Future Generation Computer Systems 163 (2025) 107536

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

PPAT: An effective scheme ensuring privacy-preserving, accuracy, and trust for worker selection in mobile crowdsensing networks

Qianxue Guo [a], Yasha He [b], Qian Li [c], Anfeng Liu [a,*], Neal N. Xiong [d], Qian He [e], Qiang Yang [a], Shaobo Zhang [f]

[a] School of Computer Science and Engineering, Central South University, Changsha 410083 PR China
[b] School of Life Sciences, Central South University, Changsha 410083, PR China
[c] School of Automation, Central South University, Changsha 410083, PR China
[d] Department of Computer Science and Mathematics, Sul Ross State University, Alpine, TX 79830, USA
[e] Guilin University of Electronic Technology, Guilin 541199, PR China
[f] School of Computer Science and Engineering, Hunan University of Science and Technology, Xiangtan 411201, PR China

ARTICLE INFO

Keywords:
Mobile crowd sensing (mcs)
Data content privacy-preserving
Data accuracy
Trust computing
Worker selection

ABSTRACT

The data content privacy protection and data accuracy are two important research issues in Mobile Crowdsensing (MCS). However, current researches have rarely been able to satisfy privacy protection as well as data accuracy at the same time, thus hindering the development of MCS. To solve the above issues, for the first time, we have proposed a Privacy Preserving, Accuracy, and Trust data collection scheme (PPAT) for MCS, which can protect the privacy of data content and maintain high accuracy at low-cost style. In PPAT scheme, First, we proposed a scrambled data privacy protection framework which can protect the data of each worker from being known to any third party, which can protect the data privacy of workers. The second, more importantly, we propose a truth value estimation method based on trust computing, which can obtain the truth value more accurately compared to the classic methods under privacy-preserving. In the proposed trust-based truth value calculation, the worker's trust is determined by comparing it with the weight of the trusted worker. Then, the truth value is calculated by the trust of the workers, so that the truth value obtained is more accurate. Through theoretical analysis, it is proved that the proposed PPAT scheme has good worker data content, worker trust, and truth value content privacy protection. Through a large number of simulation experiments, the strategy proposed in this paper has a good ability to protect data content privacy compared to the previous strategy, while improving data quality by 0.5%~5.7%, and reducing data collection costs by 35.6%~54.9%.

1. Introduction

Mobile Crowd Sensing (MCS) [1-3] has emerged as a popular and widely adopted data collection paradigm in many applications. Large numbers of people with sensing devices are called workers [4-6], they can move to the nearest location specified by the task as needed. Therefore, it is a low-cost way to obtain data in a large area [7,8]. After obtaining a large amount of data, based on the fusion and processing of these data, various applications or services are formed [7,8]. Nowadays these applications range from numerous fields [9-11].

In MCS, there are three main components [12-14]: (1) The workers, (2) the Data Requesters (DR), (3) and the platform. If a worker is interested in participating in data collection, he or she will apply for a task [15,16], perform the task after being approved, and then report the data to the platform [20,21]. Since workers' sensory data consumes certain resources, time, etc. [5,11], MCS will give workers a certain payment based on their data quality [22-23]. DR is the demander of data. It publishes the scope, time, data attributes and other requirements of data collection tasks through the platform [12,13]. After collecting the data, it fuses the data and processes it to form various applications for consumers to use [17,18], and collects certain fees from consumers [22]. The fees cover the costs it pays its workers to collect the data, as well as the costs of processing the data [23]. The platform serves as an intermediary between workers and DR, it receives data collection tasks

* Corresponding author.
E-mail addresses: 8305210706@csu.edu.cn (Q. Guo), 8305211210@csu.edu.cn (Y. He), 8214210206@csu.edu.cn (Q. Li), afengliu@mail.csu.edu.cn (A. Liu), xiongnaixue@gmail.com (N.N. Xiong), heqian@guet.edu.cn (Q. He), yangqiang030@mail.csu.edu.cn (Q. Yang), shaobozhang@hnust.edu.cn (S. Zhang).

https://doi.org/10.1016/j.future.2024.107536
Received 12 April 2024; Received in revised form 16 July 2024; Accepted 15 September 2024
Available online 27 September 2024

from DR, publishes tasks to the outside world, and then transfers the collected data to DR to assist in completing data collection. Many studies have shown that high-quality data constructs high-quality services [22,23], so they can better meet the needs of consumers, and they are willing to pay higher fees for the platform, thus promoting the development of MCS [24]. On the contrary, low services constructed with low-quality data or even malicious data will be of poor quality and may even cause serious losses to consumers [25,26]. So, one of the important functions for the platform is to recruit high-quality workers to perceive data to construct high-quality services [21-26].

Privacy protection is also one of the most important issues in MCS which has been widely studied by researchers [27-29]. Data content privacy protection is one of them [27]. The main purpose of data content privacy protection is to protect the data sensed by workers from being known by unrelated third parties [19,27], which is especially important for the development of MCS. In MCS, data generally involves workers' privacy [27-29], for example, workers' physiological data such as height, weight data, workers' economic data, such as wages, consumption expenditure data, etc. which are personal privacy [30-32]. If the privacy of these data cannot be guaranteed, workers will be unwilling to participate in task completion [21,31]. Even public data such as environmental monitoring data will involve public data privacy issues. Therefore, the content of these data cannot be made irrelevant the third party knows [19,27]. If workers do not participate in data collection, MCS will become water without a source, which will hinder the development of MCS [31,32]. Therefore, data content privacy is an important issue that needs to be studied to ensure workers actively participate in data collection [19,27].

There have been some studies on data content privacy protection [19,27]. Since MCS data usually needs to be processed through a platform, data content privacy protection has two levels of meaning: one is that workers' data kept from platforms, which is basic research [19,27]. Because in most studies, researchers assume that the platform is honest but curious [27-33], which means that the platform will honestly carry out the task publication and some data process, but will also be curious about the sensitive information of the workers [27-33]. The other type is not only not letting the platform know the worker's data, but also not letting the DR know the worker's data, which is more challenging [27]. The main idea of these studies is: DR sends tasks to the platform, then the platform releases the tasks, workers apply for tasks to the platform and then execute the tasks, and then the workers send the sensed data to the platform. Since the platform sends the data to the DR after preliminary processing [27], workers generally do not send data directly to the DR, because in this case, the worker's data will be leaked to the DR [27].

In order to achieve the above goals, researchers have proposed some methods. However, the shortcomings of these methods are: in some applications, the data must be associated with some personal information of workers, especially in medical research, human health data must be associated with age, weight, gender, and even place of birth, education, and living habits for data analysis and mining. Therefore, when so much information is combined, the probability of identifying a specific worker through query and information reasoning is very high, which can easily cause information leakage [36,37].

Data quality is another important research issue in MCS [38]. There has been some research on quality-ware data collection [22,23,38]. However, the shortcomings of these type of methods are: First, the quality of the data obtained cannot be guaranteed. Since the credibility of the workers is not identified when recruiting workers, the platform does not know the quality of the workers recruited. Therefore, the accuracy of the estimated true data is uncertain. Second, the cost of this type of method is high, and the same task requires k times cost. There are some data quality-ware studies later: high-quality workers submit high-quality data, so when recruiting workers, we should try to select high-quality workers. In this way, the problem of high-quality data collection is transformed into the problem of how to identify high-quality workers and thus select high-quality workers. The general method used to identify high-quality workers is: if a worker always submits high-quality data, it is a high-quality worker [39]. In specific research, it is generally through the historical submission data quality of workers to identify who is high-quality workers.

In MCS data collection, although there are some studies on data content privacy protection [19,25,27-32], there are also some data quality-ware studies [22,23,38]. However, there are few studies that have both data content privacy protection functions and data quality-ware. However, in actual MCS, there is an urgent need for research that has both privacy protection and data quality-ware which is very challenging. To solve the above issue, for the first time, we proposed a Privacy Preserving, Accuracy, and Trust data collection scheme (PPAT) for MCS, which can protect the privacy of data content and maintain high accuracy at low-cost style. The contributions of the proposed PPAT scheme are summarized as follows:

(1) First, we propose a content privacy protection framework based on scrambling which can protect each worker's data from being known by the platform, DR, and any third party, thus having a strong function of protecting workers' data privacy.
(2) The second, more importantly, based on the proposed content privacy protection framework, we propose a method that can make inferences based on the comparison of the weights of trustworthy workers and the weights of other workers without knowing the workers' data. Worker's weight is an indicator that identifies how credible the worker is. Based on the identification of workers' trustworthiness, we propose a selection method for recruiting workers based on workers' trustworthiness, this realizes quality-ware data collection based on data content privacy protection. To the best of our knowledge, we are the first scheme which achieves quality-ware based data collection without the platform and DR knowing the privacy protection of data content.
(3) Through theoretical analysis, it is proved that the proposed PPAT scheme can not only protect workers' data from being known by any third party, but also significantly improve the quality of collected data, which is impossible with all previous strategies. This paper confirms the effectiveness of the proposed PPAT scheme through a large number of experiments. The PPAT strategy proposed in this paper provides better data content privacy protection than previous strategies while improving data quality by 0.5%∼5.7% and reducing data collection cost by 35.6%∼54.9%.

The rest of this paper is organized as follows. In Section 2, the related works are reviewed. And the system model and problem statement are presented in Section 3. In Section 4, the PPAT scheme is proposed. Then, Section 5 provides the performance analysis. Conclusions and future work are given in Section 6.

2. Related work

With the development of micro processing technology [41], edge networks have been greatly developed [42]. The research on MCS has also received extensive attention from researchers [2-7]. Strong privacy protection and high-quality data collection are two important goals pursued by the MCS system. Due to strong privacy protection, all participants will actively participate without worrying about privacy leak [19,25,27-32], only in this way can the vitality of MCS be maintained, and only high-quality data can be used to create high-quality data-based services which can bring benefits to all parties and promote the healthy development of MCS [34-37]. This section discusses some of the work in high-quality data collection and privacy protection that is relevant to the research in this paper.

2.1. The quality-based data collection related work

High-quality data collection is very important for MCS systems [22,

23,38]. Due to the fact that high-quality data is the basis for constructing high-quality applications or services, if the underlying data for constructing applications is of low quality, or even false [42], the quality of the constructed applications will not be high [23]. And low-quality or even incorrect applications can result in poor user experience for the consumers, and can even cause losses [40]. In this way, consumers will be lost and MCS will be difficult to develop.

If the data reported by workers is closer to the value of GTD, the accuracy is higher [33]. Therefore, the goal of truth data discovery (TDD) is to make the data received by DR closer to GTD, that is, the more accurate it is [33]. A commonly used method is to recruit k workers for the same task [33], collect k data, and then calculate the average, median, and weighted average of the data. Among them, the weighted average method is the most commonly used CRH [33]. The specific execution process of the CRH method is as follows: after DR receives k pieces of data, it can first consider the truth value $\overline{x}$ as the average of the k pieces of data, then calculate the difference $(x_i - \overline{x})^2$ between each x_i and the truth value $\overline{x}$, and calculate the weight of each worker $w_i = \log\frac{\sum_{i=1}^{k}(x_i - \overline{x})^2}{(x_i - \overline{x})^2}$, update to get a new truth value $\overline{x} = \frac{\sum_{i=1}^{k}(w_i x_i)}{\sum_{i=1}^{k} w_i}$. Repeat the above process, recalculate the difference between data x_i and the true value $\overline{x}$, update worker's weight, update true value. Until the estimated true value obtained in the next iteration is the same as the estimated true value obtained in the previous iteration, the final estimated truth value $\overline{x}$ obtained [33].

However, the accuracy of the ground truth obtained above is higher when most workers in the network are trustworthy workers [33]. Its accuracy decreases as the proportion of untrustworthy workers in the network increases. The biggest disadvantage of this method is that it is vulnerable to collusion attacks by malicious workers [40]. If multiple workers join together, it is easy to form a situation where malicious workers account for the majority [40]. At this time, this makes it easy for attackers to get any data they want from the platform, which can cause great damage to MCS.

The reason why data quality is difficult to guarantee in the above method is because MCS does not identify workers when recruiting workers, which leads to the possibility that the selected workers may be malicious workers, which will result in low data quality. Therefore, selection methods based on workers' quality (or trust) have been proposed by researchers [22,23,38]. In this type of method, the MCS system identifies the quality or trust of workers, and then selects those high-quality workers for data collection [12,22]. This can improve the quality of the data received. This type of method usually includes two important steps: the first step is to identify the trust or quality of workers; the second step is to select high-quality workers for data collection. For the first step, the method used to identify the quality of workers is usually based on the data quality of workers' historical reports [12,22]. It is generally believed that workers who report high-quality data are often perceived to be of high-quality or to have high levels of trust. The more recent the worker time is, the higher the weight will be in the evaluation, and the longer the data is from the current time, the lower the weight will be in the evaluation. After evaluating the quality of workers, there is also a lot of research on how to select workers. One type of worker selection strategy called Multi-Armed Bandit (MAB) is one of the more classic methods [12,22]. In this method, the main strategy of worker selection is composed of two parts: exploration and exploitation. Exploitation means that the system selects high-quality workers based on the obtained data quality, thus ensuring that the selected workers are of high quality [12,22]. But the problem if you only use it is that the proportion of workers that the system can identify quality is small compared to the entire worker set, so even if you select high-quality workers from them, the high-quality workers obtained will only be locally optimized. While the quality of more workers is unknown, there may be higher-quality workers among these workers with unknown data quality. Therefore, the MAB strategy uses the exploration method to select a certain proportion of workers from unknown quality workers [12,22]. If the selected workers are of higher quality, it will increase the high-quality set that the system can select in the future, thus systematically improving the system's data quality [12,22]. If the quality of the workers selected this time is low, it will only affect the quality of the current worker selection.

However, most of the above strategies based on the selection of data quality workers are anchored on the assumption that after the system receives the worker's data, it knows the quality of the data and can therefore evaluate the quality of the worker [22]. But there is a problem in actual MCS known as Information Elicitation Without Verification (IEWV) [43], which is much more challenging to solve due to the lack of GTD, so the system is still unable to evaluate the quality of workers' data even after receiving the worker's data. Based on this situation, most of the currently proposed worker selection strategies actually become inapplicable. The reason for this difficulty is that the system lacks GTD to determine whether the data reported by workers is true. Therefore, some researchers have proposed using drones to collect some data as GTD to compare the data submitted by workers [20], so that the system can evaluate the quality of workers' data. However, this type of method has high costs and is not applicable in some applications [20].

What makes things even more challenging is that none of the above methods take into account the privacy protection of the data [19,25,27-32]. In the case of data privacy protection, the DR and the platform cannot be made aware of the data content submitted by the workers, making it more difficult to identify the quality of the workers [19,27]. As mentioned before, it is a challenging issue for DR and the platform to evaluate and select worker trust when they can fully obtain worker data [12,22,43]. As mentioned before, it is a challenging issue for DR and the platform to evaluate and select worker trust when they can fully obtain worker data. In the case of privacy protection, it is more difficult for DR and the platform to evaluate workers' trust when they cannot obtain worker data. As far as we know from the current research, the work of this paper is the first time to conduct research on worker selection based on data quality under the condition of privacy protection of data content, which can not only protect the data privacy of workers but also improve the data quality, which is of great significance.

2.2. *The data privacy-preserving related work*

Data privacy protection method is an important research issue in MCS [19,25,27-32]. The goal of data privacy protection is discussed differently in different studies. In most studies, the goal of data privacy protection is to protect the data content reported by workers from being leaked to the platform [19,25,27-32]. Because, the platform is usually considered cautious and curious [3-5], as a result, although the platform will honestly carry out the intended data collection behavior. However, if it stores workers' data on it, it may cause serious data leakage once it is breached [36]. Of course, furthermore, it is best not to let the DR know the data of the workers, because a lot of the data of the workers involves personal privacy, or the workers' private information can be inferred based on the workers' data, so it is best not to let the data of the workers be known to the DR [19,27]. However, if a task only recruits one worker to collect, in this kind of data collection application, the DR must know the worker data. In this 1:1 data collection model, the mechanism for protecting the privacy of data content is very complicated [36,37]. However, in this model, the anonymity method can play a certain role, because the anonymity method anonymizes the personal information of workers [36,37], so it is difficult for DR to establish a connection between the data and the workers, thus achieving a certain privacy protection function. Most of the current MCS uses a 1: k data collection model such as the CRH method [33]. In this method, since k workers are recruited for a task to collect data, the platform processes the data of these k workers and sends the estimated true data to the DR. Therefore, in this model, it is possible to neither leak workers' data to the platform nor leak workers' data content to DR. Therefore, there are many privacy

protection strategies for data content in this model:

(1) Encryption is the most commonly used method which can be found in Ref. [31]. The main encryption method used is homomorphic encryption [31]. In this method, workers use homomorphic encryption to encrypt their own data, and then send the ciphertext data to the platform [31]. After receiving the worker's data, the platform performs homomorphic encryption on the data according to the CRH method and obtains the estimated true data. Then, the estimated true data obtained by the calculation is sent to the DR. DR decrypts the encrypted estimated true data to obtain the estimated true data of the plaintext. In this method, since the platform obtains ciphertext, it cannot know the content of the worker data. DR does not obtain the encrypted data of individual workers, but the estimated true data after calculation by the platform, so it does not know the data content of each worker. This achieves the goal of protecting the privacy of both the platform and DR. Current homomorphic encryption methods cannot perform arithmetic operations like plaintext data [31]. Therefore, in these strategies, basic homomorphic ciphertext calculation methods are often used to define several operations for CRH truth value calculation. In this way, the platform can complete the same function as the plaintext CRH method to find the truth value by calling these operations. The advantage of this method is that the privacy protection function is relatively strong, which can protect the privacy of both the platform and the data content of the DR. However, the disadvantage of this method is that the calculation intensity of homomorphic encryption is high, and the truth value in the CRH method requires repeated homomorphic encryption calculations many times, so its computational complexity is high.

(2) Use secret sharing method in multi-party computation to protect data content privacy [19]. In the secret sharing of multi-party computation, a data packet is divided into $s|s \geq 2$ shares [19]. If the platform knows more than k of these s shares, the contents of the data packet can be recovered. However, if the number of obtained share is less than k, the contents of the data packet cannot be obtained. Therefore, you can use this feature to deploy two or more servers on the platform, and formulate corresponding methods for calculating the corresponding operations of the estimated true data. In this way, if the servers in the platform do not collude, the share obtained by each server will not constitute the content of the recovered data package. However, after collaborative calculations among multiple servers, the final estimated true data can be calculated and given to DR, thereby achieving data content privacy protection for the platform and DR. This type of research can be found in Ref. [19]. This type of method can reduce the large amount of calculation caused by encryption, but the disadvantage is that multi-party calculations between secret shares between servers will cause communication load, but in the current era of easy and rapid development of network communications, it is still a better method.

(3) Lightweight data privacy protection method [27]. Tang et al. believe that the current encryption method brings a large computing load to the platform, so, a lightweight data privacy protection method with less computational requirements is proposed, which is somewhat similar to the method used in this paper. However, the method proposed in this paper has the function of data quality-ware and this method does not. The method adopted by Tang et al. requires a trusted auxiliary server (Assistant Server, AS) that acts as a platform [27]. When collecting data, the AS sends two pieces of data to each worker i participating in the task, called additive noise $\alpha_i$and multiplicative noise β_i. The workers send the two numbers plus α_i and multiplied β_i to their own data $x_i$to the DR [27], respectively. Then, DR and AS cooperate with each other and use the CRH method to calculate so that DR can get the truth value without knowing the workers' data. Although there are some of the above data content privacy protection methods, the main focus of these privacy protection methods is how to protect workers' data privacy. Therefore, these methods do not involve how to collect data quality-ware. Therefore, the quality of data collection will be low. However, some data quality-ware studies that have been proposed do not have the function of data content privacy protection. Therefore, in this paper, we propose a data collection strategy that can both protect data content privacy and data quality-ware for MCS which is of great significance.

(4) The anonymous method is also one of the methods. In this method, workers submit data anonymously. Since the platform does not know the workers' information, it is difficult to connect the data with the workers, thus protecting workers' privacy [36,37].

Table 1
Description of the abbreviation.

Abbreviations	Description
MCS	Mobile Crowd Sensing
RKC	Reputation and Key Center
DPC	Data Processing Center
GTD	Ground Truth Data
ETD	Estimated Truth Data
DOT	Degrees-Of-Trust
IPV	Individual Perturbation Values
ADV	Aggregated Descrambling Values
CRH	Conflict Resolution on Heterogeneous Data

Table 2
The Notation Description.

Notations	Description
RKC	The auxiliary server of reputation and key center
DPC	The cloud server of data processing center
N	The total number of workers to participate the tasks
n	The number of workers selected to participate the tasks
pk_i	The private key of worker i
pk_0	The converged key of all workers
k_i	The participation level of worker i (1 or 0)
k_i'	The scrambled participation level of worker i
$\bar{x}$	The estimated truth
x_i	The observed data of worker i
x_{i1}'	The 1st scrambled sensed data of worker i
x_{i2}'	The 2nd scrambled sensed data of worker i
m_c	The random additive and multiplicative noises
m_{ci}'	The scrambled m_c of worker i
$d(x_i, \bar{x})$	The distance function of workers' sensed data and the estimated truth
x_i'	The scrambled difference between the workers' sensed data and the estimated truth
w_i	The weight of worker i
w_i'	The weight of worker i with random additive noises
w_0	The converged weight of all workers
w_c	The converged weight of all workers with random additive noises
g_t	The ground truth-value of round t
$\tilde{g}_t$	The estimation of $g_{i,t}$
$\mathscr{A}_t$	The accurate rate of round t
$\overline{\mathscr{A}}$	The average accurate rate of all tasks t
U_t	Sets of workers recruited of round t
U_m	Sets of workers applying to participate in tasks
U_n	Sets of workers selected to participate tasks
P_t	The cost of recruiting workers of round t
D_n	Sets of sensed data submitted by workers
T_t	Sets of workers' Degrees-Of-Trust of round t
$T_{i,t}$	The Degrees-Of-Trust of worker i in round t

3. System model and problem statement

3.1. System overview

We consider an MCS system consistent of 3 main components [12-14]: (1) The platform which is composed of 2 parts in this paper, one is the Data Processing Center (DPC) which is responsible for data processing and truth value calculation, and like most research, DPC is

considered cautious and curious [27-33], that is to say, DPC will strictly follow established procedures to process data, and curious to obtain the content of the data. The other is Reputation and Key Center (RKC) which is considered to be completely credible, such as trust center, authorization center, etc. (2) Data Requesters (DR), the DR uses the platform to recruit some workers to perform data collection tasks (e.g., collecting data on traffic flow, water quality, air quality, etc.) in urban areas. (3)A crowd of workers. They refer to the large number of people who own sensing devices in the Internet of Things. Workers earn income by completing tasks posted by the DR and assist the DR in accomplishing tasks [22-23]. However, the quality of the data submitted by workers is often uneven, and it is more likely that there is collusion in order to earn high profits. Therefore, often the quality of task completion cannot be guaranteed (Table 1).

The notations of main variables used in this paper are shown in Table 2.

The tasks and unknown workers are defined as follows:

Definition 1. **(Task, Round)**. The task is to perceive the data of a specified time, and the DR continuously publishes tasks to the platform, the total number is D. The tasks released in the round t are recorded as d_t.

Definition 2. **(Ground Truth Data)**. The truth value refers to the value actually sensing physically, which is the data that exists physically and is measured. truth value refers to true and reliable data, which is the benchmark for data sensory tasks. The truth value is used as a measure of the sensing quality of the data submitted by workers. At the same time, the work effect and quality of each round of PPAT system can be evaluated by the truth value. The truth values of all tasks in the round t are expressed as V_t^G.

Definition 3. **(Unknown Worker, Sensing Quality)**. The system recruits n worker by default in each round, U_t indicating the recruitment set in the round t, which is usually considered $|U_i| = n$.The data submitted by the worker i in the round t is set to $x_{i,t}$, and the data set submitted by all the workers participating in the task in the round t is represented by $U_{t,n} = \{x_{t,1}, x_{t,2} \ldots, x_{t,n}\}$.The index n represents the number of workers participating in the task. In this paper, not all candidates will participate in the task and have an impact on the final truth result.$k_{i,t}$ is used to indicate the participation of worker i in the round t. If $k_{i,t} = 1$, it means that the worker is selected to participate in the task; if $k_{i,t} = 0$, it means that the worker is not selected. At the same time, due to the uneven quality of workers, there is usually a certain error from the truth value, so it is necessary to introduce sensing quality to characterize the quality of the data submitted by workers. The smaller the error from the truth value, the higher the sensing quality.

Definition 4. **(Degrees-Of-Trust).** Since a worker may keep applying for tasks, we use $T_n = \{t_1, t_2, \ldots, t_n\}$to represent the set of Degrees-Of-Trust (DOT) generated by all tasks during the execution of the group intelligence sensory task. In addition, after the task t, each worker has a composite DOT, expressed as $T_n' = \{t_1', t_2', \ldots, t_n'\}$, It is the result of all past trust iterations, and it is the latest DOT of current workers. See Section 4.3 for the specific update algorithm.

Definition 5. **(Cost).** After the selected workers participate in the task, the platform will calculate the rewards that need to be given to the workers, that is, the recruitment cost of the system. After a round of tasks is over, the recruitment cost of each worker is expressed as $P_t = \{p_1, p_2, \cdots, p_n\}$.

3.2. Threat model

In this paper, it is assumed that the entities involved are semi-honest [44-46], which is specifically reflected in: on the one hand, the entity will strictly abide by the agreement and will not deliberately tamper with the agreement or calculation results; on the other hand, the entity will explore and disclose the privacy of others as much as possible. The specific discussion from various angles can be divided into:

(1) For sensory platforms, they will try to know the privacy of workers, including the true identity of workers, real data and other private information, and will try to obtain a series of calculation results such as truth value, data weight, and DOT.
(2) For workers, they will try to know the weight and DOT of their own and others' data, and may submit false data, or conspire with other workers to obtain more rewards.
(3) For other entities that are not in the system, there may be entities that are interested in sensing data and results, and they will try their best to obtain the privacy of system data and workers.
(4) There will be no collusion between sensory platforms. At the same time, this paper defaults that there is no collusion between sensory platforms and workers. In order to protect their privacy, workers will not conspire with the sensory platform.
(5) Once data leaves the data source, it is easy to have data poisoning. In this paper, we mainly use the trust mechanism to reward workers who submit high-quality data and improve their trust; To punish workers who submit low-quality data, reduce their trust, and judge the quality of workers by the cumulative trust of workers, which examines the long-term performance of workers, can effectively resist workers to cheat the trust of the system through camouflage in a short time.

3.3. Design goals

In this paper, our goal is to design a sensory system with high accuracy, low-cost and privacy preserving of the collected data. Specifically:

(1) Privacy preserving: The privacy of workers' sensory data, workers' DOT, and estimated truth data (ETD) is protected in our agreement. Specifically, the data sensed by workers should not be disclosed to cloud servers and other participants, and the DOT and ETD of workers should not be exposed to participants and cloud servers that are not involved.
(2) Maximize the collected data quality: The quality of workers is uneven, and the submitted data may also deviate greatly from the real data. Moreover, there may be collusion between workers, which makes the ETD deviate greatly from the truth value. The goal of this paper is to make the ETD calculated based on the data submitted by the workers as close to the truth value as possible.

In this paper, the ETD is used as the result and reported to DR. According to the above definition, we set the ETD of the round t to be $\widehat{g}_t$., and the actual truth value of the corresponding system to be g_t. Then the accuracy of the round t of data collection $\mathscr{A}_t$ is shown in Eq. (1),$\mathscr{A}_t \in [0, 1]$. Among them, we use constant parameters δas adjustment parameters in different application scenarios.

$$\mathscr{A}_t = \frac{1}{1 + \delta \times \left| \frac{g_t - \widehat{g}_t}{g_t} \right|} . \tag{1}$$

According to Eq. (1), the ETD $\widehat{g}_t$is to the actual truth value $g_t$of the system, the higher the quality of accuracy. When $\widehat{g}_t$ equal tog_t, the accuracy reaches the maximum value (at this time$\mathscr{A}_t = 1$). On the contrary, if the gap between $\widehat{g}_t$ and $g_t$is large, the accuracy $\mathscr{A}_t$ will also be significantly reduced. The average data quality in all cycles is shown in Eq. (2):

Table 3
The cases using CRH mechanism.

Wks	Case 1		Case 2		Case 3		Case 4		Case 5	
	x_i	w_i	x_i	w_i	x_i	w_i	x_i	w_i	x_i	w_i
s_1	**49**	**2.6**	**49**	**7.8**	**49**	**6.2**	**49**	**1.6**	**49**	**0.8**
s_2	53	1.9	53	5.5	53	4.5	53	1.4	53	0.6
s_3	48	1.7	48	6.4	48	7.0	48	1.7	8	7.9
s_4	52	3.0	52	6.4	52	4.8	52	1.4	2	4.3
s_5	47	1.0	47	5.6	47	8.3	7	3.5	7	6.7
s_6	51	5.7	51	7.7	51	5.2	11	4.3	11	7.0
s_7	49	2.6	48	7.8	9	1.2	9	3.9	9	11.9
s_8	51	5.7	51	7.7	11	1.3	11	4.3	11	7.0
s_9	52	3.0	2	0.1	2	0.9	2	2.8	2	4.3
ETD	50.6434		49.9036		45.9416		18.747		9.09703	
Accu Rate	98.73%		99.81%		92.49%		37.49%		18.19%	

$$\overline{\mathscr{A}} = \frac{1}{n}\sum_{t=1}^{n}\mathscr{A}_t \ . \tag{2}$$

Obviously, the higher the average data quality and the closer the ETD is to the truth value, the better the performance of the PPAT scheme. Therefore, our goal is to maximize the average quality of the collected data:Max($\overline{\mathscr{A}}$).

(3) Minimize the data collection cost: Regardless of the quality of the data provided by the workers, the sensory system needs to distribute rewards to each worker. Through pre-screening, only high-quality workers are recruited, thereby reducing the scale of task participants, reducing the rewards that need to be issued, and reducing costs.

This paper mainly considers the cost of worker recruitment. We use $\mathfrak{C}(.)$ to represent the cost function, which increases monotonically relative to the number of workers recruited. Simply put, we assume that the cost of recruiting different employees is equal, which p_t is used to express it. In Section 3.1, we use U_t to represent the set of workers recruited in the round t, then the cost of data collection for all cycles is calculated as shown in Eq. (3).

$$\mathscr{C} = \mathfrak{C}\left(\sum_{t=1}^{n} \| U_t \|\right) = \mathfrak{C}\left(\sum_{t=1}^{n}\left(\sum_{i=1}^{m} k_{i,t}\right) \times p_t\right) \ . \tag{3}$$

Eq. (3) $\| \ . \ \|$represents the operation of calculating the number of elements in the set, so $\| U_t\|$ represents the number of workers recruited in the round $t$of tasks.In the formula $k_{i,t}(k_{i,t} \in (0 \ or \ 1),)$ calculates the participation for the truth value of the workers in the round t of tasks, and the recruitment cost p_t for worker i in the round t of tasks. As we all know, the lower the cost of data acquisition, the better the system performance. Therefore, our goal is to minimize the cost of comprehensive data collection:Min($\mathscr{C}$). This means that we should choose trustworthy workers to participate as much as possible and exclude untrustworthy workers.

Overall, the goals of this paper are as follows:

Get Data Privacy-Preserving and Max($\overline{\mathscr{A}}$) and Min($\mathscr{C}$) as followings.

$$\begin{cases} Max(\overline{\mathscr{A}}) = Max\left(\frac{1}{n}\sum_{j=1}^{n}\mathscr{A}_t\right), \\ Min(\mathscr{C}) = Min\left(\sum_{t=1}^{n}\left(\sum_{i=1}^{m} k_{i,t}\right) \times p_t\right). \end{cases} \tag{4}$$

4. The proposed PPAT scheme

In this section, we introduce the process of the PPAT mechanism. First of all, we introduced the research motivation of this paper. Secondly, we describe the calculation of the ETD under the protection of privacy, and then introduce the calculation of workers' DOT under privacy protection. After that, we described the truth value calculation based on DOT under privacy protection, and finally introduced the worker recruitment strategy based on DOT under privacy protection.

4.1. Research motivation

As mentioned earlier, CRH is a classic truth discovery protocol with excellent performance on heterogeneous data [33]. Almost all privacy protection truth discovery protocols are based on CRH, and the protocols proposed in this paper are no exception. Therefore, we will briefly introduce the CRH composed of two stages: weight update and truth estimation [33].

Based on the data x_i sent by each worker, platform first uses the mean method to calculate and estimate the truth value $\bar{x}$:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \ . \tag{5}$$

(1) After obtaining the average$\bar{x}$, perform a weight w_i update: the sensing data x_i and ETD $\bar{x}$ are known, and the calculation of the worker's weight w_i update is shown in Eq. (6).

$$w_i = log\frac{\sum_{i=1}^{n} d(x_i, \bar{x})}{d(x_i, \bar{x})} = log\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(x_i - \bar{x})^2} \ . \tag{6}$$

(2) According to the above updated weight, update the ETD: known w_i and sensing data x_i, the ETD $\bar{x}$ of the sensing task is:

$$\bar{x} = \frac{\sum_{i=1}^{n}(w_i \times x_i)}{\sum_{i=1}^{n} w_i} \ . \tag{7}$$

Repeat the two steps (1) and (2), and use an iterative mechanism to update the weights and ETD until the ETD is the same as the ETD calculated in the previous iteration, and the resulting data is the ETD calculated by CRH [33].

However, the basic assumption for the effectiveness of this method is that the vast majority of workers have a higher sensing quality. Once the sensing quality of worker decreases or collusion occurs, the performance of CRH will deteriorate. Let's use an example to illustrate. The example given is shown in Table 3. There are 9 workers recruited in each round, and the truth value of the task is set to 50. The bold data in the table indicates the data submitted by highly trustworthy workers, the red data indicates the data submitted by untrustworthy workers, and the others

Table 4
The cases using PPAT mechanism.

Wks	Case 1		Case 2		Case 3		Case 4		Case 5	
	x_i	w_i	x_i	w_i	x_i	w_i	x_i	w_i	x_i	w_i
s_1	**49**	**2.6**	**49**	**7.8**	**49**	**6.2**	**49**	**1.6**	**49**	**0.8**
s_2	53	1.9	53	5.5	53	4.5	53	1.4	53	0.6
s_3	48	1.7	48	6.4	48	7.0	48	1.7	8	0
s_4	52	3.0	52	6.4	52	4.8	52	1.4	2	0
s_5	47	1.0	47	5.6	47	8.3	7	0	7	0
s_6	51	5.7	51	7.7	51	5.2	11	0	11	0
s_7	49	2.6	48	7.8	9	0	9	0	9	0
s_8	51	5.7	51	7.7	11	0	11	0	11	0
s_9	52	3.0	2	0	2	0	2	0	2	0
ETD	50.6434		49.9909		49.5333		50.3279		50.7143	
Accu Rate	98.73%		99.98%		99.08%		99.35%		98.59%	

are the data submitted by workers of unknown trustworthiness. Table 3 shows the presence of different numbers of untrustworthy workers in each round of tasks. The weight of each worker, the final ETD, and the accuracy of the ETD calculated according to the CRH method are shown in Table 3:

As can be seen from Table 3, with the increase of the number of untrustworthy workers, the deviation between the ETD calculated by the CRH method and the truth value is getting larger and larger. It can be seen that the CRH method, when there are untrustworthy workers in it, or when workers conspire to attack, the accuracy of the data obtained cannot be guaranteed. The reason for this is that the CRH method does not identify the DOT of workers, but randomly selects a number of workers, and finally weights the worker data to obtain the final result. As shown in Fig. 2 is the situation in the third round of Table 3. As a result, the data quality cannot be guaranteed.

To solve the above issues, for the first time, we have proposed a Privacy Preserving, Accuracy, and Trust data collection scheme (PPAT) for MCS, which can protect the privacy of data content and maintain high accuracy at low-cost style. In this paper, we propose a trust-based worker selection method. The method we propose is shown in Fig. 3. We assume that at the beginning, we know that a small number of workers are trustworthy, and other workers have unknown trustworthiness. Then, we use the weights calculated by this small group of workers as a benchmark to test the weights of other workers. Based on the comparison results between the two, if the error between the weights calculated by unknown workers and the weights of highly trustworthy workers is less than a predetermined threshold, then increase their DOT, and when their DOT rises beyond the predetermined threshold, they become highly trustworthy workers. Conversely, if the error between the data submitted by the unknown worker and the data of the highly trustworthy worker is greater than a predetermined threshold, the worker's DOT will be reduced. If the DOT is lower than a certain low threshold, it will be attributed to the collection of untrustworthy workers. As shown in Fig.3, in a round of data collection, we know that the first worker s_1 is trustworthy, and the other workers are workers with unknown trustworthiness. Comparing the data submitted by them (see case 3 in Table 3) will increase the DOT of workers $s_1 \sim s_6$ and reduce the DOT of workers $s_7 \sim s_9$ accordingly.

After updating the DOT of the workers, in the calculation of the ETD, we also calculate the ETD based on the DOT of the workers. Because we know that worker s_1 is trustworthy. The data weights of workers $s_2 \sim s_6$ are very close to them, so they can also be considered credible. In this way, we calculate their weights normally, which are 4.5,7.0,4.8,8.3,5.2, If the weight of the other three workers $s_7 \sim s_9$is set to 0, it can be calculated that the final ETD is 49.5333. In such a situation, even if there are 3 malicious workers, the algorithm we propose can still get an accurate ETD.

Thus, we use the same data as Table 3, and use the PPAT mechanism to re-calculate the ETD for 5 rounds of tasks.

We can see that the first worker s_1 in each task is known to the system from a trusted collection, and the remaining workers are from a collection with unknown DOT. For the weights calculated by CRH in Table 3, we take the weights of highly trustworthy worker s_1 as the benchmark and adjust the weights of other workers (closer to the benchmark, increase their DOT; if the deviation with the benchmark is large, reduce their DOT). At the same time, for workers with a large benchmark deviation, set their weight to 0 (as shown in the data marked in red in Table 4, that is, the worker s_9 in the second round, the workers $s_7 \sim s_9$in the third round, the workers $s_5 \sim s_9$in the fourth round, and the workers $s_3 \sim s_9$ in the fifth round), so the data submitted by these untrustworthy workers will not be used in the calculation of the final ETD.

As can be seen from Table 4, no matter how many untrustworthy workers there are, the accuracy of the ETD calculated by the PPAT mechanism is more than 98%. Comparing the calculation results of the CRH mechanism in Table 3, the PPAT mechanism has a significant resistance effect when the data of untrustworthy workers increases.

Therefore, we want to always have a sufficient number of highly trustworthy workers in the system for us to choose to participate in the task. However, it is unfortunate that at the beginning of the system, the workers were all workers with unknown DOT. So, a set of DOT update algorithms is needed to iterate the DOT of all workers, so as to distinguish between highly trustworthy people and untrustworthy people in the crowd. PPAT proposes to calculate and update the DOT of workers based on the quality of the data submitted each time. Through continuous iteration, the crowd can be distinguished. In order to show the evolution of workers, there are the following experiments: We obtained a set of weather and temperature data. Among the 1000 workers, 50 were randomly selected every hour to provide an observation value. In this set of data, we randomly divided workers into three categories. The first category is called "honesty", which usually sensors an accurate temperature with a true error range of 10% from the ground, and the number of people in the honest category accounts for 50% of all participants; the second category is called "dishonesty", which usually sensors inaccurate temperature, with an error range of 30% of the ground truth, and the number of people in the dishonest category is 50%.In this iteration process, the trend of changes in the DOT of workers from 1 to 5 under the PPAT mechanism is shown in Fig. 4.

The initial situation at the beginning of the system, when the default is all trustworthiness unknown population. We believe that untrustworthy participants will keep submitting untrustworthy data. According to Algorithm 2, the DOT will gradually decrease with a decreasing rate, and the lower the quality of the data submitted by workers, the faster the DOT decreases, such as worker3, worker4, and worker5 in Fig.4; similarly, we think that the trustworthy participants will keep submitting trustworthy data, and the DOT of trustworthy participants will gradually increase with an increasing rate, and the higher the quality of the data

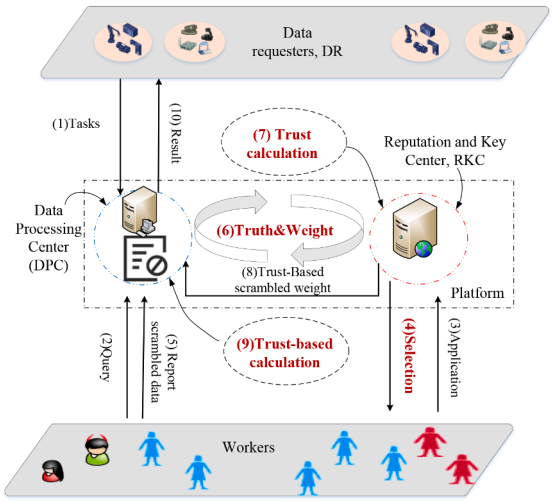


Fig. 1. MCS system structure.

Fig. 2. Data and weight of workers in case 3 using CRH.

Fig. 3. Data and weight of workers in case 3 using PPAT.

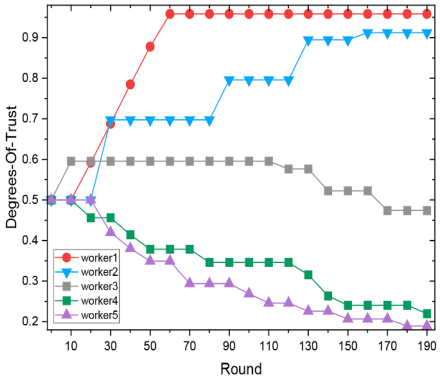


Fig. 4. The changing trend of workers' DOT under the PPAT mechanism.

submitted by workers, the faster the DOT increases, such as worker1, worker2 in Fig.4.

After a period of iteration, the DOT of the workers has stabilized except for a very small number of workers due to insufficient number of iterations, and the DOT of some workers will continue to increase and the DOT of some workers will continue to decrease. Therefore, according to the comparison with the pre-set thresholds, the population will be gradually differentiated into trustworthy workers and untrustworthy workers. During the iteration process, the proportion of various groups of all workers is shown in Fig.5. It can be seen that after iteration, workers with high DOT and low DOT have been effectively distinguished.

Above, we have explained the main mechanism of PPAT, and the

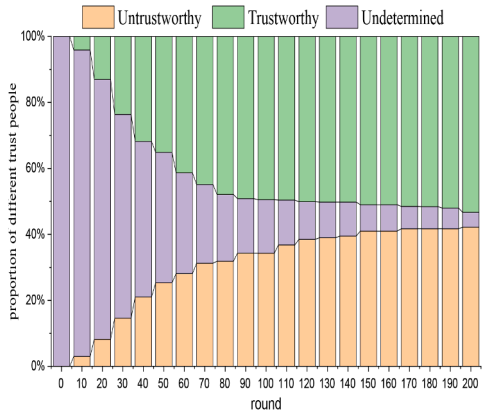


Fig. 5. Trends in the proportion of various groups of people under the PPAT mechanism.

specific mechanism of the system is shown in Fig 1. There are three types of people in the crowd: people with high trustworthiness, people with unknown trustworthiness, and people who are not trusted.

(1)Tasks. Data requesters generate a series of tasks as needed. Subsequently, DR will submit these tasks to the system platform and wait for workers to apply.
(2)Query. Workers who wish to participate in a perceived task will query the platform for the task with a view to obtaining a list of tasks so that they can select the task to complete and receive a reward.
(3)Application. After workers get the task list, they will choose their favorite task from it, and then submit a task application to the platform.
(4)Selection. After RKC receives the task application submitted by the workers, it will sort the DOT of the workers, and then select a number of workers from high to low to participate in the task. See Section 4.5 for the specific selection mechanism.
(5)Report scrambled data. After the worker recruitment is completed, the selected workers will complete the data collection. For the purpose of protecting privacy, the workers will scramble the original data, and finally submit the scrambled data to DPC for the next step of processing. See Section 4.2 for the scrambling processing formula.
(6)Truth & Weight. After DPC receives the scrambled data submitted by the workers, it will calculate the initial ETD, and RKC will be responsible for the initial weight calculation. The two platforms will continue to iteratively calculate and update the initial weight and estimate the truth value until they converge. See 4.2 for the specific computing mechanism.
(7)Trust calculation. When DPC and RKC calculations are completed, RKC will update the DOT of all workers participating in the task based on the initial weight of highly trustworthy workers. See Section 4.3 for the specific update mechanism.
(8)Trust-Based scrambled weight. When RKC has updated the DOT of all workers participating in the task, the scrambling weights of all workers will be generated and sent to DPC for the final truth value calculation. See Section 4.4 for the formula for generating the scrambling weight.
(9)Trust-based calculation. After the DPC receives the scrambled DOT from all workers, it will calculate the estimated truth value again based on the scrambled DOT on top of the initial ETD to get the final result. See Section 4.4 for specific computing mechanism.
(10)Result. After DPC completes the truth value calculation based on the DOT, the final ETD will be generated. Then DPC will send the final ETD to DR as the final result of this sensory task, and the task is completed at this time.

4.2. Privacy-preserving calculation of estimated truth data

Previously mentioned truth estimation algorithms are calculated in plaintext. This may threaten the privacy security of workers. To protect workers' privacy, this paper uses a scrambling strategy so that DPC gets the plaintext sum of all sensed data instead of all data itself. Therefore, DPC will not learn the real data submitted by workers. RKC generates the individual perturbation values (IPV) of the workers and produces the aggregated descrambling values (ADV) required by DPC, whereby DPC can learn the sum of all encrypted data, but cannot learn the specific data of each worker. In this way, after iterative calculation, DPC can get the ETD without knowing the specific weights of the workers; RKC can get the weights of the workers without knowing the ETD. Thus, the privacy of workers is effectively protected. The specific process is shown below.

(1) RKC generates, for each worker, a corresponding IPV$\{pk_1, pk_2, ...,pk_n\}$ (generate random numbers with hash function), which are sent to the corresponding workers. At the same time, the ADV of DPC is generated according to Eq. (8) and sent to DPC.

$$pk_0 = \sum_{i=1}^{n} pk_i \ . \tag{8}$$

(2) Each worker processes the acquired sensory data x_i according to the IPV pk_i to obtain the scrambled sensory data $x_{i1}{}'$, which is sent to DPC.

$$x'_{i1} = x_i + pk_i \ . \tag{9}$$

(3) DPC receives the scrambled data $\{x_{11}{}', x_{21}{}', ...,x_{n1}{}'\}$ from each worker. combined with the ADV pk_0, and processed using Eq. (10) to obtain the ETD:

$$\overline{x} = \frac{1}{n} \times \left(\sum_{i=1}^{n} x'_{i1} - pk_0 \right) \ . \tag{10}$$

On the basis of Eqs. (8)-(10), the following can be deduced:

$$\begin{aligned} \overline{x} &= \frac{1}{n} \times \left(\sum_{i=1}^{n} x'_{i1} - pk_0 \right) = \frac{1}{n} \times \left(\sum_{i=1}^{n} (x_i + pk_i) - pk_0 \right) \\ &= \frac{1}{n} \times \left(\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} pk_i - \sum_{i=1}^{n} pk_i \right) = \frac{1}{n} \times \sum_{i=1}^{n} x_i \end{aligned} \tag{11}$$

According to the calculation results of Eq. (11), it can be seen that DPC, without being informed of the real sensory data of the workers, obtains the ETD of the system through the algorithm provided in this paper, and therefore effectively protects the privacy of the workers.

(4) DPC relies on the received scrambled data $\{x_{11}{}', x_{21}{}', ...,x_{n1}{}'\}$ sent from each worker, as well as the ETD $\overline{x}$, is processed using Eq. (12) to obtain the scrambled deviation data $\{x_1{}', x_2{}', ...,x_n{}'\}$, which is sent to RKC.

$$x'_i = x'_{i1} - \overline{x} \ . \tag{12}$$

On the basis of Eqs. (9) and (12), the following can be deduced:

$$x_i - \overline{x} = x'_{i1} - pk_i - \overline{x} = x_i{}' - pk_i \ . \tag{13}$$

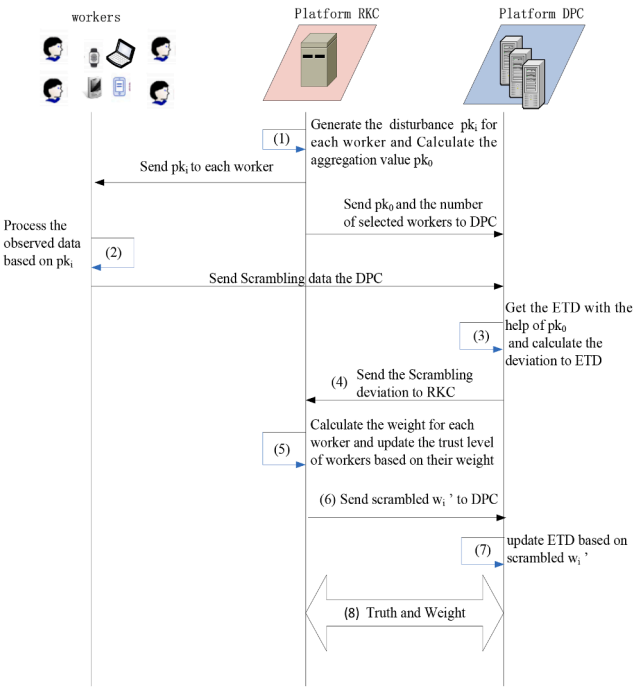


Fig. 6. The flowchart for calculating the ETD of privacy preserving.

(5) RKC receives the scrambled deviation data of each worker from the ETD $\{x_1{}', x_2{}', ...,x_n{}'\}$ from DPC, according to Eq. (13), the actual deviation value of each worker from the ETD can be calculated and obtained to update the weights:

$$d(x_i, \overline{x}) = (x_i - \overline{x})^2 = (x_i{}' - pk_i)^2 \ . \tag{14}$$

$$w_i = log \frac{\sum_{i=1}^{n} d(x_i, \overline{x})}{d(x_i, \overline{x})} = log \frac{\sum_{i=1}^{n} (x_i{}' - pk_i)^2}{(x_i{}' - pk_i)^2} \ . \tag{15}$$

According to the calculation result of Eq. (15), at this time, RKC learns the weight information of all workers without knowing the real sensory data of workers and the real ETD of the system.

Algorithm 1
The PP_ETD Algorithm.

Input: D_n, $\{x_i | i \in D_n\}$, pk_i, n
Output: $\overline{x}$, w_i

Platform RKC generates pk_i
RKC Calculates the pk_0 by Eq.(8)
Send the value of pk_0 to Platform DPC
foreach $i \in D_n$ **do**
  $x_{i1}{}' = x_i + pk_i$;
  Work i sends $x_{i1}{}'$ to Platform DPC
end for
Platform DPC Calculates the $\overline{x}$ by Eq.(10)
foreach $i \in D_n$ **do**
  $x_i{}' = x_{i1}{}' - \overline{x} = x_i + pk_i - \overline{x}$;
  Platform DPC sends $x_i{}'$ to Platform RKC
end for
RKC Calculates the w_i by Eqs.(14) and (15)
Initialize $\overline{x_p}$
While $\overline{x_p} \ != \overline{x}$ **do**
  $\overline{x_p} = \overline{x}$
  $k_i = 1$;
  $\overline{x}$= *PPAT_Truth*(D_n, T_t, w_i, k_i)
  $x_i{}' = x_{i1}{}' - \overline{x} = x_i + pk_i - \overline{x}$;
  RKC Calculates the w_i by Eqs.(14) and (15)
end while
return $(\overline{x}, w_i)$;

At this time, RKC learns the weight information of all workers, and DPC learns the ETD of the system. However, by default, RKC and DPC will not be complicit, i.e., they will not exchange their known information with each other, so that the weight information and ETD will not be known by the same platform, and at the same time, any platform will not be able to get the real sensory data of the workers, which effectively protects the privacy of the workers.

(6) RKC gets the scrambled weights w_i' (set k_i are all 1 at this point) according to Eq. (22) in Section 4.4 and sends them to DPC.
(7) DPC recalculates the ETD based on the updated weights w_i' according to Eq. (25) in Section 4.4.
(8) Steps (4)-(7) are repeated, using an iterative mechanism for weight and truth value updating, until the ETD is the same as the ETD calculated in the previous round of iteration. The flowchart for calculating the ETD of privacy preserving is shown in Fig. 6.

4.3. Privacy-Preserving calculation of workers' degrees-of-trust

After obtaining the weights of the workers in Section 4.2, the DOT of the other workers is updated using the weights of the trusted workers as a criterion. The input to the algorithm in this section is the output of Algorithm 1: the weights of each worker, and the DOT of the workers is updated based on the output.

The privacy-preserving ETD calculation effectively protects the worker's privacy and initially calculates the worker's weight information. However, due to the varying sensory quality of workers, the ETD may suffer from low quality. Therefore, we introduce a DOT correlation algorithm to select highly trustworthy workers to improve the accuracy of the ETD. In order to efficiently select highly trustworthy workers, we need to compute the DOT of the current worker based on the data submitted by each worker and iteratively update the total DOT of the workers.

After RKC receives the scrambled deviation data of each worker from DPC, it calculates the weight of each worker and then updates the DOT of each worker. In this paper, we consider that the workers with higher DOT provide more trustworthy data and the truth value accuracy of the tasks participated by them is higher, thus they should be selected to participate in the tasks more. The rule for computing the trust of a worker is as follows:

When the system is initialized, workers are always set as unknown perceivers, and the initial value of trust is set to ψ.

When the unknown worker's DOT is computed by the following rule, when it is greater than ξ_2, it can be transformed into a highly trustworthy worker; if it is lower than ξ_1, it is transformed into an untrustworthy worker.

The result of each task, whose DOT is updated and computing according to the following rules:

When there is a highly trusted worker in the task, the weights of the other workers are compared with the weights of the highly trusted worker, and if the result is within the acceptable error range τ, the trust index $A_{i,t}$ of the ith worker is increased by 1. If there is no highly trusted worker in the task when there is only an unknown worker, the weights of the other workers are compared with the weights of the trusted worker with the highest DOT, and if the result of the comparison is within the error range τ, the $A_{i,t}$ is increased by 1 as well; the comparison accuracy of the ith worker at time t is counted as in the following equation, where $A_{i,tot}$ is the number of the total number of comparisons:

$$\sigma_r = 1 - \frac{A_{i,t}}{A_{i,tot}} \ . \tag{16}$$

$T_{i,t}$ represents the overall DOT of the ith worker at time period t, while $T_{i,t-1}$ denotes the DOT of the worker at time period t-1. $T_{i,0}$ represents the initial value of the overall DOT of the ith worker, and the overall DOT can be increased by a maximum of $\frac{1}{\rho}$ (when $\sigma_r = 1$) per update. A threshold σ_s is defined to determine whether to perform a DOT increase or decrease, e.g., if σ_r is 0.6 indicating that the probability of a worker collecting data accurately at time period t is 0 0.6, then greater than 0.6 it is considered trustworthy and increases its DOT. if σ_r is greater than σ_s, the worker is considered trustworthy, ρ is a variable greater than 1, which is used to control how fast or slow the DOT is updated each time. If σ_r is greater than σ_s, the equation for updating the worker's DOT is as follows:

$$T_{i,t} = T_{i,t-1} + \frac{\sigma_r}{\rho} * \left(1 - T_{i,t-1}\right) \ . \tag{17}$$

If σ_r is less than σ_s, the equation for updating the worker's DOT is as follows:

$$T_{i,t} = T_{i,t-1} + (\sigma_r - 1) * \left(\frac{T_{i,t-1}}{\rho}\right) \ . \tag{18}$$

Algorithm 2
The DOT_Update Mechanism.

Input: T_{t-1}, $\{T_{i,t-1} | T_{i,t-1} \in T_{t-1}\}$, w_i
Output: T_t, $\{T_{i,t} | T_{i,t} \in T_t\}$
$w_{max} = \text{argmax}_{i \in D_n}(w_i)$
foreach $i \in D_n$ **do**
  if $|w_i - w_{max}| \le \tau$ **then**
    $A_{i,t}$++;
  end if
  Platform RKC Calculates the σ_t by Eq.(16)
  if $\sigma_t > \sigma_s$ **then**
    Calculate the $T_{i,t}$ by Eq.(17)
  else
    Calculate the $T_{i,t}$ by Eq.(18)
  end if
end for
return $(T_t, \{T_{i,t} | T_{i,t} \in T_t\})$;

Algorithm 3
The PPAT_Truth Algorithm.

Input: D_n, $\{x_i | i \in D_n\}$, T_t, $\{T_{i,t} | T_{i,t} \in T_t\}$, w_i, k_i
Output: $\overline{x}$
Platform RKC generates pk_i and m_c
foreach $i \in D_n$ **do**
  if $k_i! = 1$ **then**
    if $T_{i,t} \ge \xi_2$ **then**
      $k_i = 1$;
    else
      $k_i = 0$;
    end if
  end if
  $w_i' = w_i \times k_i'$
  $k_i' = k_i + m_c$;
  $m_{ci}' = w_i \times m_c$
  Send the value of pk_i and m_{ci}' to work i
end for
RKC Calculates w_0 and w_c by Eq.(23) and (24)
RKC sends w_0 and w_c to Platform DPC
foreach $i \in D_n$ **do**
  $x_{i1}' = x_i + pk_i$;
  $x_{i2}' = m_{ci}' \times (x_i + pk_i)$
  Work$i$ sends x_{i1}' and x_{i2}' to DPC
end for
Platform DPC Calculates the $\overline{x}$ by Eq.(25)
return $(\overline{x})$;

4.4. Privacy-preserving calculation of trust-based truth value

Algorithm 2 updates the DOT of the workers and Algorithm 3, based on the updated DOT of the workers, processes the workers' data again by increasing the weight of the data submitted by highly trustworthy

workers, i.e., setting it to 1, and decreasing the weight of the data submitted by lowly trustworthy workers, i.e., setting it to 0. Through this strategy, the ETD is updated again and the final result of the ETD is calculated.

(1) RKC outputs the scrambled DOT according to the DOT generated from the latest round of iteration based on the trust-based truth value calculation rule (set k_i to 1 for those involved in the calculation of the truth value, and 0 otherwise), and also generates a random mask m_c and outputs scrambling DOT $k_i^{'}$.

$$k_i^{'} = k_i + m_c;\ k_i = \begin{cases} 1, //\textit{trusty and undertermined} \\ 0. //\textit{untrustworthy} \end{cases} \tag{19}$$

(2) RKC utilizes random masks m_c and weights w_i to output scrambling masks $m_{ci}^{'}$ to be sent to the corresponding workers:

$$m_{ci}^{'} = w_i \times m_c\ . \tag{20}$$

(3) Each worker processes the acquired sensory data according to the scrambling mask m_{ci}' and IPV pk_i. And sends the obtained scrambled data $x_{i2}^{'}$ to DPC.

$$x_{i2}' = m_{ci}^{'}\ \times (x_i + pk_i)\ . \tag{21}$$

(4) RKC sends the weights w_i, plus scrambling k_i, to DPC as scrambled weights w_i', and sends the aggregated weights w_0 and w_c to DPC:

$$w_i^{'} = w_i \times k_i^{'}\ . \tag{22}$$

$$w_0 = \sum_{i=1}^{n} (w_i \times k_i)\ . \tag{23}$$

$$w_c = \frac{\sum_{i=1}^{n} (w_i \times k_i \times pk_i)}{w_0}\ . \tag{24}$$

Algorithm 4
The WorkerSelection Mechanism.

Input: U_m, T_t, $\{T_{i,t} | T_{i,t} \in T_t\}$
Output: U_n

```
Initialize S_high = 0, S_unknown = 0
foreach i ∈ U_m do
  if (T_i,t ≥ ξ_2) then S_high++;
  else if (T_i,t ≥ ξ_1) then S_unknown++;
  end if
end for
if (Trustworthy +Untrustworth) ≤ β_1 then
  Select all of participates
  else if(Trustworthy +Untrustworth) ≤ β_2 then
  if (S_high ≥ φ) then
    Select φ works from high to low levels of trust
  else if (S_high + S_unknown ≥ φ) then
    Select φ works from high to low levels of trust
  else
    return false; // there are no enough trustworthy works, this round is cancelled.
  end if
else
  if (S_high + S_unknown ≥ α) then
    Select α works from high to low levels of trust
  else
    return false; // there are no enough trustworthy works, this round is cancelled.
  end if
end if
return (U_n);
```

(5) DPC receives the scrambling weights for each worker and updates the ETD.

$$\begin{aligned} \overline{x}\ &= \frac{\sum_{i=1}^{n} ((w_i^{'}\ \times x_{i1}^{'}) - x_{i2}')}{w_0} - w_c \\ &= \frac{\sum_{i=1}^{n} ((w_i \times (k_i + m_c)\ \times\ (x_i + pk_i) - w_i \times m_c \times (x_i + pk_i))}{w_0} - w_c \\ &= \frac{\sum_{i=1}^{n} (w_i \times k_i \times x_i)}{\sum_{i=1}^{n} (w_i \times k_i)} + \frac{\sum_{i=1}^{n} (w_i \times k_i \times pk_i)}{w_0} - w_c \\ &= \frac{\sum_{i=1}^{n} (w_i \times k_i \times x_i)}{\sum_{i=1}^{n} (w_i \times k_i)}\ . \end{aligned} \tag{25}$$

In the process of the ETD calculation, the reliability of the ETD is effectively enhanced because the most trustworthy workers are selected to participate in the task; for the collusion of malicious workers, the data submitted by these malicious workers will be excluded due to their relatively low DOT, which avoids the bias of the ETD caused by the collusion and protects the motivation of ordinary workers to participate in the task.

4.5. Privacy-preserving trust-based worker selection

In the initial stage of the system, all workers in the system are participants whose DOT is unknown. At this time, to ensure the accuracy of the ETD, all the candidates should be selected to participate in the task. Worker differentiation training can also be performed at this time at a faster rate to further facilitate system iteration.

After the task is released, the high trustworthy workers are preferentially selected among the applicants, so that this group of people can be used as the truth value reference standard to ensure the effectiveness of the truth value estimation. And a number of workers whose DOT is unknown are selected to make up the required number. The reason for selecting workers whose DOT is unknown is that the number of initially set high trustworthy workers is small, which may not be able to meet the required number of people standard for the task, so a part of the workers whose DOT is unknown is needed to complete the task. At the same time, the workers whose DOT is unknown can be trained iteratively based on the data of the highly trustworthy workers to update the DOT, with a view to selecting new highly trustworthy workers as well as eliminating the untrustworthy workers from them. In the long run, when the iterative process continues, the number of highly trustworthy workers in the network will be more and more, the proportion is higher and higher, and the number of workers needed for each task will be less and less, eventually converging to the minimum number, so the scale of task participants can be reduced, and fewer rewards can be issued, thus reducing the cost; and because all the selected people are trustworthy, the DOT and accuracy of the ETD computed by iteration is higher.

The criteria for the selection of workers are as follows:

(1) The system has a portion of highly trustworthy workers, the data requester publishes a task that requires the number of participants is at least α, and m of the workers in the network apply for the task from the platform after obtaining the task;
(2) After receiving a task participation request from m workers, the perception platform selects workers according to the following rules:
(a) If there is more than one worker with the DOT above the trustworthiness comparison threshold ξ_2 in each round of workers, and the number of workers is greater than the headcount threshold φ, then φ workers are selected from them.

(b) If there are fewer than φ workers with DOT above the trustworthiness comparison threshold ξ_2 in each round of workers, all workers with DOT above the trustworthiness comparison threshold ξ_2 are selected and some workers whose DOT are unknown are selected to complement the headcount threshold φ.
(c) If the task is applied only by untrustworthy workers, no workers will be selected for the task and the round will be nullified, waiting for the next round of task to be released.

RKC notifies the selected workers to collect data and submit sensory data to DPC.

After the system has converged iteratively, i.e., when the sum of the proportion of highly trustworthy people and the proportion of untrustworthy people reaches a high proportion threshold β_2, at this point we consider that there are only a relatively small number of training-uncompleted DOT unknowns in the system. This represents the presence of a sufficient number of highly trustworthy people among the respondents for each task. Due to the high quality of the data submitted by the highly trustworthy people, we can assume that the error between their submitted data and the true value is negligible. Assuming that the minimum number of workers required to complete a task is α, the number of people we need to select for each task at this point is the minimum number of workers required to complete the task. If the sum of the number of highly trustworthy people and the number of people with unknown DOT among the applicants in this round is less than α, then the task is nullified in this round. The mechanism for worker selection is shown in Algorithm 4.

4.6. Efficiency and privacy protection capability analysis

4.6.1. Theorem and proof

Theorem 1. *The PPAT mechanism is truthful and low-cost.*

Proof. For $\forall$ worker $s_i \in S$ submits its observed data $x_{i,t}$ to complete tasks in $D_{i,t}$. According to Eq.(6) $w_i = log\frac{\sum_{i=1}^{n} d(x_i,\bar{x})}{d(x_i,\bar{x})} = log\frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{(x_i-\bar{x})^2}$, the closer the data submitted by the worker s_i is to the ETD $\bar{x}$, the higher the weight is; if the deviation between the data submitted by the worker s_i and the ETD $\bar{x}$ is larger, the lower the weight is. According to the relevant content in Section 4.3, when the deviation between the weights and the highly trustworthy population is within the margin of error τ, the worker's DOT will be increased; on the contrary, the worker's DOT will be decreased.

As workers' DOT are iterated, the system will distinguish between the highly trustworthy and untrustworthy populations. The data submitted by the untrustworthy population, as defined in Eq. (19), will not be used (i.e. $k_i = 0$) in the calculation of the ETD. Thereby:

(1) The greater the deviation between the data submitted by the worker s_i and the ETD $\bar{x}$, the lower the worker weighting.
(2) The lower the weight of the worker s_i, the lower the DOT will be, and when the DOT is below the threshold ξ_1, the worker s_i is considered as an untrustworthy population and the data submitted by him/her will not be used in the calculation of the ETD (i.e. $k_i = 0$).

According to Eq. (25) $\bar{x} = \frac{\sum_{i=1}^{n}(w_i \times k_i \times x_i)}{\sum_{i=1}^{n}(w_i \times k_i)}$, the weight of the data submitted by untrustworthy workers will be reduced to a minimum value of 0, and the authenticity of the ETD $\bar{x}$ will be effectively increased.

Meanwhile, the data collection cost of the cycle is calculated as shown in Eq. (3) $\mathscr{C} = \mathfrak{C}\left(\sum_{t=1}^{n}\left(\sum_{i=1}^{m} k_{i,t}\right) \times p_t\right)$, and the overall cost of the system will be effectively reduced since the data submitted by the untrustworthy worker s_i will not be used in the calculation of the ETD (i.e. $k_i = 0$), and the corresponding incentives of the workers will be canceled.

Theorem 2. *PPAT mechanism that effectively protects the privacy of the sensory data submitted by the worker, the worker's DOT, and the ETD calculated by the system.*

Proof. Confidentiality of the system environment: throughout the process of sensing-based data submission and calculation, the data transmission between the sensing platform and the workers is carried out through an encrypted channel, so that entities external to the system do not have access to the worker's sensing data and DOT and other relevant information.

(1) We first demonstrate the security of worker sensed data.

In the data submission and truth value calculation phase, workers obtain the corresponding sensed values and then scramble them with their own Individual Perturbation Values (IPV); after receiving the scrambled data from all workers, DPC processes them using the Aggregated Descrambling Values (ADV), and obtains the sum of data from all workers. The IPV of a single worker is not shared with DPC. Therefore, DPC cannot obtain the real sensing data of individual workers.

Although RKC has the IPV of the worker, DPC only transmits for it the scrambled deviation of the worker's sensing data from the ETD, and the ETD is only retained in DPC, and will not be divulged to RKC, therefore, RKC cannot obtain the real sensing data of the worker through the IPV.

(2) Second, we demonstrate the security of worker's DOT.

RKC will use random masks to scramble the DOT of the workers and will not tell the workers about the random masks they generate, so the workers will not be able to get information about their true DOT.

Although DPC calculates the scrambled deviation between the worker's sensed data and the ETD and transmits the scrambled deviation to RKC, this scrambled deviation requires the IPV of each worker to decrypt, and DPC is unable to obtain the true deviation data between each worker and the ETD through the IPV, and is unable to derive the DOT of each worker accordingly.

(3) Finally, we demonstrate the security of the system's ETD.

DPC receives the scrambled data from all the workers and processes it using the ADV to obtain the sum of the data from all the workers and calculates the ETD accordingly, which is retained only in DPC, and the workers and RKC do not have the opportunity to individually access the ETD calculated by DPC.

At the same time, DPC only transmits the scrambled deviation of the worker's sensory data from the ETD for RKC, and RKC does not have access to the worker's real sensory data, which makes it impossible to derive the ETD of the system.

In summary, the worker's sensory data, the DOT of each worker, and the ETD of the system can be protected in the PPAT mechanism, which is used in this paper to realize the privacy protection of all three.

4.6.2. Fault tolerance analysis

The more users are involved, the greater the likelihood that a participating user in the truth-value discovery process will drop out at any given moment due to the limited energy or network conditions of the smart device. In this case, PPAT utilizes a design mechanism to ensure the correctness of the truth value computation of the aggregation results.

First, during the encrypted data collection phase, each participating user exports his/her own IPV, and for various reasons, only some users in the subset succeed in submitting their personal data in time. In order to eliminate the mask introduced by the exiting users into the sum aggregation, the DPC server sends a list of users who have received the

data and requests the ADV from the RKC. The server RKC constructs the corresponding ADV from the received list of users and then uses it to accurately compute the masks introduced into the sum aggregation by the exiting users. As a result, the server DPC successfully aggregates the sum of input data from online participating users.

Second, during the truth value computation process, as seen in Algorithm 3, especially Eq. (25), the key challenge for the reliability of the truth value computation in mobile scenarios is that the set of users in the weighted aggregation $\sum_{i=1}^{n}(w_i \times x_i)$ is guaranteed to be the same as the set of users in the user-weighted aggregation $\sum_{i=1}^{n} w_i$. To solve this problem, the PPAT scheme combines the weighted data values and the weight values of each user into a vector, and obtains both sums $\sum_{i=1}^{n}(w_i \times x_i)$ and $\sum_{i=1}^{n} w_i$ simultaneously in one aggregation operation through a secure aggregation protocol that supports variable dimensional vector aggregation. The truth value can be calculated correctly according to Eq. (25). Therefore, the withdrawal of some of the participating users also has no effect on the correctness of the final truth value calculation.

In summary, the PPAT scheme is robust to the errors caused by the withdrawal of participating users at any time, and can guarantee the correctness of the aggregation results and the truth value calculation of the sensed data streams.

5. Performance analysis

In this section, the proposed truth discovery mechanism based on privacy preservation and DOT evaluation is experimentally validated. The interaction mechanism in this paper is mainly divided into two modules, one is the interaction between the data requester and the platform, and the other is the interaction between the platform and the workers. Two algorithms are proposed for this mechanism. The first one is the MEAN calculation method [47], which is based on the calculation of the mean value of all the data involved in the truth value calculation, and the result of the calculation is directly used as the ETD; the second one is the CRH method [33], which is based on the calculation method in Section 4.1.

5.1. Experiment setup

This section details the experimental environment setup used for the experiment. This article conducts simulation experiments on the PyCharm software and MATLAB platform, using the Python language to construct data collection scenarios in PyCharm and simulate the trust evolution process on the MATLAB platform. And the hardware platform is at Intel i7–1165G7 CPU@2.80 GHz, 16 GB DDR4, 1 TB SSD hard disk on Windows 11. In the experiment, it consists of three parties: data requesters, platforms and workers. We set the number of workers to 1000, the platform defaults to 2, the workers are initially set to 50, and the default iteration is 100 rounds, i.e., for the current population, the task is issued 100 times by default. Workers will gradually reduce the required number of workers after the system iteration. In each round of tasks, initially, the platform first releases tasks and workers apply for them. After receiving the worker's application, the platform decides whether to confirm the application. If rejected, the worker does not participate in the current round of tasks, and once confirmed becomes a participant of the task.

Workers have three categories of identity according to their DOT, which are high trustworthy crowd, trustworthiness unknown crowd and untrustworthy crowd. In the experiment, workers whose trustworthiness is higher than the trustworthiness comparison threshold ξ_2 are set as the high trustworthy population, those whose DOT is between the trustworthiness comparison threshold ξ_2 and the lower limit of trustworthiness ξ_1 are considered as the trustworthiness unknown population, and those whose DOT is lower than the lower limit of trustworthiness ξ_1 are considered as the untrustworthy population. In the task, these three populations submitted data of different quality and are considered to be associated with DOT between [0,1]. In the experiment, the quality of the data provided by all workers is considered to follow a normal distribution, and the overall quality of the data was characterized by the level of the mean of the normal distribution, with groups with a high mean having a high overall quality of data and groups with a low mean having a low overall quality of data. The DOT of each worker is not shared between platforms. The platforms select 0 or some or all of the workers to participate in the truth estimation operation based on the current requirement. The requirements are as follows: if all the workers enlisted under the current task call are untrustworthy people, the current round is nullified and no truth value estimation is performed but DOT updating is done; if all the workers enlisted under the current task call are not all untrustworthy people, these people are excluded from the task participation, and the selection is made from the remaining people. If in the pre-training phase, all data are selected to participate in the truth estimation and rewards are given to the workers who participate in the truth estimation to give DOT updates to all workers. If in the late iteration phase, all data provided by workers with DOT above the trustworthiness comparison threshold ξ_2 are selected to participate in the truth value estimation and rewards are given to the workers who participated in the truth value estimation and DOT updates are given to all workers.

In this experiment, the algorithm proposed in this paper requires updating the worker's DOT by comparing it with the baseline value, so as to iterate on the worker's DOT and complete the categorization and differentiation of the three categories of identities. The specific method is as follows: after each worker submits data, the platform will judge what kind of update to make according to the conditions. The specific rules are as follows: after each worker submits data, the platform will compare the worker's weight with the benchmark value. If the absolute value of the difference between the workers' weight in the set and the benchmark value is less than or equal to the estimated truth deviation γ, the DOT of the workers will be increased; if the absolute value of the difference between the workers' weight in the set and the benchmark value is greater than the estimated truth deviation γ, the DOT of the workers will be decreased. The estimated truth deviation γ is an adjustable parameter.

After receiving all the data submitted by workers, the platform processes them and then updates the DOT. The platform calculates the worker's weight based on the quality of the data submitted by the worker, and then uses the weight as an indicator of the quality of the data submitted by the worker, and multiplies the trust adjustment step λ by the weight, and determines the update range of the worker's trustworthiness based on the quality of the data. If the worker's own weight is lower, but still the data error is within the allowable range, then its adjustment step is smaller compared to the higher weight, i.e., the DOT rises slower; if the data error is not within the allowable range, then its adjustment step is larger, i.e., the DOT declines faster, and the trust level adjustment step λ is an adjustable parameter.

5.2. Algorithms for comparison

This paper compares four truth estimation algorithms.

(1) MEAN [47]: data from all workers are sent to the platform in a uniform way, then the mean value of the data is calculated, and the mean value is taken as the ETD.
(2) CRH [33]: After collecting the data submitted by all workers, the ETD is calculated with reference to the equation in Section 4.1.
(3) PPAT MEAN: the PPAT strategy is added to the mean method to include trust computing (the ETD are calculated according to Eq. (11)).
(4) PPAT CRH: The PPAT strategy is added to the CRH method to include trust computing (the ETD and weights are calculated according to Eqs (15) and (25), respectively)

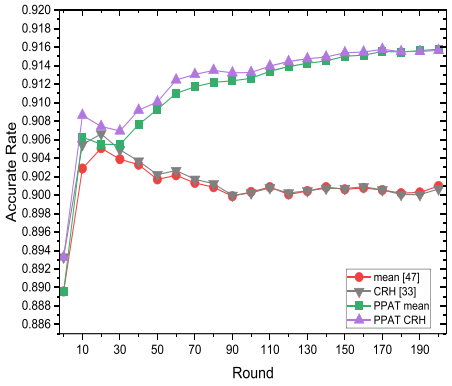


Fig. 7. Comparison of the average accuracy of the ETD of the four algorithms (iterated from initial).

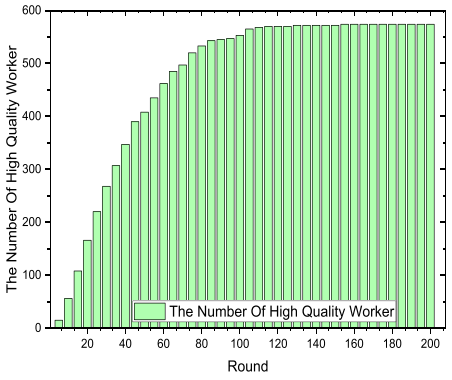


Fig. 8. Trend of the number of highly trustworthy workers after iterative training.

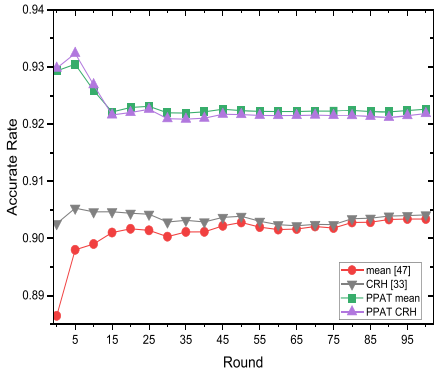


Fig. 9. Comparison of the average accuracy of the ETD of the four algorithms (after iterative stabilization).

5.3. Overview of experiments results

First, we use the initial experimental parameter values, i.e., the default quality of data provided by all workers follows a normal distribution with a mean μ of 0.9 and a variance δ of 0.1. Also, the default trustworthiness comparison threshold ξ_2 is 0.85, the trustworthiness adjustment step λ is 0.1, the estimated truth deviation γ is 0.1, and the lower bound of the trust level is 0.3. The results after conducting 200 rounds of iterative experiments are shown in Fig. 7. From the figure, it can be seen that the results of both PPAT MEAN algorithm and PPAT CRH are unstable at the very beginning training phase, which is due to the fact that the DOT of all workers starts from 0.5 at the very beginning of the training, so the algorithms' advantages are not very obvious. And

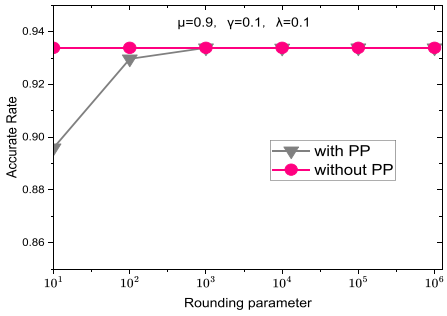


Fig. 10. Truth value accuracy (different rounding factors).

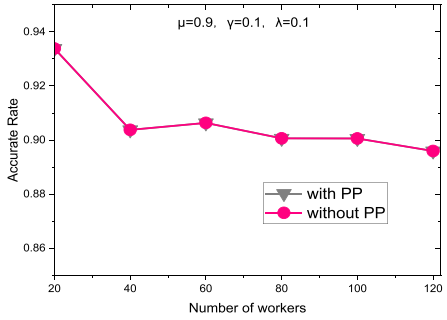


Fig. 11. Truth value accuracy (different number of workers, rounding factor = 10^5).

in the later iteration stage, it can be seen that the PPAT MEAN algorithm and PPAT CRH algorithm are obviously better than the MEAN algorithm and CRH algorithm. Fig. 8 shows the trend of highly trustworthy workers in the iterative training process, and it can be seen that as the number of iteration rounds increases, the highly trustworthy workers also increase gradually, and finally stabilize after the iteration is carried out to about 120 rounds. Fig. 9 then shows that after the iterations are stabilized, the PPAT MEAN algorithm and the PPAT CRH algorithm are significantly better than the MEAN algorithm and the CRH algorithm from the very beginning due to having a sufficient number of highly trustworthy workers.

5.4. The accuracy of data comparison

According to the fault-tolerance analysis in Section 4.6.2, the withdrawal of participating users has no effect on the accuracy of the truth-value computation. The only factor that affects the accuracy of the final result is the rounding factor for converting from the plaintext domain to the ciphertext domain [48].

The experiment first verifies that the PPAT scheme design does not affect the accuracy of the plaintext domain truth value calculation experiment. The number of workers is set to 20 and the rounding factor is varied from 10^1 to 10^6. Fig.10 shows that unless the rounding factor parameter is too small (e.g., 10^1), the truth computation error of the PPAT scheme is almost the same as that of the plaintext domain algorithm.

When the rounding factor is set to 10^5, and different numbers of participating users are set in the experiments, the experimental results shown in Fig.11 further support the conclusion that the truth values computed by the PPAT scheme have the same accuracy as the plaintext-domain truth value computation regardless of the number of workers.

The main parameters of the PPAT algorithm that affect the accuracy of the truth estimate are as follows:

(1) The statistical mean μ of the sensed

Table 5
Simulation settings.

Parameter name	Values
Mean value μ	1.0, 0.9, 0.8, 0.75, 0.7
deviation to ETD γ	01, 0.15, 0.2, 0.25, 0.3
trust level adjustment steps λ	0.08, 0.1, 0.15, 0.3, 0.5
untrustworthy threshold ξ_1	0.5, 0.4, 0.3, 0.2, 0.1
trustworthy threshold ξ_2	0.95, 0.85, 0.8, 0.75, 0.7

values provided by each worker: the size of the mean μ is closely related to the quality of the data provided by the worker, with the smaller the mean μ the lower the quality.

(2) Comparison thresholds in the DOT update algorithm:

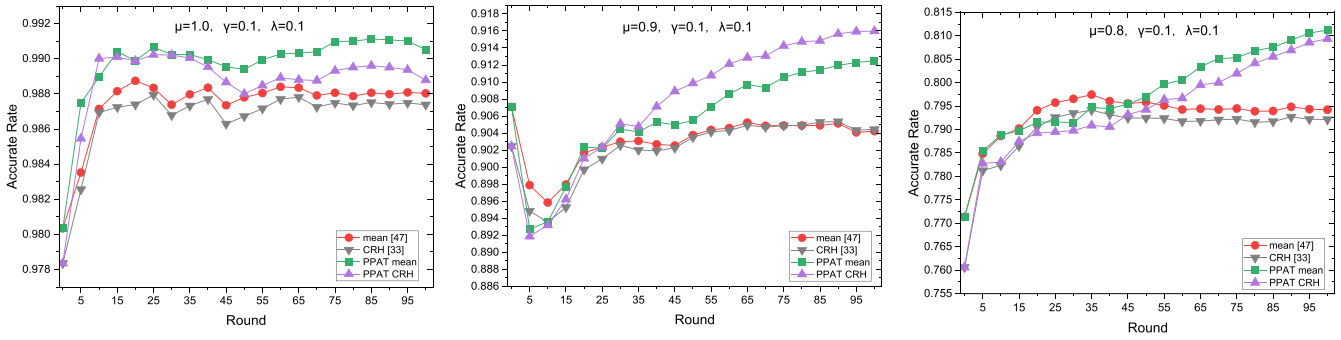
(a) γ: DOT is increased when the error between the worker's weight and the weight of highly trustworthy worker is less than this.

(b) λ: Percentage of steps to increase or decrease DOT.

(3) Several comparison thresholds for worker selection:

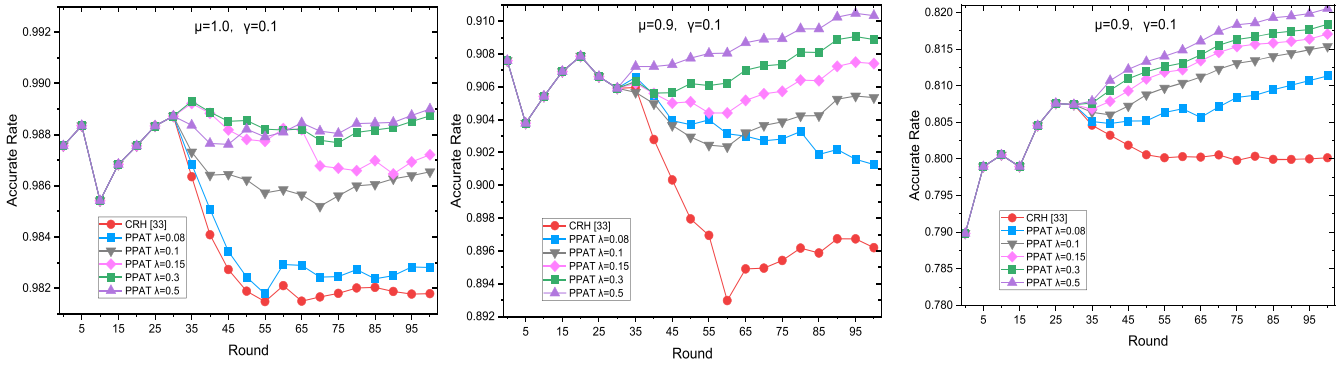
(a) ξ_2:Higher than this DOT, the worker will be included in the highly trustworthy crowd.

(b) ξ_1:Below than this DOT, the worker will be listed as untrust-



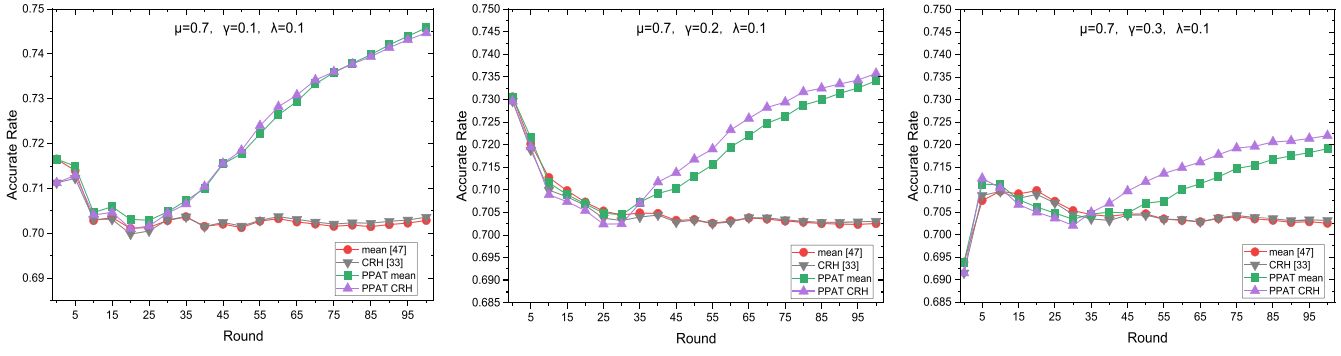
(a) μ=1.0, γ=0.1, λ=0.1 (b) μ=0.9, γ=0.1, λ=0.1 (c) μ=0.8, γ=0.1, λ=0.1

Fig. 12. Comparison of the average accuracy of the ETD of the four algorithms in an experimental setting where the mean is changed and the rest of the variables are the same.



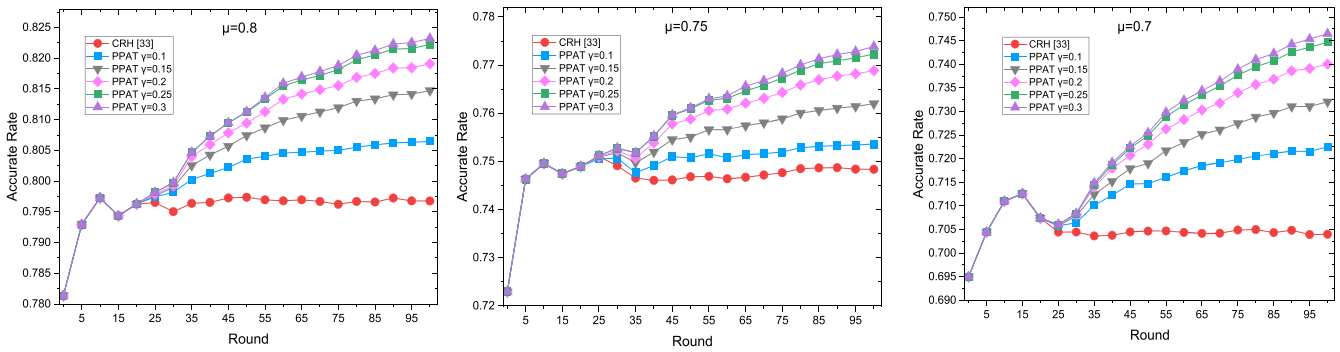
(a) μ=1.0, γ=0.1 (b) μ=0.9, γ=0.1 (c) μ=0.8, γ=0.1

Fig. 13. Comparison of the average accuracy of the ETD of different λ values in an experimental setting where the mean is changed and the rest of the variables are the same.



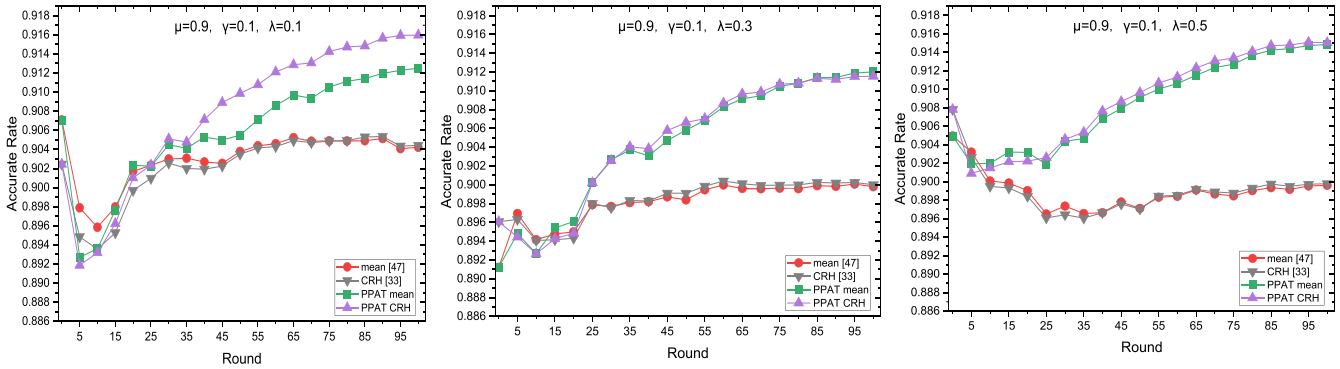
(a) μ=0.7, γ=0.1, λ=0.1 (b) μ=0.7, γ=0.2, λ=0.1 (c) μ=0.7, γ=0.3, λ=0.1

Fig. 14. Comparison of the average accuracy of the four algorithms in estimating the truth value in an experimental setting where the bias of the ETD is changed and the rest of the variables are the same.



(a) μ=0.8 (b) μ=0.75 (c) μ=0.7

Fig. 15. Comparison of the average accuracy of the ETD of different γ values in an experimental setting where the mean is changed and the rest of the variables are the same.



(a) μ=0.9, γ=0.1, λ=0.1 (b) μ=0.9, γ=0.1, λ=0.3 (c) μ=0.9, γ=0.1, λ=0.5

Fig. 16. Comparison of the average accuracy of the ETD of the four algorithms in an experimental setting where the trustworthiness adjustment step is changed and the rest of the variables are the same.

worthy workers and will no longer accept submissions from them.

For the above parameters, design the corresponding values (Table 5):

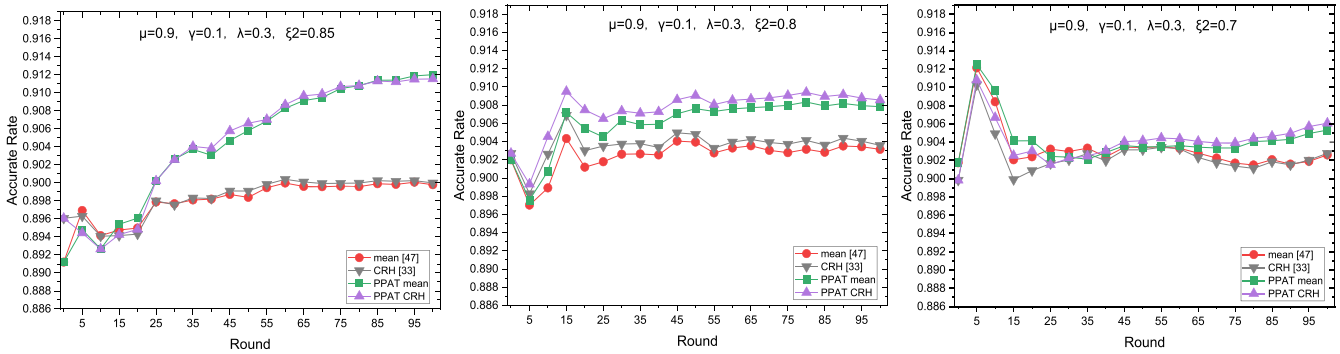
The experiment and its comparative results are given as follows:

(1) The effect of the accuracy of the mean μ and the step λ on the quality of the ETD

Accuracy of mean μ: The smaller the mean value proves that the data quality is more unstable, and estimating a higher quality DOT is more difficult. Fig.12 (a)~(c) show the accuracy comparison of the four algorithms under different mean value results. It can be seen that no matter how the mean value changes, the performance of the two strategies proposed in this paper is ultimately better than the MEAN algorithm and the CRH algorithm, so it proves the excellence and robustness of the strategy in this paper. Meanwhile, the trend in the figure shows that the advantage of this paper's strategy is more obvious as the mean value decreases and the data quality decreases.

Fig.13 (a)~(c) show the accuracy comparison of the different λ values under different mean value results. It can be seen that the larger the DOT adjustment step λ, the faster the DOT is updated and the more rapidly it converges. The truth values estimated by the strategy in this paper have higher quality. So, it can be seen that the faster the DOT converges, the faster the strategy of this paper stabilizes.

(2) Impact of estimated true value bias γ on the ETD quality



(a) μ=0.9, γ=0.1, λ=0.3, ξ_2=0.85 (b) μ=0.9, γ=0.1, λ=0.3, ξ_2=0.8 (c) μ=0.9, γ=0.1, λ=0.3, ξ_2=0.7

Fig. 17. Comparison of the average accuracy of the ETD of the four algorithms in an experimental setting where the trustworthiness comparison threshold is changed and the rest of the variables are the same.

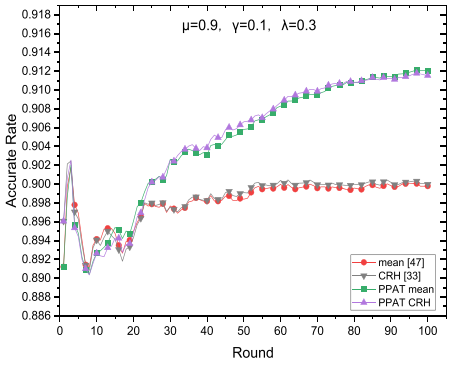


Fig. 18. Comparison of average accuracy of ETD under four algorithms without complicity.

Estimated truth deviation γ: the smaller the estimated truth deviation, the less tolerance for the deviation of the data submitted by the worker from the baseline value, and the higher the quality requirement for the data submitted by the worker. Fig. 14 (a)~(c) show the accuracy comparison of the four algorithms under different estimated truth deviation. It can be seen that no matter how the estimation truth deviation varies, the performance of the two strategies proposed in this paper is ultimately better than the MEAN algorithm and the CRH algorithm. Fig. 15 (a)~(c) show the accuracy comparison of CRH and PPAT under different estimated truth deviation. It can be seen that no matter how the estimation truth deviation varies, the performance of PPAT is ultimately better than CRH. The larger the deviation value to ETD γ, the faster the convergence speed of iteration, the higher the accuracy rate. So, it proves the excellence and robustness of the strategy in this paper.

(3) Effect of trustworthiness adjustment step λ on stability of results (how many rounds to stabilize)

The larger the trustworthiness adjustment step, the faster the trustworthiness update speed and the more rapid the convergence. Fig. 16 (a)~(c) demonstrate the accuracy comparison of the four algorithms under different trust degree adjustment steps. It can be seen that no matter how the trustworthiness adjustment step changes, the performance of the two strategies proposed in this paper is ultimately better than the MEAN algorithm and the CRH algorithm, so it proves that the strategy of this paper is excellent and robust.

(4) The effect of trustworthiness comparison threshold ξ_2 on the quality of ETD

The larger the trustworthiness comparison threshold ξ_2, the higher the requirement for worker trustworthiness and the higher the data quality. Fig. 17 (a)~(c) demonstrate the accuracy comparison of the four algorithms under different trustworthiness contrast thresholds. It can be seen that no matter how the trustworthiness comparison threshold changes, the performance of the two strategies proposed in this paper is ultimately better than the MEAN algorithm and the CRH algorithm, so it proves that the strategy of this paper is excellent and robust.

(5) Comparison and rationale under malicious conspiracy attack

If complicity occurs during the initial training phase, there is a certain probability that this will cause the trustworthiness of the complicit population to rise briefly, at which point the accuracy will decrease. However, when the training continues subsequently, the portion of the population that does not conspire is trained normally and will be judged as untrustworthy because the quality of the submitted data is too low, resulting in lower DOT in subsequent iterations, thus counteracting some of the effects of the conspiracy. If the conspiracy occurs in the subsequent iteration stage, then it must be the less trustworthy people who conspire, at this time, this part of the population does not participate in the estimation of the truth value, but the DOT will be further reduced, and thus at this time, it can also be resisted against the conspiracy.

Fig. 18 demonstrates the ETD accuracy of the four algorithms in the absence of conspiracy. Fig.19-20 demonstrate that when the last 10 rounds produce conspiracies, the strategy proposed in this paper still maintains a high ETD accuracy, while the ETD accuracy of the traditional algorithms decreases faster. This demonstrates that the strategy in this paper has a better defense against conspiracy. The effectiveness and accuracy of this paper's strategy is demonstrated.

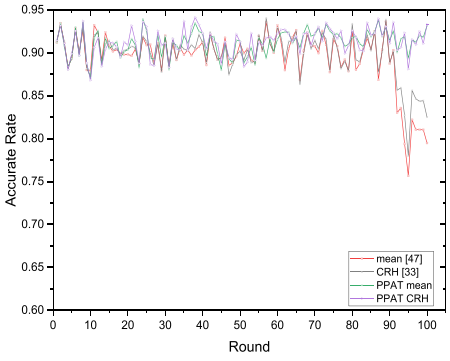


Fig. 19. . Comparison of the accuracy of the four algorithms in estimating the truth value per round under the last 10 rounds of conspiracy.

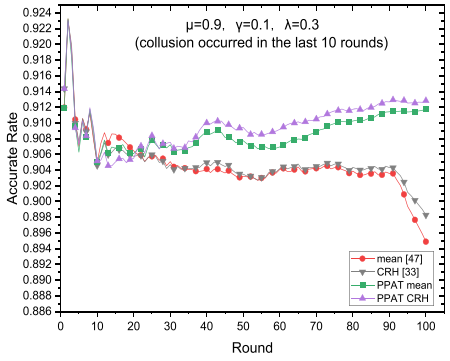


Fig. 20. Comparison of the accuracy of the four algorithms in estimating the truth value per round under the last 10 rounds.

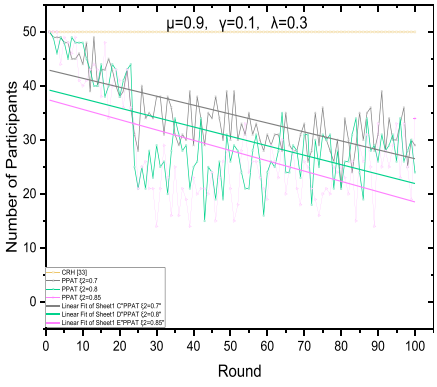


Fig. 21. Comparison of system cost per round for PPAT CRH algorithm with different trustworthiness comparison thresholds ξ_2.

5.5. *The data collection cost comparison*

(1) The state of the system from initialization to Iterative basic convergence

In the strategy proposed in this paper, in the later iteration stage, only the workers whose DOT is higher than the trustworthiness

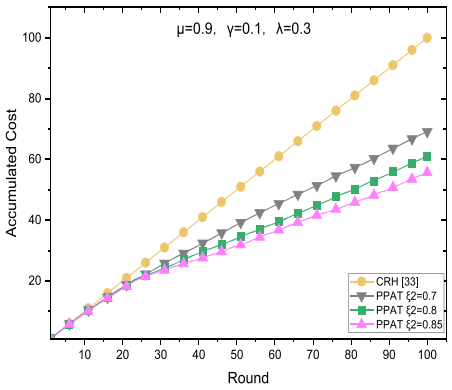


Fig. 22. Comparison of current total system cost of PPAT CRH algorithm with different trustworthiness comparison thresholds ξ_2.

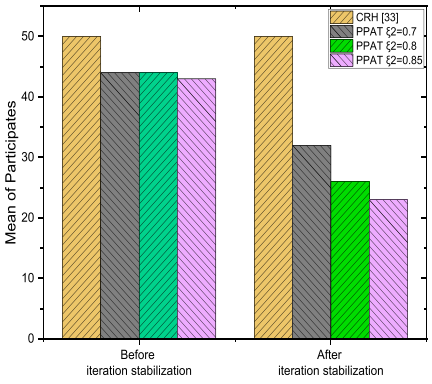


Fig. 23. Comparison of average system cost of PPAT CRH algorithm before and after trustworthiness stabilization under different trustworthiness comparison thresholds ξ_2.

comparison threshold is selected to participate in the truth estimation session, so it is only necessary to give incentives to these selected workers to participate in the truth estimation at this time. According to the accuracy comparison chart listed in Fig. 17 "The effect of trustworthiness comparison threshold on the quality of the ETD", the cost in various cases is shown in Fig. 21-23. It can be seen that under the premise of each threshold, the number of people used in the strategy proposed in this paper is about 20, while at this time the MEAN algorithm and the CRH algorithm still need 50 people, which proves that the cost of the strategy proposed in this paper can be greatly reduced, and it has a good saving.

The per-round overhead statistics of the strategy proposed in this paper under different trustworthiness comparison thresholds ξ_2 are shown in Fig. 21, while the CRH algorithm is added as a reference. In addition to the raw data, trend lines are added to aid in understanding. As can be seen from the figure, the overhead of the system is getting smaller and smaller as the trustworthiness comparison threshold ξ_2 is raised and the iteration proceeds. And it can be significantly seen that the system overhead of this paper's strategy has been smaller than that of the CRH algorithm, demonstrating the savings of this paper's strategy.

The current total overhead statistics of this paper's proposed strategy under different trustworthiness comparison thresholds ξ_2 are shown in Fig. 22, while the CRH algorithm is added as a reference. As can be seen from the figure, the growth of the current total overhead of the system for this paper's strategy decreases significantly with the increase of the trustworthiness comparison threshold ξ_2 and is significantly less than that of the CRH algorithm, demonstrating the frugality of paper's strategy.

In Fig. 23, the overhead statistics generated by the proposed strategy

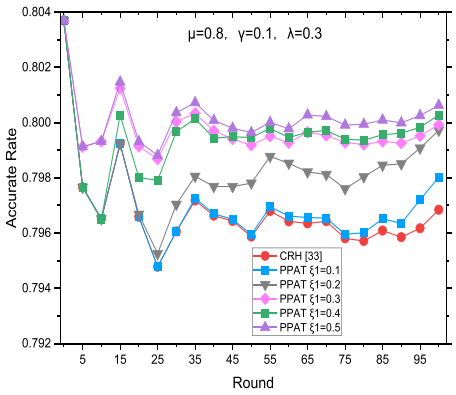


Fig. 24. Comparison of the average accuracy of the ETD for PPAT CRH algorithm with different trustworthiness comparison thresholds ξ_1.

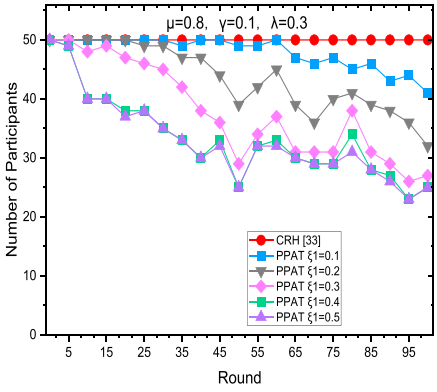


Fig. 25. Comparison of system cost per round for PPAT CRH algorithm with different trustworthiness comparison thresholds ξ_1.

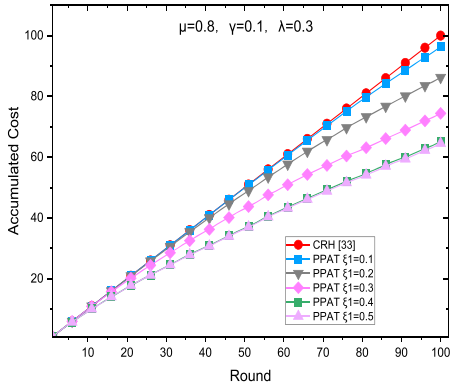
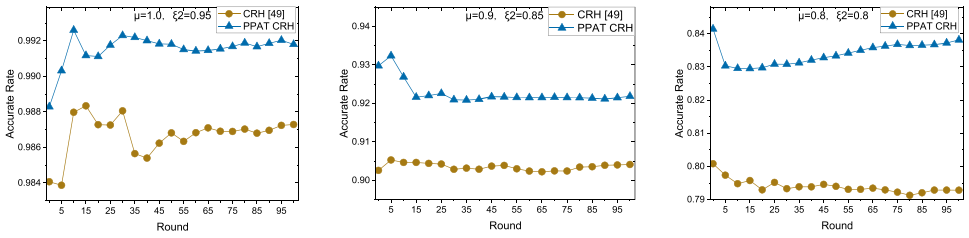


Fig. 26. Comparison of current total system cost of PPAT CRH algorithm with different trustworthiness comparison thresholds ξ_1.

in this paper before and after the DOT is stabilized under different trustworthiness comparison thresholds ξ_2 are shown, while the CRH algorithm is added as a reference. As can be seen from the figure, as the trustworthiness comparison threshold ξ_2 increases, the system overhead incurred after the DOT is stabilized gradually decreases and is smaller than the CRH algorithm regardless of whether the DOT is stabilized or not. This demonstrates the savings of this paper's strategy, and also proves that as the DOT continues to iterate, the system overhead also continues to decrease, reflecting the effectiveness of this paper's strategy.

Fig. 24-26 show the experimental results under different thresholds ξ_1. Fig. 24 shows the accuracy of the estimated truth value of the system



(a) μ=1.0,ξ_2=0.95 (b) μ=0.9,ξ_2=0.85 (c) μ=0.8,ξ_2=0.75

Fig. 27. Comparison of the average accuracy of the ETD algorithms of the CRH and PPAT CRH algorithms after iterative stabilization.

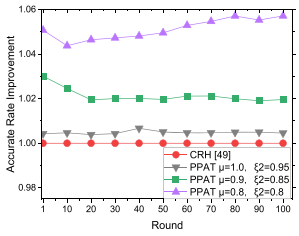


Fig. 28. Comparison of accuracy improvement between CRH algorithm and PPAT algorithm after iterative stabilization.

before iteration stabilization for different thresholds ξ_1, the larger the threshold ξ_1 is, the more data of untrustworthy users are excluded, the higher the accuracy of the system estimation of the truth value is; Fig. 25 shows the cost of the system in each round before iteration stabilization for different thresholds ξ_1, the larger the threshold ξ_1 is, the less the system cost is. Fig. 26 shows the total system cost before iteration stabilization for different thresholds ξ_1, the larger the threshold ξ_1 is, the more data of untrustworthy users are excluded, the lower the total system cost is.

(2) After the system iteration reaches stabilization

After the system has gone through several iterations and reached a steady state, experiments are conducted on CRH and PPAT under three different conditions such as $\mu = 1.0, \xi_2 = 0.95$; $\mu = 0.9, \xi_2 = 0.85$ and $\mu = 0.8, \xi_2 = 0.75$. The results are shown in Figs. 27-30, which clearly demonstrate the comparison of the two mechanisms in terms of the average accuracy in estimating the truth value, the number of workers per round, and the average overhead of the system.

From the figure, it can be seen that the PPAT mechanism, in terms of accuracy, PPAT shows a strong improvement; at the same time, the number of workers in PPAT shows a simultaneous reduction trend. It is fully shown that the PPAT mechanism can effectively reduce the number of workers while improving the accuracy, which improves the efficiency of the system and saves the cost of the system.

This fully reflects the effectiveness of the strategy proposed in this paper, and the PPAT mechanism achieves a good balance of accuracy and cost by optimizing the algorithm and the worker selection mechanism, proving the superiority of the PPAT mechanism in practical applications.

5.6. Comparison with the recently published studies

According to the above experimental results, it can be seen that after adding the trust index, PPAT significantly improves the accuracy of estimating truth values for both CRH and Mean algorithms. Truth estimation in Mobile Crowd Sensing network has always been a popular field, and three are some recently published studies, i.e. MESTD [49] and SinNA [50].

To compare the effectiveness of the truth value algorithm, we conducted relevant experiments, and the results are shown in Fig. 31. From the results, it can be seen that both MESTD and SinNA have certain improvements on the basis of the original CRH, but our algorithm has higher accuracy in estimating truth values and has obvious advantages under various mean values.

6. Conclusion and the future work

In this paper, we focus on the problem of privacy of worker data in crowd sensing, and cost reduction, while considering worker conclusion, we proposed a Privacy Preserving, Accuracy, and Trust data collection scheme (PPAT) for MCS, which can protect the privacy of data content and maintain high accuracy in a low-cost way. It is a more robust variant

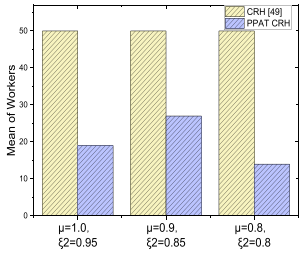
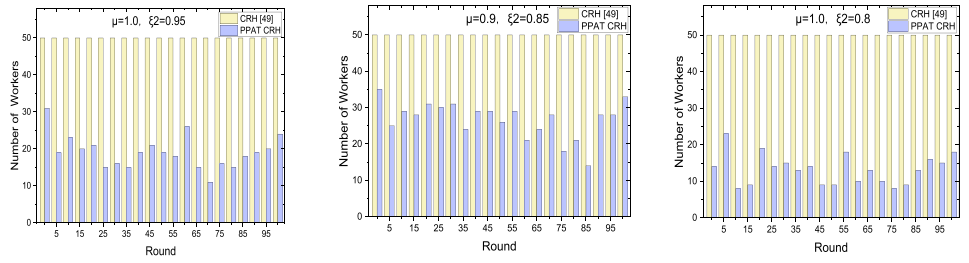
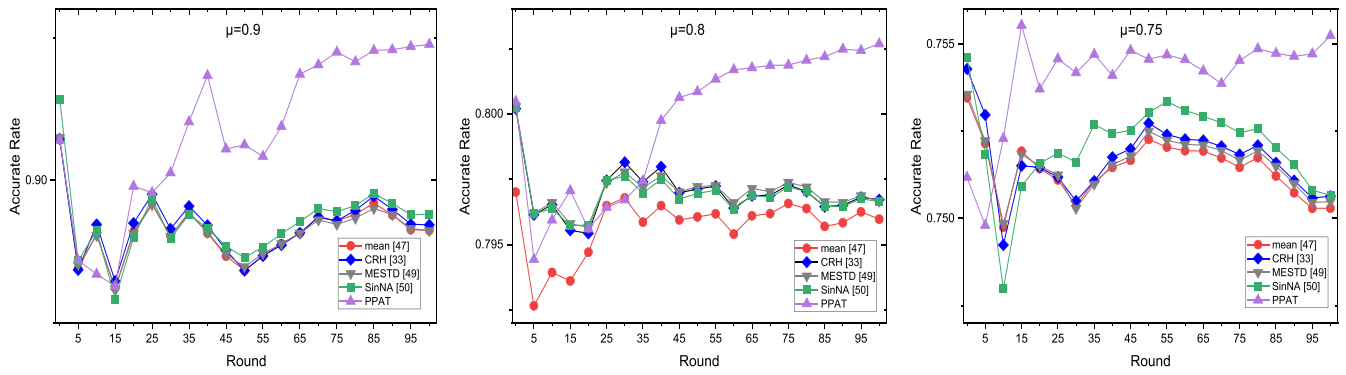


Fig. 30. Comparison of average overhead between CRH algorithm and PPAT algorithm after iteration stabilization.



(a) μ=1.0,ξ_2=0.95 (b) μ=0.9,ξ_2=0.85 (c) μ=0.8,ξ_2=0.75

Fig. 29. Comparison of the number of workers per round between the CRH algorithm and the PPAT CRH algorithm after iterative stabilization.



(a) μ=0.9 (b) μ=0.8 (c) μ=0.75

Fig. 31. Comparison of the average accuracy of the several algorithms in estimating the truth value in an experimental setting where mean value is changed and the rest of the variables are the same.

of the existing truth discovery mechanism based on privacy preservation and DOT evaluation. We add a privacy-preserving mechanism to the truth-value discovery process to protect the privacy of the task participants. Meanwhile, this paper introduces a worker selection and DOT iterative updating mechanism, which introduces the DOT and effectively improves the accuracy of truth value estimation, as well as reduces the cost in group intelligence perception.

Finally, we performed extensive simulations on the dataset to demonstrate the effectiveness of our mechanism. The results show that our algorithm significantly outperforms other algorithms in terms of the ETD accuracy, and system overhead. The advantage is especially significant when malicious workers are in the majority, and when worker complicity occurs.

In the future, we are committed to researching more effective truth discovery mechanisms for MCS. At the same time, we are also committed to combining our truth discovery mechanism with worker selection based on the balance of exploration and exploitation, especially considering the application scenario of malicious conspiracy. We will conduct further in-depth research on schemes to enhance data privacy by combining the advantages of DP and other privacy-preserving schemes while guaranteeing the accuracy of truth value discovery. we also consider other methods that can be used to further combat data poisoning, such as clustering algorithms and machine learning. Combining federated analysis methods, extending our mechanism to distributed systems will also be the direction of further research in the future. We hope that the worker selection and truth discovery obtained by the PPAT mechanism can be effectively fed back to the data platform to help improve the accuracy of data collection and reduce the platform cost.

CRediT authorship contribution statement

Qianxue Guo: Writing – original draft, Software, Methodology, Investigation, Formal analysis, Conceptualization. **Yasha He:** Visualization, Software, Investigation, Data curation. **Qian Li:** Validation, Resources, Investigation. **Anfeng Liu:** Writing – review & editing, Supervision, Resources, Funding acquisition, Conceptualization. **Neal N. Xiong:** Writing – review & editing, Supervision, Project administration. **Qian He:** Data curation. **Qiang Yang:** Investigation, Formal analysis. **Shaobo Zhang:** Software, Resources.

Declaration of competing interest

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of, the manuscript entitled, “PPAT: An Effective Privacy-Preserving, Accuracy, and Trust Scheme for Worker Selection in Mobile Crowdsensing Networks”.

Data availability

The data that has been used is confidential.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China (62072475).

References

[1] J. Huang, L. Kong, L. Cheng, HN. Dai, M. Qiu, G. Chen, X. Liu, G. Huang, BlocskSense: towards trustworthy mobile crowdsensing via proof-of-data blockchain, IEEe Trans. Mob. Comput. 23 (2) (2024) 1016–1033.
[2] W. Wang, Y. Yang, Z. Yin, K. Dev, X. Zhou, et al., BSIF: Blockchain-based secure, interactive, and fair mobile crowdsensing, IEEE J. Selected Areas Commun 40 (12) (2022) 3452–3469.
[3] SA. Moqurrab, A. Anjum, A. Khan, M. Ahmed, A. Ahmad, G. Jeon, Deep-confidentiality: An IoT-enabled privacy-preserving framework for unstructured big biomedical data, ACM Transac. Internet Technol. (TOIT) 22 (2) (2022) 1–21.
[4] C. Peng, D. He, J. Chen, N. Kumar, MK. Khan, EPRT: an efficient privacy-preserving medical service recommendation and trust discovery scheme for eHealth system, ACM Transac. Internet Technol. (TOIT) 21 (3) (2021) 1–24.
[5] J. Gao, L. Zhao, X. Shen, Network utility maximization based on incentive mechanism for truthful reporting of local information, IEEe Trans. Veh. Technol. 67 (8) (2018) 7523–7537.
[6] R. Zhang, A. Liu, T. Wang, N. Xiong, A. Vasilakos, A trust active and trace back based trust management system about effective data collection for mobile IOT services, Info. Sci 664 (2024) 120329.
[7] B. Yang, A. Liu, N. Xiong, T. Wang, S. Zhang, LC-TDC: a low cost and truth data collection scheme by using missing data imputation in sparse mobile crowdsensing, Info. Sci 662 (2024) 120274.
[8] Y. Liu, Z. Yu, B. Guo, Q. Han, J. Su, et al., CrowdOS: A ubiquitous operating system for crowdsourcing and mobile crowd sensing, IEEe Trans. Mob. Comput. 21 (3) (2022) 878–894.
[9] L. Sarkar, V. Ramasamy, A. Majumder, et al., I-health: SDN-based fog architecture for IIoT applications in healthcare, IEEE/ACM. Trans. Comput. Biol. Bioinform. (2022), https://doi.org/10.1109/TCBB.2022.3193918.
[10] Z. Cai, Z. Duan, W. Li, Exploiting multi-dimensional task diversity in distributed auctions for mobile crowdsensing, IEEe Trans. Mob. Comput. 20 (8) (2021) 2576–2591.
[11] Y. Liu, A. Liu, T. Wang, et al., An intelligent incentive mechanism for coverage of data collection in cognitive Internet of Things, Future Gen. Comp. Sys 100 (2019) 701–714.
[12] X. Gao, G.Chen S.Chen, MAB-based reinforced worker selection framework for budgeted spatial crowdsensing, IEEe Trans. Knowl. Data Eng. 34 (3) (2022) 1303–1316.
[13] M. Karaliopoulos, E. Bakali, Optimizing mobile crowdsensing platforms for boundedly rational users, IEEe Trans. Mob. Comput. 21 (4) (2022) 1305–1318.
[14] J. Tang, K. Fan, P. Yin, Z. Qu, A. Liu, N. Xiong, T. Wang, M. Dong, S. Zhang, DLFTI: a deep learning based fast truth inference mechanism for distributed spatiotemporal data in mobile crowd sensing, Info. Sci 644 (2023) 119245.

[15] D. Sarma, A. Das, P. Dutta, et al., A cost minimization resource allocation model for disaster relief operations with an information crowdsourcing-based mcdm approach, IEEe Trans. Eng. Manage 69 (5) (2022) 2454–2474.
[16] A. Hamrouni, H. Ghazzai, T. Alelyani, et al., Low-complexity recruitment for collaborative mobile crowdsourcing using graph neural networks, IEEe Internet. Things. J. 9 (1) (2021) 813–829.
[17] M. Kadadha, H. Otrok, S. Singh, et al., Two-sided preferences task matching mechanisms for blockchain-based crowdsourcing, J. Netw. Comp. Applic 191 (2021) 103155.
[18] S. née Müller, C. Tekin, M. van der Schaar, et al., Context-aware hierarchical online learning for performance maximization in mobile crowdsourcing, IEEE/ACM Transac. Netw 26 (3) (2018) 1334–1347.
[19] Q. Feng, D. He, M. Luo, X. Huang, KKR. Choo, EPRICE: an efficient and privacy-preserving real-time incentive system for crowdsensing in industrial internet of things, IEEE Transac. Comp 72 (9) (2023) 2482–2495.
[20] J. Guo, G. Huang, Q. Li, N. Xiong, S. Zhang, T. Wang, STMTO: a smart and trust multi-UAV task offloading system, Info. Sci 573 (2021) 519–540.
[21] Y. Wang, Z. Cai, ZH. Zhan, YJ. Gong, et al., An optimization and auction-based incentive mechanism to maximize social welfare for mobile crowdsourcing, IEEe Trans. Comput. Soc. Syst. 6 (3) (2019) 414–429.
[22] J. Tang, F. Han, K. Fan, W. Xie, P. Yin, Z. Qu, A. Liu, S.Zhang N.Xiong, T. Wang, Credit and quality intelligent learning based multi-armed bandit scheme for unknown worker selection in multimedia MCS, Info. Sci 647 (2023) 119444.
[23] G. Ji, Z. Yao, B. Zhang, C. Li, Quality-driven online task-bundling-based incentive mechanism for mobile crowdsensing, IEEe Trans. Veh. Technol. 71 (7) (2022) 7876–7889.
[24] H. Wang, A. Liu, N. Xiong, S. Zhang, T. Wang, TVD-RA: a truthful data value discovery based reverse auction incentive system for MCS, IEEe Internet. Things. J. 11 (4) (2024) 5826–5839.
[25] Z. Wang, J. Li, J. Hu, et al., Towards privacy-driven truthful incentives for mobile crowdsensing under untrusted platform, IEEe Trans. Mob. Comput. 22 (2) (2023) 1198–1212.
[26] C. Huang, H. Yu, R.A. Berry, J. Huang, Using truth detection to incentivize workers in mobile crowdsourcing, IEEe Trans. Mob. Comput. 21 (6) (2022) 2257–2270.
[27] J. Tang, S. Fu, X. Liu, et al., Achieving privacy-preserving and lightweight truth discovery in mobile crowdsensing, IEEe Trans. Knowl. Data Eng. 34 (11) (2022) 5140–5153.
[28] G. Xu, H. Li, S. Liu, et al., Efficient and privacy-preserving truth discovery in mobile crowd sensing systems, IEEe Trans. Veh. Technol. 68 (4) (2019) 3854–3865.
[29] C. Zhang, L. Zhu, C. Xu, et al., Reliable and privacy-preserving truth discovery for mobile crowdsensing systems, IEEe Trans. Depend. Secure Comput. 18 (3) (2019) 1245–1260.
[30] S. Gao, X. Chen, J. Zhu, et al., TrustWorker: A trustworthy and privacy-preserving worker selection scheme for blockchain-based crowdsensing, IEEe Trans. Serv. Comput. 15 (6) (2022) 3577–3590.
[31] R. Ganjavi, AR. Sharafat, Edge-assisted public key homomorphic encryption for preserving privacy in mobile crowdsensing, IEEe Trans. Serv. Comput. 16 (2) (2023) 1107–1117.
[32] L. Ma, X. Liu, Q. Pei, Y. Xiang, Privacy-preserving reputation management for edge computing enhanced mobile crowdsensing, IEEe Trans. Serv. Comput. 12 (5) (2019) 786–799.
[33] Q. Li, Y. Li, J. Gao, B. Zhao, W. Fan, J. Han, Resolving conflicts in heterogeneous data by truth discovery and source reliability estimation, in: Proc. ACM SIGMOD Int. Conf. Manag. Data (SIGMOD/PODS), 2014, pp. 1187–1198.
[34] K. Li, S. Wang, X. Cheng, et al., A misreport-and collusion-proof crowdsourcing mechanism without quality verification, IEEe Trans. Mob. Comput. 21 (9) (2022) 3084–3095.
[35] Y. Zhao, X. Gong, X. Chen, Privacy-preserving incentive mechanisms for truthful data quality in data crowdsourcing, IEEe Trans. Mob. Comput. 21 (7) (2022) 2518–2532.
[36] Y. Cheng, J. Ma, Z. Liu, L. Wang, Z. Ying, X. Chen, Efficient anonymous authentication and privacy-preserving reliability evaluation for mobile crowdsensing in vehicular networks, IEEe Internet. Things. J. 10 (17) (2023) 14925–14939.
[37] HT. Wu, Y. Zheng, B. Zhao, J. Hu, An anonymous reputation management system for mobile crowdsensing based on dual blockchain, IEEe Internet. Things. J. 9 (9) (2021) 6956–6968.
[38] S. Yang, F. Wu, S. Tang, X. Gao, B. Yang, G. Chen, On designing data quality-aware truth estimation and surplus sharing method for mobile crowdsensing, IEEE J. Selec. Areas Commun 35 (4) (2017) 832–847.
[39] H. Tian, W. Sheng, H. Shen, et al., Truth finding by reliability estimation on inconsistent entities for heterogeneous data sets, Knowl. Based. Syst. 187 (2020) 104828.
[40] Y. Zhao, X. Gong, F. Lin, X. Chen, Data poisoning attacks and defenses in dynamic crowdsourcing with online data quality learning, IEEe Trans. Mob. Comput. 22 (5) (2023) 2569–2581.
[41] G. Soatti, S. Savazzi, M. Nicoli, MA. Alvarez, S. Kianoush, et al., Distributed signal processing for dense 5G IoT platforms: Networking, synchronization, interference detection and radio sensing, Ad. Hoc. Netw. 89 (2019) 9–21.
[42] R. Liu, M. Xie, A. Liu, H. Song, Joint Optimization risk factor and energy consumption in IoT networks with TinyML-enabled internet of UAVs, IEEe Internet. Things. J. (2023), https://doi.org/10.1109/JIOT.2023.3348837.
[43] B. Waggoner, Y. Chen, Output agreement mechanisms and common knowledge, in: Proc 2nd AAAI Conf. Human Comput. Crowdsourcing 2, 2014, pp. 220–226.
[44] H. Li, D. Liu, Y. Dai, T.H. Luan, Engineering searchable encryption of mobile cloud networks: When QoE meets QoP, IEEE Wirel. Commun 22 (4) (2015) 74–80.
[45] H. Li, D. Liu, Y. Dai, T. Luan, S. Yu, Personalized search over encrypted data with efficient and secure updates in mobile clouds, IEEE Trans. Emerg. Topics Comput. 6 (1) (2018) 97–109.
[46] X. Li, Y. Zhu, J. Wang, Z. Liu, Y. Liu, M. Zhang, On the soundness and security of privacy-preserving SVM for outsourcing data classification, IEEE Trans. Depend. Secure Comput. 15 (5) (2018) 906–912.
[47] S. Ye, J. Wang, H. Fan, et al., Probabilistic model for truth discovery with mean and median check framework, Knowl. Based. Syst. 233 (2021) 107482.
[48] Y Zheng, H Duan, X Yuan, et al., Privacy-aware and efficient mobile crowdsensing with truth discovery, IEEe Trans. Depend. Secure Comput. 17 (1) (2020) 121–133.
[49] Chenfei Hu, Zihan Li, Yuhua Xu, Chuan Zhang, Ximeng Liu, Daojing He, Liehuang Zhu, Multi-round Efficient and Secure Truth Discovery in Mobile Crowdsensing Systems, IEEe Internet Things. J. 11 (10) (2024) 2327–4662.
[50] Shaohua Zeng, Bin Cai, Xiaohu Li, Zhen Zhang, A truth value discovery algorithm for conflict sensor data in sensor network, in: 2023 35th Chinese Control and Decision Conference (CCDC), 2023, pp. 4435–4440.

Qianxue Guo is currently a student at the School of Computer Science and Engineering, Central South University, China. Her research interests include mobile crowd sensing, reinforcement learning and incentive mechanism. E-mail: 8305210706@csu.edu.cn.

Yasha He is currently a student at the School of Life Sciences, Central South University, China. Her research interest is biostatistics. E-mail: 8305211210@csu.edu.cn.

Qian Li is currently a student at the School of Automation, Central South University, China. Her research interest is embedded AI system. E-mail: 8214210206@csu.edu.cn.

Anfeng Liu received the M.*Sc.* and Ph.D. degrees from Central South University, China, in 2002 and 2005, respectively, both in computer science. He is currently a professor of the School of Information Science and Engineering, Central South University, China. His major research interest is wireless sensor networks, Internet of Things, information security, edge computing and crowdsourcing. Dr. Liu has published 4 books and over 200 international journal and conference papers, among which there are more than 30 ESI highly-cited papers. Some of his works were published in IEEE Transactions on Information Forensics & Security, IEEE Transactions on Mobile Computing, IEEE Transactions on Parallel and Distributed Systems, IEEE Transactions on Vehicular Technology, IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. His research has been supported by the National Basic Research Program of China (973 Program) and the National Natural Science Foundation of China for five times. He was a recipient of the First Prize of Scientific Research Achievement of Colleges from the Ministry of Education of China in 2016, and the Second Prize of Science and Technology Award from China Nonferrous Metal Industry Association in 2005. He has served as the Editor of the IEEE Networking Letters. E-mail: afengliu@mail.csu.edu.cn.

Neal N. Xiong (S'05–M'08–SM'12) is current a Professor, Computer Science Program Chair, at Department of Computer, Mathematical and Physical Sciences, Sul Ross State University, Alpine, TX 79,830, USA. He received his both PhD degrees in Wuhan University (2007, about sensor system engineering), and Japan Advanced Institute of Science and Technology (2008, about dependable communication networks), respectively. Before he attended Sul Ross State University, he worked in Georgia State University, Northeastern State University, and Colorado Technical University (full professor about 5 years) about 15 years. His research interests include Cloud Computing, Security and Dependability, Parallel and Distributed Computing, Networks, AI, and Optimization Theory. Prof. Xiong published over 300 IEEE journal papers and over 200 international conference papers. Some of his works were published in IEEE JSAC, IEEE or ACM transactions, ACM Sigcomm workshop, IEEE INFOCOM, ICDCS, and IPDPS. He has been a General Chair, Program Chair, Publicity Chair, Program Committee member and Organizing Committee member of over 100 international conferences, and as a reviewer of about 100 international journals, including IEEE JSAC, IEEE SMC (Park: A/B/C), IEEE Transactions on Communications, IEEE Transactions on Mobile Computing, IEEE Trans. on Parallel and Distributed Systems. He is serving as an Editor-in-Chief, Associate editor or Editor member for over 10 international journals (including Associate Editor for IEEE Tran. on Systems, Man & Cybernetics: Systems, Associate Editor for IEEE Tran. on Network Science and Engineering, Associate Editor for Information Science, Editor-in-Chief for Journal of Internet Technology (JIT), and Editor-in-Chief for Journal of Parallel & Cloud Computing (PCC)), and a guest editor for over 10 international journals, including Sensor Journal, WINET and MONET. He has received the Best Paper Award in the 10th IEEE International Conference on High Performance Computing and Communications (HPCC-08) and the Best student Paper Award in the 28th North American Fuzzy Information Processing Society Annual Conference (NAFIPS2009). Dr. Xiong is the Chair of "Trusted Cloud Computing" Task Force, IEEE Computational Intelligence Society (CIS), HYPERLINK "http://www.cs.gsu.edu/~cscnxx/index-TF.html", and the Industry System Applications Technical Committee, HYPERLINK "http://ieee-cis.org/technical/isatc/"; He is a Senior member of IEEE Computer Society from 2012. E-mail: xiongnaixue@gmail.com.

Qian He received the Ph.D. degree in computer science from the Beijing University of Posts and Telecommunications, China. He completed postdoctoral research with the National University of Defense Technology, Changsha, China. He worked as a Visiting Scholar with The University of Manchester, U.K., and The Hong Kong Polytechnic University. He is currently a Full Professor with the Guilin University of Electronic Technology and the Deputy Director of the National and Local Joint Engineering Research Center for Satellite Navigation, Positioning and Location Services. His research interests include network security and distribute computing. He is a member of the ACM. He is a Distinguished Member of CCF. E-mail: heqian@guet.edu.cn.

Qiang Yang is currently working toward the PhD degree with the School of Computer Science and Engineering, Central South University, China. His research interests include mobile crowd sensing, reinforcement learning and incentive mechanism. E-mail: yangqiang030@csu.edu.cn.

Shaobo Zhang received the B.*Sc.* and M.*Sc.* degree in computer science both from Hunan University of Science and Technology, Xiangtan, China, in 2003 and 2009 respectively, and the Ph.D. degree in computer science from Central South University, Changsha, China, in 2017. He is currently a Professor at the School of Computer Science and Engineering of the Hunan University of Science and Technology, China. His research interests include edge computing, privacy computing and cyberspace security. E-mail: shaobozhang@hnust.edu.cn.