery can ensure that other parties cannot obtain users' sensing data except users themselves. Finally, we calculate every user's data quality according to the weight and distribute the reward.

In summary, the main contributions of this paper are as follows:

- We propose a method to judge whether the data is in the qualified range. And this method will not disclose users' data and the qualified interval in the implementation process.
- We design a security truth discovery scheme, which can compute ground truth and users' weight. In this scheme, any party can not get the user's sensing data except himself. And the method can allow users to drop out.
- Our PAID accomplishes reward distribution according to data quality. The data quality is calculated by the weight.

2 Problem Statement

In this section, we introduce the background of truth discovery and our design goals.

2.1 Truth Discovery

Truth discovery [13] is widely used in the MCS system to solve the conflicts between sensing data collected from multiple sources. Although the methods of estimating weights and calculating ground truth are different, their general processes are similar. Specifically, truth discovery initializes a random ground truth and then iteratively updates the weight and ground truth until convergence.

Weight Update: Suppose that the ground truth of the object is fixed. If the user's sensing data is close to the ground truth, a higher weight should be assigned to the user. The weight w_i of each user u_i can be iteratively updated as follows:

$$w_i = log\left(\frac{\sum_{i'=1}^{|U|} d_{ist}(x_{i'}, x^*)}{d_{ist}(x_i, x^*)}\right) \tag{1}$$

where $d_{ist}(x_i, x^*)$ is a distance function, and $d_{ist}(x_i, x^*) = (x_i - x^*)^2$. We use U to represent the set of users, and $|U|$ is the number of users in the set U. The sensing data collected by the user u_i is denoted as x_i, which i is the number of u_i. And x^* is the estimated ground truth.

Truth Update: Similarly, we assume that the weight w_i of each user u_i is fixed. Then we can calculate the ground truth x^* as follows.

$$x^* = \frac{\sum_{i=1}^{|U|} w_i \cdot x_i}{\sum_{i=1}^{|U|} w_i} \tag{2}$$

The final ground truth x^* is obtained by iteratively running the weight update and the truth update when the convergence condition is satisfied.

2.2 Design Goals

In this section, we introduce the design goals of our PAID, which are divided into privacy and security goals and property goals.

The privacy goals can protect the user's private data, and the security goals can avoid malicious attacks. The details are as follows.

- **Privacy goals.** PAID can protect user's location privacy, data privacy, and identity privacy. Specifically, the location and sensing data of a user can not be obtained by any other parties except the user himself. And users' real identities would not be disclosed when performing a sensing task.

– **Security goals.** In our PAID, users can avoid the denial of payment attack (DoP) of TP. The server $\mathcal{S}$ cannot initiate an inference attack (IA) on users. The server $\mathcal{S}$ can resist the data pollution attack (DPA) launched by malicious users. And our PAID guarantees fairness by resisting the Sybil attack (SA).

Our PAID also requires the following property goals.

– **Eligibility.** If users' data do not meet the eligibility requirements, they cannot pass the eligibility assessment. In other words, the sensing data adopted by our PAID must be eligible.
– **Zero-knowledge.** When the server $\mathcal{S}$ assesses whether users' data meets the eligibility requirements, it cannot obtain the content of users' private data.

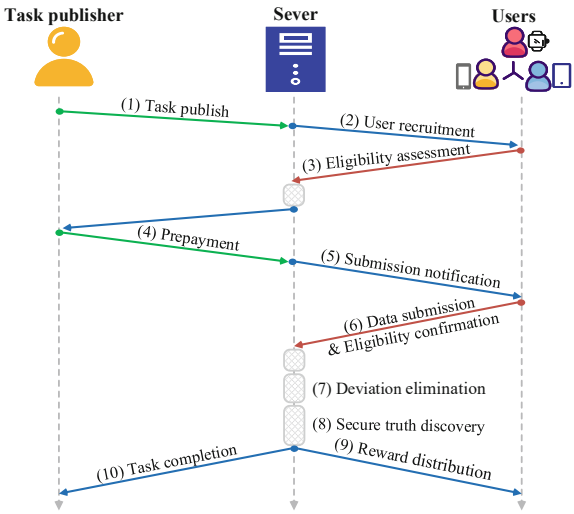

Fig. 1. System model of PAID.

3 Preliminaries

In this section, we review the cryptographic primitives used in our PAID.

3.1 Secret Sharing

We use Shamir's t-out-of-N secret sharing protocol, which can split each user's secret s into N shares, where any t shares can be used to reconstruct s. Still, it is impossible to get any information about s if the shares obtained by attackers are less than t.

Assume that some integers can be identified with distinct elements in a finite field $\mathcal{F}$, where $\mathcal{F}$ is parameterized with a size of $l > 2^k$ (which k is the security

parameter). These integers can represent all users' IDs, and we use a symbol U to denote the set of users' IDs. Then the Shamir's secret sharing protocol consists of two steps as below.

- **Shamir.share**$(s, t, U) \rightarrow \{(u_i, s_i)\}_{u_i \in U}$. The inputs of the sharing algorithm are a secret s, a threshold $t \leq |U|$, and a set U of N field elements denoting the users' ID, where $|U| = N$. It outputs a set of shares s_i, each of which is associated with its corresponding the user u_i.
- **Shamir.recon**$(\{(u_i, s_i)\}_{u_i \in \mathcal{M}}, t) \rightarrow s$. The inputs of the reconstruction algorithm are the shares corresponding to a subset $\mathcal{M} \subseteq U$ and a threshold t, where $t \leq |\mathcal{M}|$, and it outputs the secret s.

3.2 Key Agreement

We utilize Diffie-Hellman key agreement called SIGMA [14] in our PAID to generate a session key between two users. Typically, SIGMA is described in three parts as follows.

- **KA.param**$(k) \rightarrow (\mathbb{G}, g, q, H)$. The algorithm's input is a security parameter k. It samples a group $\mathbb{G}$ of prime order q, along with a generator g and a hash function H, where H is set as SHA-256 for practicability in our model.
- **KA.gen**$(\mathbb{G}, g, q, H) \rightarrow (x, g^x)$. The algorithm's inputs are a group $\mathbb{G}$ of prime order q, along with a generator g and a hash function H. It samples a random $x \leftarrow Z_q$ and g^x, where x and g^x will be marked as the secret key $\mathtt{SK}_i$ and the public key $\mathtt{PK}_i$ in the following sections.
- **KA.agree**$(sign_j(g^{x_i}, g^{x_j}), MAC_k(u_j), x_i, g^{x_j}, i, j) \rightarrow s_{i,j}$. The algorithm's inputs are the user u_i's secret key x_i, the user u_j's public key g^{x_j}, signed signature $sign_j(g^{x_i}, g^{x_j})$ and $MAC_{k_v}(u_j)$ from the user u_j, where k_v is used as the MAC key. It outputs a session key $s_{i,j}$ between user u_i and user u_j. For simplicity, we use **KA.agree**$(x_i, g^{x_j}) \rightarrow s_{i,j}$ to represent the above process in the following sections.

3.3 Paillier Cryptosystem

The Paillier Cryptosystem is a probabilistic public key Cryptosystem. It consists of three parts as follows.

- **Paillier.gen**$(N, g) \rightarrow (sk_p, pk_p)$. The key distribution algorithm inputs are a number N and $g \leftarrow Z^*_{N^2}$, where N is the product of two large primes p, q. It outputs a secret key sk_p and a public key pk_p, where pk_p is computed by (N, g), and $sk_p = lcm(p-1, q-1)$.
- **Paillier.enc**$(m, pk_p) \rightarrow c$. The encryption algorithm inputs are a plaintext m (which $m < N$) and a public key pk_p. It outputs a ciphertext c.
- **Paillier.dec**$(c, sk_p) \rightarrow m$. The decryption algorithm inputs are a ciphertext c (which $c < N^2$) and a secret key sk_p. It outputs a plaintext m.

The Paillier cryptosystem has the property of homomorphic addition.

$$E_{pk}(a + b) = E_{pk}(a) \cdot E_{pk}(b) \pmod{N^2}, \tag{3}$$

We assume that E is an encryption function.

4 Technical Intuition

In this section, we first introduce how the interval judgment scheme can judge users' data eligibility under protecting users' privacy. Then, we notice that truth discovery mainly involves the aggregation of multiple users' data in a secure manner. Therefore, we require that the server $\mathcal{S}$ only get the sum of users' input, not content. And we propose a double-masking scheme to achieve this goal. Finally, we introduce the process of secure truth discovery.

4.1 Interval Judgment Scheme for Privacy Protection

In our PAID, we use the interval judgment scheme [15] based on the Paillier cryptosystem to determine the sensing data eligibility. Every user u_i provides a sensing data x_i, and the server provides a continuous integer interval $[y_1, y_2]$ $(y_1, y_2 \leftarrow Z^*)$. The server $\mathcal{S}$ can judge whether the user u_i's sensing data x_i meets the interval range $[y_1, y_2]$ without knowing the data x_i. The user u_i also cannot obtain any information about the integer interval. The scheme is divided into four steps as follows.

- The user u_i gets $(pk_p, sk_p) \leftarrow \textbf{Paillier.gen}(N, g)$. Then u_i computes $E(x_i)$ using pk_p and sends it to $\mathcal{S}$.
- The server $\mathcal{S}$ picks two random numbers k, b $(k, b \leftarrow Z^*)$ to construct a monotone increasing (or decreasing) function $f(x_i) = kx_i + b$. Then the server $\mathcal{S}$ computes $f(y_1), f(y_2), c = E(x_i)^k E(b) = E(kx + b)$, and sends them to u_i.
- After receiving the information from the server $\mathcal{S}$, the user u_i gets $f(x_i) \leftarrow \textbf{Paillier.dec}(c, sk)$, then compares the size of $f(y_1)$, $f(y_2)$, and $f(x_i)$. Next, send the message to the server $\mathcal{S}$.
- After receiving the message from u_i, the server $\mathcal{S}$ judges whether $f(y_1) < f(x_d) < f(y_2)$. If so, we can know $x_i \in [y_1, y_2]$ because of the monotonicity of the function $f(x_i) = kx_i + b$, i.e., the user u_i passes the data eligibility assessment. Otherwise, the user u_i fails to pass the eligibility assessment of the server $\mathcal{S}$.

For simplicity, We formulate the above process as an interval judgment function denoted by $ins(x_i, y_1, y_2)$. If the user u_i passes the eligibility assessment of the server $\mathcal{S}$, $ins(x_i, y_1, y_2) = 1$, otherwise $ins(x_i, y_1, y_2) = 0$.

4.2 One-Masking to Protect Security

Assume that all users are represented in sequence as integers $1, \ldots, n$. And any pair of users (u_i, u_j), $i < j$ agree on a random value $r_{i,j}$. Let's add $r_{i,j}$ to the user u_i's data x_i and subtract $r_{i,j}$ from the user u_j's data x_j to mask all users' raw data. In other words, each user u_i computes as follows.

$$y_i = x_i + \sum_{u_j \in U: i<j} r_{i,j} - \sum_{u_j \in U: i>j} r_{j,i} \pmod{R}, \tag{4}$$

where we assume x_i and $\sum_{u_j \in U} r_{i,j}$ is in Z_R with order R for simplicity.

Then, each user u_i submits y_i to the server $\mathcal{S}$, and $\mathcal{S}$ computes:

$$\begin{aligned} z &= \sum_{u_i \in U} y_i \\ &= \sum_{u_i \in U} \left(x_i + \sum_{u_j \in U : i<j} r_{i,j} - \sum_{u_j \in U : i>j} r_{j,i} \right) \\ &= \sum_{u_i \in U} x_i \pmod{R}. \end{aligned} \tag{5}$$

However, this approach has two shortcomings. The first one is that every user u_i needs to exchange the value $r_{i,j}$ with all other users, which will result in quadratic communication overhead ($|U|^2$) if done naively. The second one is that the protocol will fail if any user u_i drops out since the server can't eliminate the value $r_{i,j}$ associated with u_i in the final aggregated results z.

4.3 Double-Masking to Protect Security

To solve these security problems, we introduce a double-masking scheme [16].

Every user u_i can get a session key $r_{i,j}$ with other user u_j by engaging the Diffie-Hellman key agreement after the server $\mathcal{S}$ broadcasting all of the Diffie-Hellman public keys.

We use the threshold secret sharing scheme to solve the issue that users are not allowed to drop out. Every user u_i can send his secret's shares to other users. Once some users cannot submit data in time, other users can recover masks associated with these users by submitting shares of these users' secrets to $\mathcal{S}$, as long as the number of dropped users is less than t (i.e., threshold of Shamir's secret sharing).

However, there is a problem that may lead to users' data leaked to $\mathcal{S}$. There is a scenario where a user u_i is very slow to send data to the server $\mathcal{S}$. The server $\mathcal{S}$ considers that the user u_i has dropped and asks for their shares of the user u_i's secret from all other users. Then, the server $\mathcal{S}$ receives the delayed data y_i after recovering u_i's mask. At this time, the server $\mathcal{S}$ can remove all the masks $r_{i,j}$ and get the plaintext x_i.

To improve the scheme, we introduce an additional random seed $\mathtt{n}_i$ to mask the data. Specifically, each user u_i selects a random seed $\mathtt{n}_i$ on the round of generating $r_{i,j}$, then creates and distributes shares of $\mathtt{n}_i$ to all other users during the secret sharing round. Now, users calculate y_i as follows:

$$\begin{aligned} y_i = & x_i + \mathbf{PRG}(\mathtt{n}_i) + \sum_{u_j \in U : i<j} \mathbf{PRG}(r_{i,j}) \\ & - \sum_{u_j \in U : i>j} \mathbf{PRG}(r_{j,i}) \pmod{R}. \end{aligned} \tag{6}$$

Note that an honest user will never reveal both kinds of shares of the same user to the server $\mathcal{S}$. During the recovery round, the server $\mathcal{S}$ can request either a share of $r_{i,j}$ or a share of $\mathtt{n}_i$ from each surviving user u_j. After gathering at least t shares of $r_{i,j}$ for all dropped users and t shares of $\mathtt{n}_i$ for all surviving users, the server $\mathcal{S}$ can eliminate the remaining masks to reveal the sum.

4.4 Secure Truth Discovery

In the secure truth discovery scheme [6], data exchange is between users and the server $\mathcal{S}$. The user u_i needs to collect sensing data x_i, perform the double-masking scheme to mask the raw input data, and then send the masked input data to $\mathcal{S}$. The server $\mathcal{S}$ receives masked input data from each user u_i and aggregates the input data of online users. The main process can be summarized as follows.

Part 1 (Key Generation). A trusted third party creates three key pairs for each user u_i signature, session key, and noise value. Then, each user u_i generates shares of $\mathtt{n}_i$ using secret sharing protocol and sends the encrypted information to $\mathcal{S}$.

Part 2 (Masking Data). Each user u_i uses the double-masking scheme to mask his input data and sends it to $\mathcal{S}$.

Part 3 (Unmasking). After receiving the masking data, the server $\mathcal{S}$ performs a summation operation to obtain the sensing data aggregation result of surviving users. For dropped users, the server $\mathcal{S}$ restores their noise using the secret sharing protocol then eliminates the impact on the aggregation results.

Part 4 (Computing Ground Truth and Weight). After the server $\mathcal{S}$ gets the aggregation result, the server $\mathcal{S}$ iteratively calculates the ground truth x^* and weight w_i of every user u_i according to Formula 1 and Formula 2 until convergence. And the server $\mathcal{S}$ initializes a random ground truth x^* in the first calculation.

5 Our Proposed Scheme

In this section, we introduce the process of our model. For convenience, we introduce a simple case. We set up a sensing task $\mathcal{T}$ to collect the temperature of urban roads in the evening. There are range requirements for time, location, and sensing data (i.e., temperature). To be more precise, the time range is required to be 5–8 pm on February 3rd, the location range is required to be 12.45–12.55 E and 41.79–41.99 N, and the temperature requirement is 10–15°C. In our PAID, we consider the range requirement as the data eligibility requirement $\mathcal{E}$. The data $\mathcal{D}_i$ ($\mathcal{D}_i = (x_i, \tau_i, \hat{\iota_i}, \tilde{\iota_i})$) collected by a user u_i meet the eligibility requirements $\mathcal{E}$, meaning that $10 \leq x_i \leq 15, 5 \leq \tau_i \leq 8, 12.45 \leq \hat{\iota_i} \leq 12.55, 41.79 \leq \tilde{\iota_i} \leq 41.99$. Figure 1 shows the flow of our PAID. And the specific steps are as follows.

Step 1 (Task Publish). The task publisher TP initializes a public key $pk_{\mathcal{T}}$ and a private key $sk_{\mathcal{T}}$, a reward control parameter π (π is a decimal number), a task budget B, the number of users N, and eligibility requirements $\mathcal{E}$ for a sensing task $\mathcal{T}$. The public key $pk_{\mathcal{T}}$ is used to encrypt the information that the server $\mathcal{S}$ needs to send to the TP, and the TP decrypts the ciphertext using the private key $sk_{\mathcal{T}}$. Then the TP sends the information $\{\mathcal{T}, pk_{\mathcal{T}}, \pi, N, B, \mathcal{E}\}$ to $\mathcal{S}$ as a task request.

Step 2 (User Recruitment). The server $\mathcal{S}$ broadcasts the sensing task information $\{\mathcal{T}, \pi, N, B\}$ and recruits N users who request to participate in the sensing task. Then $\mathcal{S}$ generates a key pair $\{\mathtt{PK}_{\mathcal{S}}^{i}, \mathtt{SK}_{\mathcal{S}}^{i}\}$ using the key agreement scheme for every user u_i and sends $\mathtt{PK}_{\mathcal{S}}^{i}$ to u_i.

Step 3 (Eligibility Assessment). Each user u_i confirms whether $c_i \leq \frac{B-\pi}{N}$, where c_i denotes the sensing cost of u_i, and the posted lowest reward is denoted as $\frac{B-\pi}{N}$. If $c_i \leq \frac{B-\pi}{N}$, u_i starts the sensing task and collects the data $\mathcal{D}_i$. The user u_i then generates a key pair $\{\mathtt{PK}_i, \mathtt{SK}_i\}$ using the key agreement scheme and computes a session key $k_i \leftarrow \mathbf{KA.agree}(\mathtt{SK}_i, \mathtt{PK}_{\mathcal{S}}^{i})$ as u_i's anonymous identity information. Then the user u_i performs the interval judgment scheme $ins(\mathcal{D}_i, \mathcal{E})$ and sends the public key $\mathtt{PK}_i$ to $\mathcal{S}$. Specifically, $ins(\mathcal{D}_i, \mathcal{E})$ is divided into $ins(x_i, \mathcal{E})$, $ins(\tau_i, \mathcal{E})$, $ins(\hat{\iota}_i, \mathcal{E})$, $ins(\tilde{\iota}_i, \mathcal{E})$.

Step 4 (Prepayment). After recruiting N eligible users, the server $\mathcal{S}$ requests TP to prepay a budget reward B for the sensing task $\mathcal{T}$ to prevent the denial of payment attack. And the server $\mathcal{S}$ calculates the session key $k_i \leftarrow \mathbf{KA.agree}(\mathtt{SK}_{\mathcal{S}}^{i}, \mathtt{PK}_i)$ with the eligible user u_i.

Step 5 (Submission Notification). After getting the budget reward B, the server $\mathcal{S}$ informs the eligible user u_i ($1 \leq i \leq N$) to submit data.

Step 6 (Data Submission & Eligibility Confirmation). After receiving the submission notification, each user u_i performs double masking scheme to mask the sensing data x_i and get y_i, at the same time, execute eligibility confirmation $ins(\mathcal{D}_i, \mathcal{E})$ to prevent malicious users from modifying data. Then u_i encrypts the data y_i using the symmetric encryption algorithm and sends the ciphertext $\mathbf{SEnc}(y_i, k_i)$ to $\mathcal{S}$. The session key k_i is the key of symmetric encryption.

Step 7 (Secure Truth Discovery). The server $\mathcal{S}$ computes the surviving user u_i's weight w_i and the ground truth x^* of the sensing object utilizing the truth discovery algorithm. The detailed algorithm process will be introduced later.

Step 8 (Reward Distribution). The server $\mathcal{S}$ calculates the sensing data quality $q_i = \frac{w_i}{\sum_{i=1}^{m} w_i}$ of u_i, where $\sum_{i=1}^{m} q_i = 1$, m is the number of online users. Then $\mathcal{S}$ pays a monetary reward $p_i = \frac{B}{m} + \pi \cdot (q_i - \bar{q})$ for u_i, where $\pi \cdot (q_i - \bar{q})$ denotes the payment parameter, $m \leq N$, and $1 \leq i \leq m$.

Step 9 (Task Completion). The server $\mathcal{S}$ encrypts the ground truth x^* using $pk_{\mathcal{T}}$ and sends $\mathbf{Enc}(x^*, pk_{\mathcal{T}})$ to TP. And the TP can decrypt the data using $sk_{\mathcal{T}}$, i.e., $x^* = \mathbf{Dec}(\mathbf{Enc}(x^*, pk_{\mathcal{T}}), sk_{\mathcal{T}})$.

6 Analysis

In this section, we introduce property analysis, privacy analysis, and security analysis to illustrate the feasibility of our PAID.

6.1 Property Analysis

In this section, we introduce eligibility, zero-knowledge of our PAID.

Theorem 1 (Eligibility). If the data $\mathcal{D}_i$ ($\mathcal{D}_i = (x_i, \tau_i, \hat{\iota_i}, \tilde{\iota_i})$) collected by users do not meet the eligibility requirement $\mathcal{E}$, these users cannot pass the eligibility assessment.

Proof. We assume that the user's data are denoted as s, and the eligibility requirement interval is $[a, b]$. The user gets ciphertext $E(s)$ using homomorphic encryption. Then $\mathcal{S}$ picks different random k, b, and constructs a monotone increasing (or decreasing) function $f(x) = kx + b$. Then $\mathcal{S}$ computes $f(a)$, $f(b)$, and $c = E(s)^k E(b) = E(ks + b)$. When receiving $f(a)$, $f(b)$, c from $\mathcal{S}$, the user decrypts c to get $f(s)$ and compare the sizes of $f(a)$, $f(b)$, $f(s)$. Because the user does not know the monotonicity of the function, it is impossible to determine the size relationship among the three numbers. Therefore, if the user's data is not qualified, then it cannot pass the qualification judgment.

Theorem 2 (Zero-knowledge). The server $\mathcal{S}$ can determine whether the user's data meets the eligibility requirements, but it cannot know the user's specific data content.

Proof. Similar to the description in Theorem 1, we assume that the user's data is s, and the server $\mathcal{S}$ can receive the user's homomorphic encrypted ciphertext $E(s)$. Since the Paillier Cryptosystem is indistinguishable under the chosen plaintext attack, a malicious user has no way to recover the plaintext s. The server $\mathcal{S}$ may be curious about each user's data, but it cannot obtain each user's data s without knowing the secret key.

6.2 Privacy Analysis

In this section, we demonstrate the protection of user sensing data, location, and identity privacy in our PAID.

Theorem 3 (Data and location privacy protection). In addition to the user himself, other parties cannot obtain the user's sensing data and location data.

Proof. In PAID, the objects that steal users' data and location privacy are mainly the server $\mathcal{S}$ and external attackers. Specifically, the server $\mathcal{S}$ may obtain users' sensing data and location privacy in eligibility assessment and truth discovery. External attackers steal data and location privacy by eavesdropping on the communication between the server $\mathcal{S}$ and users.

According to Theorem 2, we can know that our PAID is zero-knowledge, so the server $\mathcal{S}$ cannot learn users' sensing data and location data in the eligibility assessment. In truth discovery, users' sensing data is sent to $\mathcal{S}$ after double-masking. However, the server $\mathcal{S}$ can't recover users' raw sensing data by double-masking sensing data. Furthermore, before the communication between the user u_i and $\mathcal{S}$, the data is encrypted by AES symmetric encryption function $\mathbf{SEnc}(y_i, k_i)$. Therefore, as long as $\mathbf{SEnc}(y_i, k_i)$ is secure, external attackers cannot steal the data y_i by eavesdropping communication.

Theorem 4 (Identity privacy protection). When users participate in a sensing task, they use an anonymous identity rather than their real identity. Therefore, any PPT adversary cannot distinguish the users' identities.

Proof. In PAID, the anonymous identity of a user u_i is represented by $k_i \leftarrow \mathbf{KA.agree}(\mathtt{SK}_i, \mathtt{PK}_{\mathcal{S}}^i)$, and the real identity of u_i is $\mathtt{SK}_i$ where $\mathtt{SK}_i = x_i \leftarrow Z_q$, and $\mathtt{PK}_{\mathcal{S}}^i = g^{x_{\mathcal{S}}^i}$ ($\mathtt{PK}_{\mathcal{S}}^i$ is a token assigned by $\mathcal{S}$). The user u_i uses an anonymous identity k_i rather than a real identity $\mathtt{SK}_i$ to participate in a sensing task. Because of the DDH problem, the PPT adversary cannot get the real identity $\mathtt{SK}_i$ of the user u_i by the anonymous identity k_i. We omit the detailed proof, and interested readers can learn more details in the literature [14].

6.3 Security Analysis

In this section, we describe the attacks our PAID can resist, including *Denial of Payment attack* (DoP), *Inference attack* (IA), *Data pollution attack* (DPA), and *Sybil attack* (SA).

(1) *Resistance to denial of payment attack* (DoP). We use the prepayment mechanism in our PAID. At the beginning of a sensing task, the task publisher TP pays the monetary rewards of users to $\mathcal{S}$ in advance. If a malicious TP refuses to pay the monetary reward after receiving the data, $\mathcal{S}$ can pay the reward to users according to the reward distribution formula. Therefore, the TP cannot refuse to pay users the reward.
(2) *Resistance to inference attack* (IA). The server $\mathcal{S}$ cannot initiate an inference attack against users' data since our PAID is zero-knowledge.
(3) *Resistance to Data pollution attack* (DPA). Our PAID introduces eligibility assessment, and the unqualified data submitted by users are not used in the truth discovery algorithm. Therefore, our PAID can resist the Data pollution attack (DPA).
(4) *Resistance to Sybil attack* (SA). The anonymous identity k_i of a user u_i needs the information $\mathtt{PK}_i$ provided by the user and the token $\mathtt{PK}_{\mathcal{S}}^i$ assigned by $\mathcal{S}$. Each user can only obtain one token from $\mathcal{S}$, then get the anonymous identity k_i using the key agreement algorithm. Hence, untrusted users cannot forge vast fake identities to launch the Sybil attack (SA).

Table 1. Performance comparison between PAID and related work

Protocol	Computational overhead	Communication overhead
PAID	$4M_{N^2}$	3
Protocol 2 in [17]	$8M_{N^2}$	6

7 Performance Evaluation

In this section, we analyze the computational and communication overhead in the eligibility assessment. And Table 1 shows the performance comparison between our PAID and related work.

7.1 Computational Overhead

Since we use the Paillier homomorphic encryption in eligibility assessment, we use the modular exponentiation as the computational overhead indicator and ignore other operations. For convenience, the modular exponentiation in Paillier homomorphic encryption is denoted as M_{N^2}. The server $\mathcal{S}$ requires two encryptions, and users perform one encryption and one decryption. Therefore, the computational overhead of the eligibility assessment is $4M_{N^2}$.

7.2 Communication Overhead

Typically, we measure the communication overhead by communication rounds in secure multiparty computation. In our eligibility assessment, the interaction between server and user is 3 rounds.

8 Related Work

Truth discovery is an effective technology that can calculate the ground truth and users' quality from conflicting sensing data. Li et al. [13] Proposed a general truth discovery scheme, but privacy protection is not in their work scope. To protect users' privacy data, Miao et al. [18] proposed the first privacy-preserving truth discovery scheme using the Paillier cryptosystem, but the computational and communication costs are huge. Later, some works [19] improve the communication cost and privacy protection of truth discovery. However, these works do not take into account the failure of the MCS system caused by users' exit. And most existing works do not combine the incentive mechanism.

Another previous work related to this paper is the incentive mechanism in the MCS system. Some works [20] utilize the game theory model, such as the auction model, to implement incentive mechanisms but do not consider users' privacy leakage. In [12], the author designs privacy protection in the incentive mechanism. However, these works do not include the assessment of users who provide unqualified data in advance. Zhao et al. [8] presented an incentive mechanism

model to evaluate the reliability of users' data while protecting data privacy. However, the users' sensing data needs to be submitted to the task publisher, so the sensing data privacy protection is still insufficient.

Different from existing work, we design an incentive mechanism based on truth discovery, which can remove unqualified users in advance. The incentive mechanism ensures that enough users participate in the sensing task and improve truth discovery accuracy.

9 Conclusion

In this paper, we propose a privacy-preserving incentive mechanism based on truth discovery in the MCS system. Specifically, we design an eligibility assessment scheme to estimate whether the data submitted by users are qualified. Next, the truth discovery scheme calculates the ground truth and the weight of each user using these qualified sensing data. Then we quantify the data quality of users by weight and distribute the rewards. Besides, we also demonstrate that PAID meets eligibility, zero-knowledge. And the analysis shows that our PAID can resist the Denial of Payment attack, Inference attack, Data pollution attack, and Sybil attack. In future work, we will demonstrate the efficiency of our model through experiments.

Acknowledgments. This work was supported by the National Nature Science Foundation of China (No. 61962022 and No. 62062034), the Key Research and Development Plan of Jiangxi Province (No. 20192BBE50077), and the Postgraduate Innovation Special Fund Project of Jiangxi Province (No. YC2020-S365).

References

1. Xiong, J., Zhao, M., Bhuiyan, M.Z.A., Chen, L., Tian, Y.: An AI-enabled three-party game framework for guaranteed data privacy in mobile edge crowdsensing of IoT. IEEE Trans. Industr. Inf. **17**(2), 922–933 (2019)
2. Wei, X., Sun, B., Cui, J.: Task replica assignment in mobile self-organized crowdsensing. Int. J. Performability Eng. **16**(1), 152–162 (2020)
3. Xiong, J., Chen, X., Yang, Q., Chen, L., Yao, Z.: A task-oriented user selection incentive mechanism in edge-aided mobile crowdsensing. IEEE Trans. Netw. Sci. Eng. **7**(4), 2347–2360 (2020)
4. Zhang, S., Li, H., Dai, Y., Li, J., He, M., Lu, R.: Verifiable outsourcing computation for matrix multiplication with improved efficiency and applicability. IEEE Internet Things J. **5**(6), 5076–5088 (2018)
5. Ouyang, R.W., Srivastava, M., Toniolo, A., Norman, T.J.: Truth discovery in crowdsourced detection of spatial events. IEEE Trans. Knowl. Data Eng. **28**(4), 1047–1060 (2015)
6. Xu, G., Li, H., Liu, S., Wen, M., Lu, R.: Efficient and privacy-preserving truth discovery in mobile crowd sensing systems. IEEE Trans. Veh. Technol. **68**(4), 3854–3865 (2019)

7. Jin, H., Su, L., Chen, D., Nahrstedt, K., Xu, J.: Quality of information aware incentive mechanisms for mobile crowd sensing systems. In: Proceedings of the 16th ACM International Symposium on Mobile Ad Hoc Networking and Computing, pp. 167–176 (2015)
8. Zhao, B., Tang, S., Liu, X., Zhang, X.: PACE: privacy-preserving and quality-aware incentive mechanism for mobile crowdsensing. IEEE Trans. Mob. Comput. **20**(5), 1924–1939 (2020)
9. Dharminder, D., Mishra, D.: LCPPA: lattice-based conditional privacy preserving authentication in vehicular communication. Trans. Emerg. Telecommun. Technol. **31**(2), e3810 (2020)
10. Xiong, J., et al.: A personalized privacy protection framework for mobile crowdsensing in IIoT. IEEE Trans. Industr. Inf. **16**(6), 4231–4241 (2019)
11. Zhao, B., Tang, S., Liu, X., Zhang, X., Chen, W.N.: IronM: privacy-preserving reliability estimation of heterogeneous data for mobile crowdsensing. IEEE Internet Things J. **7**(6), 5159–5170 (2020)
12. Wang, Z., Li, J., Hu, J., Ren, J., Li, Z., Li, Y.: Towards privacy-preserving incentive for mobile crowdsensing under an untrusted platform. In: IEEE INFOCOM 2019-IEEE Conference on Computer Communications, pp. 2053–2061. IEEE (2019)
13. Li, Q., Li, Y., Gao, J., Zhao, B., Fan, W., Han, J.: Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation. In: Proceedings of the 2014 ACM SIGMOD International Conference on Management of Data, pp. 1187–1198 (2014)
14. Krawczyk, H.: SIGMA: the 'SIGn-and-MAc' approach to authenticated Diffie-Hellman and its use in the IKE protocols. In: Boneh, D. (ed.) CRYPTO 2003. LNCS, vol. 2729, pp. 400–425. Springer, Heidelberg (2003). https://doi.org/10.1007/978-3-540-45146-4_24
15. Chen, Z., Li, S., Chen, L., Huang, Q., Zhang, W.: Fully privacy-preserving determination of point-range relationship. SCIENTIA SINICA Informationis **48**(2), 187–204 (2018)
16. Bonawitz, K., et al.: Practical secure aggregation for privacy-preserving machine learning. In: proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security, pp. 1175–1191 (2017)
17. Guo, Y., Zhou, S., Dou, J., Li, S., Wang, D.: Efficient privacy-preserving interval computation and its applications. Chin. J. Comput. **40**(39), 1–17 (2016)
18. Miao, C., et al.: Cloud-enabled privacy-preserving truth discovery in crowd sensing systems. In: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, pp. 183–196 (2015)
19. Zhang, C., Zhu, L., Xu, C., Liu, X., Sharif, K.: Reliable and privacy-preserving truth discovery for mobile crowdsensing systems. IEEE Trans. Dependable Secure Comput. (2019)
20. Zhang, X., Yang, Z., Zhou, Z., Cai, H., Chen, L., Li, X.: Free market of crowdsourcing: incentive mechanism design for mobile sensing. IEEE Trans. Parallel Distrib. Syst. **25**(12), 3190–3200 (2014)