Computer Modeling in
Engineering & Sciences

Tech Science Press

DOI: 10.32604/cmes.2024.049964

ARTICLE

Blockchain-Assisted Unsupervised Learning Method for Crowdsourcing Reputation Management

Tianyu Wang[1,2] **and Kongyang Chen**[2,3,*]

[1]Department of Cyberspace Security, Guangzhou University, Guangzhou, 51006, China

[2]Institute of Artificial Intelligence, Guangzhou University, Guangzhou, 51006, China

[3]Center of Young Scholars, Pazhou Lab, Guangzhou, 510335, China

*Corresponding Author: Kongyang Chen. Email: kychen@gzhu.edu.cn

Received: 23 January 2024 Accepted: 22 March 2024 Published: 08 July 2024

ABSTRACT

Crowdsourcing holds broad applications in information acquisition and dissemination, yet encounters challenges pertaining to data quality assessment and user reputation management. Reputation mechanisms stand as crucial solutions for appraising and updating participant reputation scores, thereby elevating the quality and dependability of crowdsourced data. However, these mechanisms face several challenges in traditional crowdsourcing systems: 1) platform security lacks robust guarantees and may be susceptible to attacks; 2) there exists a potential for large-scale privacy breaches; and 3) incentive mechanisms relying on reputation scores may encounter issues as reputation updates hinge on task demander evaluations, occasionally lacking a dedicated reputation update module. This paper introduces a reputation update scheme tailored for crowdsourcing, with a focus on proficiently overseeing participant reputations and alleviating the impact of malicious activities on the sensing system. Here, the reputation update scheme is determined by an Empirical Cumulative distribution-based Outlier Detection method (ECOD). Our scheme embraces a blockchain-based crowdsourcing framework utilizing a homomorphic encryption method to ensure data transparency and tamper-resistance. Computation of user reputation scores relies on their behavioral history, actively discouraging undesirable conduct. Additionally, we introduce a dynamic weight incentive mechanism that mirrors alterations in participant reputation, enabling the system to allocate incentives based on user behavior and reputation. Our scheme undergoes evaluation on 11 datasets, revealing substantial enhancements in data credibility for crowdsourcing systems and a reduction in the influence of malicious behavior. This research not only presents a practical solution for crowdsourcing reputation management but also offers valuable insights for future research and applications, holding promise for fostering more reliable and high-quality data collection in crowdsourcing across diverse domains.

KEYWORDS

Crowdsourcing; reputation management; blockchain

1 Introduction

With the rapid development of 5G mobile communication technology and mobile smart devices, crowdsourcing has become an important research direction in a world gradually transitioning into a new era of ubiquitous sensing, connectivity, and intelligence. According to Huawei's "Global Industry Vision (GIV) 2025" forecast, by 2025, the global number of personal smart terminals (such as smartphones, smartwatches, portable computers, etc.) will reach 4 billion, the total number of connected devices worldwide will reach 100 billion, the global annual data generated will reach 1800 exabytes (EB), the adoption rate of AI in enterprises will reach 86%, and the prevalence of intelligent personal assistants will reach 90%. As a result, crowdsourcing systems have gained widespread attention in both industry and academia and have found applications in areas such as medical diagnostics [1], environmental monitoring [2], road surveillance [3], and intelligent vehicular networks [4]. The rapid evolution of 5G mobile communication technology and mobile smart devices has propelled crowdsourcing into a pivotal research domain in a world transitioning towards a new era of pervasive sensing, connectivity, and intelligence. Huawei's "Global Industry Vision (GIV) 2025" anticipates that by 2025, the global count of personal smart terminals (e.g., smartphones, smartwatches, portable computers) will reach 4 billion, worldwide connected devices will total 100 billion, the annual global data generation will hit 1800 exabytes (EB), the adoption rate of AI in enterprises will reach 86%, and intelligent personal assistants will be prevalent in 90% of cases. Consequently, crowdsourcing systems have garnered extensive attention in both industry and academia, finding applications in diverse areas such as medical diagnostics [1], environmental monitoring [2], road surveillance [3], and intelligent vehicular networks [4].

Traditional crowdsourcing systems are typically composed of three integral components: the task demander, sensing platform, and sensing users, and their operational workflow [5] is delineated as follows: 1) the task demander disseminates tasks to the sensing platform; 2) sensing users assess and decide whether to accept the tasks; 3) the sensing platform aggregates task data from enlisted users; and 4) the sensing platform analyzes the data and furnishes it to the task demander. Nevertheless, conventional crowdsourcing systems contend with frequent data transmissions between the platform and users, resulting in network congestion and operational inefficiencies [6]. Additionally, incentive mechanisms [7] have been introduced to motivate users to engage with crowdsourcing systems. Fu et al. [8] presented a task assignment scheme (PCTA-SG) based on employee location privacy protection and an employee elite selection mechanism but lacked a reputation management module to ensure the authenticity of user-uploaded data. Jiang et al. [9] proposed an incentive mechanism protecting user rights using uncertain and hidden bids but could not effectively safeguard task demander interests due to the absence of a reputation management module. Huang et al. [10] researched potential combinations of traditional Automated Passenger Counters (APC) and a novel source capable of collecting detailed mobile demand data but did not include a reputation management module to prevent malicious data uploads. However, these mechanisms face several challenges in traditional crowdsourcing systems: 1) platform security lacks robust guarantees and may be susceptible to attacks [11]; 2) there exists a potential for large-scale privacy breaches [12]; and 3) incentive mechanisms relying on reputation scores may encounter issues as reputation updates hinge on task demander evaluations, occasionally lacking a dedicated reputation update module.

To address the first two challenges, we propose the integration of a blockchain-based crowdsourcing framework. Blockchain, functioning as a decentralized and immutable distributed digital ledger [13,14], emerges as an ideal fit for crowdsourcing systems. The incorporation of smart contracts [15] within the blockchain system enables the efficient execution of diverse crowdsourcing workflow tasks within complex transactional environments [16]. The entire framework operates on smart contracts, obviating the need for trust in third-party entities. Whether pertaining to task publication, submission, reward collection, or distribution, the entire process unfolds with transparency, rigor, and immutability.

In data transactions, identity privacy pertains to the sensitive personal information stored in the blockchain, encompassing the user's actual identity and transaction account addresses (referred to as pseudonyms or nicknames). The disclosure of this information is inconvenient and necessitates confidentiality. When users engage in blockchain services using nicknames or account addresses, frequent input and output operations can enable malicious individuals to link these activities to the user's real identity, resulting in the exposure of sensitive information. While blockchain address generation does not mandate real-name authentication and allows users to create multiple transaction account addresses for added privacy protection, the transparency of all transaction paths permits attackers to trace a user's data based on addresses, analyze correlations, and combine external information to obtain the user's identity and private details. Transaction data privacy pertains to the confidentiality of the transaction content stored within the blockchain, including transaction amounts and participants. The inherent openness and transparency of the blockchain pose a risk to transaction data privacy. For instance, in supply chain scenarios, blockchain technology facilitates the provision of open and transparent information, aiding enterprises in making well-informed decisions swiftly. However, if sensitive data such as cash flows is leaked and seized by competing companies, it can lead to significant losses. Therefore, safeguarding transaction data privacy within the blockchain is of utmost importance. To protect transaction data privacy, cryptographic techniques, such as homomorphic encryption algorithms, searchable encryption algorithms, and attribute-based encryption algorithms, are commonly employed. Thus, in our framework, we leverage the Cheon-Kim-Kim-Song (CKKS) encryption scheme to ensure data security [17]. CKKS is recognized as a leading homomorphic encryption scheme due to its exceptional performance and effectiveness in preserving user privacy.

However, a challenge persists within this framework, namely the absence of a reputation update module. To remedy this, we have enhanced the task publication module and introduced a reputation management module, proposing a reputation management scheme based on the Empirical Cumulative distribution-based Outlier Detection method (ECOD) [18]. Presently, research on blockchain-based crowdsourcing often fixates on specific aspects, neglecting a holistic perspective. For instance, in the implementation of user incentive mechanisms, emphasis is frequently placed on utilizing reputation for incentives, overlooking the rationality and authenticity of updating user reputations. Our ECOD-based reputation management scheme circumvents reliance on subjective evaluations from task demanders, rendering it more reasonable and reliable. Reputation updates are executed through smart contracts, thereby preventing manual alterations.

This paper introduces an ECOD-based reputation update scheme tailored for crowdsourcing, with a focus on proficiently overseeing participant reputations and alleviating the impact of malicious activities on the sensing system. Our scheme embraces a blockchain-based crowdsourcing framework utilizing a homomorphic encryption method to ensure data transparency and tamper-resistance. Computation of user reputation scores relies on their behavioral history, actively discouraging undesirable conduct. Additionally, we introduce a dynamic weight incentive mechanism that mirrors alterations in participant reputation, enabling the system to allocate incentives based on user behavior and reputation. Our scheme undergoes evaluation on 11 datasets, revealing substantial enhancements in data credibility for crowdsourcing systems and a reduction in the influence of malicious behavior.

The main contribution of this paper can be summarized as follows:

1. We propose a reputation-based update scheme for ECOD. ECOD is an unsupervised anomaly detection algorithm that does not require extensive model training and can be applied to multi-dimensional data. It outperforms other anomaly detection algorithms in terms of accuracy and

computing speed. By leveraging ECOD's ability to detect anomalies in user-submitted data, we update user reputation scores and provide corresponding incentives based on these scores.

2. We present a novel crowdsourcing framework based on smart contracts, effectively addressing trust and third-party intermediary issues. The CKKS encryption scheme ensures privacy protection during the data transmission process, while the ECOD-based reputation update scheme protects the interests of both task demanders and users. This prevents malicious data uploads by users, thereby reducing potential losses. Task demanders can also provide incentives for negative evaluations that impact user reputation.
3. We evaluate our scheme across 11 datasets, demonstrating significant improvements in data credibility for crowdsourcing systems and a reduction in the influence of malicious behavior.

The remainder of this paper are organized as follows: In Section 2, we delineate some drawbacks of existing crowdsourcing frameworks and encapsulate our contributions. Section 3 furnishes a comprehensive description of the entire crowdsourcing framework workflow and the ECOD-based reputation update scheme. In Section 4, we present experimental results. Finally, Section 5 concludes the article, offering insights into the future directions of the crowdsourcing framework.

2 Related Work

Existing crowdsourcing frameworks often confront challenges in implementing effective reputation update schemes and ensuring user privacy during the incentive process. Typically, these frameworks rely on methods where task demanders offer positive or negative evaluations of user data to update reputation scores, and some lack a reputation update module entirely.

Tian et al. [19] proposed a distributed numerical estimation mechanism that achieved user privacy protection but relied on adding Gaussian noise to the data, which could not entirely eliminate the impact of Gaussian noise, often yielding suboptimal results in practice. Wu et al. [20] introduced a blockchain-based data truth estimation mechanism using additive homomorphic encryption, allowing user participation without knowledge of real data but could not ensure privacy security during the user incentive process. Huang et al. [21] presented an incentive mechanism based on complete information dynamic games, using homomorphic watermark technology for digital rights protection but lacked a reputation module, making it unable to guarantee user behavior as non-malicious. Zhang et al. [22] proposed a vehicle-based mobile crowdsourcing system using homomorphic encryption for privacy protection and efficient user incentive mechanisms but did not implement a reputation module to identify and penalize low-reputation malicious users. Xie et al. [23] introduced a drone-assisted mobile crowdsourcing framework based on reputation incentives and edge computing but relied on task demander evaluations for reputation updates, without accounting for malicious task demanders. Sun [24] proposed a task diffusion solution based on a social network influence propagation model but lacked a reputation management module, failing to effectively prevent malicious user feedback. Fu et al. [8] presented a task assignment scheme (PCTA-SG) based on employee location privacy protection and an employee elite selection mechanism but lacked a reputation management module to ensure the authenticity of user-uploaded data. Jiang et al. [9] proposed an incentive mechanism protecting user rights using uncertain and hidden bids but could not effectively safeguard task demander interests due to the absence of a reputation management module. Huang et al. [10] researched potential combinations of traditional Automated Passenger Counters (APC) and a novel source capable of collecting detailed mobile demand data but did not include a reputation management module to prevent malicious data uploads. Tutsoy et al. [25] proposed an AI based long-term policy making algorithm aiming to maximize the number of the students attending the schools while minimizing the number of the casualties. Tutsoy et al. [26] proposed a study that can contribute to marketing science by presenting a strong estimation of future consumer behavior in tourism through machine-learning-based predictions. Corrochano et al. [27] proposed a new hybrid physics-based machine learning model with a simple, robust and generalizable architecture, which allows reconstructing databases from very few sensors and with a very low computational cost.

Hence, this paper introduces the following contributions: We propose a reputation update scheme based on ECOD. ECOD, as an unsupervised anomaly detection algorithm, does not necessitate lengthy model training and can be applied to multi-dimensional data. Its accuracy and operational speed surpass those of other anomaly detection algorithms. Leveraging ECOD's ability to detect anomalies in user-submitted data, we update user reputation scores and provide corresponding incentives based on these scores.

The entire crowdsourcing framework operates based on smart contracts, effectively eliminating trust issues associated with third-party intermediaries. The CKKS encryption scheme ensures privacy protection during data transmission, and the ECOD-based reputation update scheme safeguards the dual interests of task demanders and users. It prevents users from maliciously uploading data that could lead to losses for task demanders and task demanders from providing negative evaluations that impact user reputation and incentives.

3 Our Crowdsourcing System

In this paper, we present a custom reputation update scheme designed for crowdsourcing, emphasizing the efficient management of participant reputations and mitigation of the impact of malicious activities on the sensing system. The reputation update scheme is based on the Empirical Cumulative Distribution-based Outlier Detection method (ECOD). Our framework employs blockchain technology in crowdsourcing, integrating a homomorphic encryption method to ensure data transparency and resistance to tampering. User reputation scores are computed based on their historical behavior, actively discouraging undesirable actions. Furthermore, we propose a dynamic weight incentive mechanism that adjusts based on changes in participant reputation, allowing the system to distribute incentives according to user behavior and reputation.

As shown in Fig. 1, both data demanders and data providers are assigned reputation scores, which are influenced by their behavior. Task demanders who consistently send rewards punctually and engage in other positive behaviors see an increase in their reputation scores. Similarly, users who consistently provide valid data and exhibit positive behaviors experience an increase in their reputation scores. Conversely, those who do not engage in these positive behaviors face a decrease in their reputation scores.

Task demanders must specify the required number of users for a task (i.e., ranging from a minimum to a maximum), the minimum task waiting time (i.e., tasks start to be allocated after this time if the number of participants meets the minimum requirement or reaches the maximum requirement), and can optionally set a minimum reputation score requirement.

Users have the discretion to accept a task based on the reputation score of the demander. Additionally, users need to provide the cost required to complete the task through encrypted bidding. Smart contracts consider both the user's reputation score and bidding amount to select an appropriate number of users for the task.

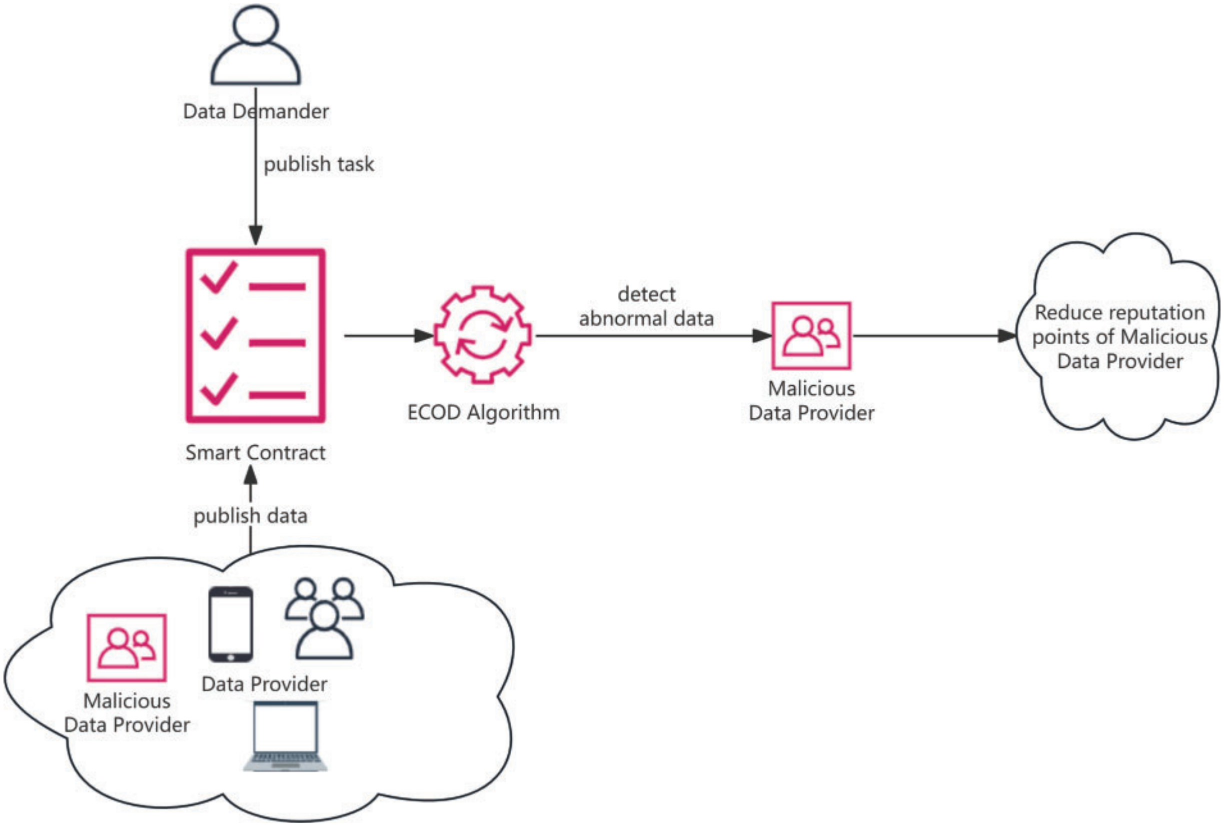

Figure 1: System framework

3.1 Intelligent Crowdsourcing System

In this framework, each data demander (i.e., task demander) r_i has specific task requirements l_i. These sensing tasks necessitate data providers (i.e., users) to collect data on platforms such as smart terminals and transmit the sensed data to the data demander via the blockchain. The data collected by data providers u_j concerning the task l_i is denoted as $d_{i,j}$. Smart contracts aggregate the data sent by data providers to derive a result d_i, considered an estimation of the task's ground truth d_i^*. It is important to emphasize that the true values d_i^* of the tasks remain unknown to both data demanders and data providers. These critical symbols are detailed in Table 1.

Table 1: Description of important symbols

Symbol	Description
U	The data provider set
R	The demander set
L	The task set
S_R	The winning set of demanders
S_{U_i}	The winning set of providers corresponding to the demanders

(Continued)

Table 1 (continued)

Symbol	Description
a_i,	Demander r_i's bidding
$b_{i,j}$	Provider u_j's bidding
$\hat{a}_i$,	a_i's ciphertext
$\hat{b}_{i,j}$	$b_{i,j}$'s ciphertext
p_{r_i}	Fees paid by demanders r_i
p_{u_j}	Fees received by providers u_j
d_i	The task l_i's data estimate
$\hat{d}_i$	d_i's ciphertext
v, B, N, h, ℓ	Encryption parameters

Since smart contracts are transparent and open to the public, it is crucial to employ encryption services to safeguard the data (i.e., collected data) and information (i.e., privacy of the transacting parties). This encryption ensures the confidentiality of the data and the privacy of the participants involved in the transactions.

The primary steps of our intelligent crowdsourcing system are illustrated in Fig. 2 and outlined as follows:

1. Bidding Encryption and Task Publication & Acceptance:
   (a) Data demanders r_i encrypt their bids a_i using the public key of the encryption service center. The bid a_i signifies the maximum reward they are willing to pay if the task l_i is successfully executed (Step 1).
   (b) Data demanders submit a data collection request to a smart contract, including the sensing task l_i, the encrypted bid $\hat{a}i$, the number of users to accept the task min, max, the minimum waiting time t, and an optional minimum reputation score requirement rep (Step 2).
   (c) The smart contract publishes the collected task set L to the public (Step 3).
   (d) Data providers u_j assess the reputation score of the data demander to decide whether to accept the task. Upon acceptance, data providers need to send their bid bi,j to the encryption service center, where $b_{i,j}$ denotes the cost of u_j required for completing the task l_i (Step 4).
   (e) Data providers then submit the set of tasks $D_j \subseteq L$ they are willing to perform and the corresponding $l_i \in D_j$ encrypted bids $\hat{b}_{i,j}$ to the smart contract (Step 5).
2. Data Provider Selection and Incentive Mechanism:
   (a) With the assistance of the encryption service center, smart contracts use the received encrypted bids $\hat{a}i$ to determine a set of winning data demanders S_R.
   (b) When the number of participants reaches the maximum max or the minimum waiting time t is met with the minimum number of participants min, the smart contract combines the encrypted bids $\hat{b}i,j$ of data providers with their reputation scores to determine the set of winning data providers S_{U_i} for each winning data demander (($r_i \in S_R$)'s task l_i.

(c) The fees p_{r_i} to be paid by each winning data demander r_i and the rewards p_{u_j} to be paid to each winning data provider u_j are calculated (Step 6).

(d) In reality, $p_{u_j} = \sum_{i;l_i \in D_j} p_{i,j}.p_{i,j}$ represents the reward obtained by data providers u_j for each task l_i they execute.

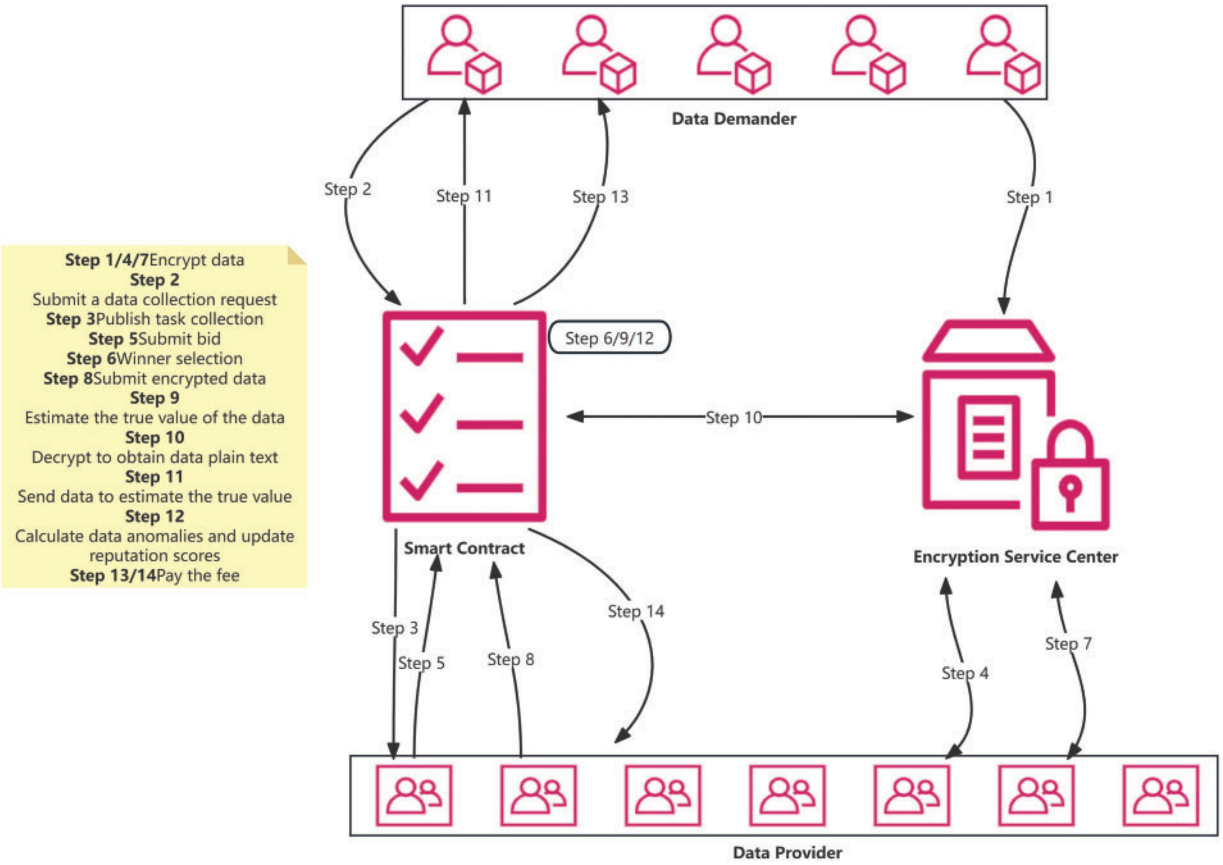
Data Demander
Step 2
Step 11
Step 13
Step 1
Step 1/4/7Encrypt data
Step 2
Submit a data collection request
Step 3Publish task collection
Step 5Submit bid
Step 6Winner selection
Step 8Submit encrypted data
Step 9
Estimate the true value of the data
Step 10
Decrypt to obtain data plain text
Step 11
Send data to estimate the true value
Step 12
Calculate data anomalies and update reputation scores
Step 13/14Pay the fee
Step 6/9/12
Step 10
Smart Contract
Encryption Service Center
Step 3
Step 5
Step 8
Step 14
Step 4
Step 7
Data Provider

Figure 2: Intelligent crowdsourcing system framework

3. Data Encryption:
   (a) Each selected data provider $u_j \in S_{U_i}$ encrypts the data $d_{i,j}$ collected for each task $l_i \in D_j$ using the public key of the encryption service center (Step 7).
   (b) The encrypted data $\hat{d}_{i,j}$ is submitted to the smart contract (Step 8).
4. Data Fusion:
   (a) The smart contract computes the fusion result $\hat{d}_i$ from the encrypted data submitted by winning data providers for each task l_i (Step 9).
   (b) It decrypts these results using the encryption service center, obtaining plaintext values d_i (Step 10).
   (c) The plaintext outcomes are subsequently forwarded to the respective winning data demanders $r_i \in S_R$ (Step 11), where d_i represents an estimation of the ground truth d_i^* for the task l_i.

5. Reputation Score Update:
   (a) Following the acquisition of plaintext values d_i, the smart contract employs the ECOD algorithm to identify data anomalies and adjust the reputation scores of both data demanders and providers accordingly (Step 12).
6. Reward Collection:
   (a) Finally, the smart contract retrieves fees p_{r_i} from winning data demanders r_i (Step 13) and disburses rewards p_{u_j} to winning data providers u_j (Step 14).

This comprehensive process ensures privacy, oversees reputation, and establishes a secure, transparent, and efficient crowdsourcing environment. It leverages encryption, smart contracts, and reputation-based mechanisms to streamline data collection and processing while safeguarding the interests of both data demanders and providers.

3.2 Anomaly Detection with ECOD

The ECOD (Elliptic Curve-Based Outlier Detection) algorithm stands out as an unsupervised anomaly detection method, recognized for its exceptional performance in this domain. This section offers a concise insight into the ECOD algorithm's procedure, while Section 4 provides comprehensive experimental results, detailing its accuracy. The workflow of the ECOD algorithm is succinctly summarized in Algorithm 1.

Algorithm 1: Anomaly detection with the ECOD algorithm

```
Input: input data $\mathbf{X} = X^{n}_{i=1} \in \mathbb{R}^{n\times d}$ with $n$ samples and $d$ features; $X^{(j)}_i$ refers to the $j$-$th$ feature of the $i$-$th$ sample.
Output: outlier scores $\mathbf{O} := ECOD(X) \in \mathbb{R}^n$.
for each dimension $j$ in 1,..., $d$ do
    Estimate left and right tail ECDFs:
        left tail ECDF: $\hat{F}^{(j)}_{left}(z) = \frac{1}{n}\sum^{n}_{i=1}\mathbb{I}\{X^{(j)}_i\} \leq z$ for $z \in \mathbb{R}$,
        right tail ECDF: $\hat{F}^{(j)}_{right}(z) = \frac{1}{n}\sum^{n}_{i=1}\mathbb{I}\{X^{(j)}_i\} \geq z$ for $z \in \mathbb{R}$.
        where $\mathbb{I}\{\bullet\}$ is the indicator function that is 1 when its argument is true and is 0 otherwise.
    Compute the sample skewness coefficient for the $j$-$th$ feature's distribution:
        $\gamma_j = \dfrac{\frac{1}{n}\sum^{n}_{i=1}(X^{(j)}_i - \overline{X^{(j)}})}{\left[\frac{1}{n}\sum^{n}_{i=1}(X^{(j)}_i - \overline{X^{(j)}})^2\right]^{\frac{3}{2}}}$,
        where $\overline{X^{(j)}} = \frac{1}{n}\sum^{n}_{i=1} X^{(j)}_i$ is the sample mean of the $j$-$th$ feature.
end for
for each sample $i$ in 1,..., n do
    Aggregate tail probabilities of $X_i$ to obtain outlier score $O_i$:
        $O_{left-only}(X_i) = -\sum^{d}_{j=1} log(\hat{F}^{(j)}_{left}(X^{(j)}_i))$,
        $O_{right-only}(X_i) = -\sum^{d}_{j=1} log(\hat{F}^{(j)}_{right}(X^{(j)}_i))$,
        $O_{auto}(X_i) = -\sum^{d}_{j=1}[\mathbb{I}\{\gamma_j < 0\}log(\hat{F}^{(j)}_{left}(X^{(j)}_i)) + \mathbb{I}\{\gamma_j \geq 0\}log(\hat{F}^{(j)}_{right}(X^{(j)}_i))]$.
    Set the final outlier score for point $X_i$ to be $O_i = maxO_{left-only}(X_i),\ O_{right-only}(X_i),\ O_{auto}(X_i)$.
```

(Continued)

Algorithm 1 (continued)

```
    end for
Return outlier scores $\mathbf{O} = (O_1, ..., O_n)$.
```

3.3 Reputation Update with ECOD

Within the task allocation process, task demanders establish a minimum reputation score (rep) requirement for each task to be accepted. The smart contract evaluates whether the reputation score ($repu_j$) of each potential data provider (u_j) in the pool of candidates for task (l_i) acceptance (S_L) exceeds the specified minimum reputation score (rep). If this criterion is satisfied, the user is included in the set of winning data providers (SU_i) for the task (l_i).

This reputation update scheme ensures that only data providers with reputation scores surpassing the defined threshold are eligible for task participation. Leveraging ECOD-based reputation scoring offers an effective method for assessing the trustworthiness and reliability of data providers. It serves to exclude users with low reputation or potential malicious intent from task participation, thereby enhancing the overall quality and reliability of crowdsourced data. As previously highlighted, the ECOD algorithm proves to be a robust tool for unsupervised anomaly detection. Its capability to accurately pinpoint data providers with abnormal behavior or low reputation scores makes it a valuable asset within the reputation management module of our crowdsourcing framework. By integrating ECOD-based reputation scoring, the framework ensures the selection of only reputable and trustworthy data providers for tasks, contributing significantly to the overall success and reliability of the crowdsourcing system. This reputation update mechanism plays a pivotal role in upholding a high standard of data quality and user trust within the platform. Algorithm 2 identifies users eligible to accept the task based on their reputation scores.

Algorithm 2: Identify users eligible to accept the task based on their reputation scores

```
Input: task collection L, alternate data provider set $S_L$.
Output: the set of winning data providers $S_{U_i}$ corresponding to the task $l_i$.
for each $l_i \in L$ $(i = 1,\ 2,\ ...,\ m)$ do
    for each member $u_j$ of the set of candidate data providers $S_{l_i}$ corresponding to each task $l_i$ do
        if $rep_{u_j} > rep$ then
            add the user $u_j$ to the set of winning data providers $S_{U_i}$ for the task $l_i$
        end if
    end for
end for
Return the winning set of data providers $S_{U_i}$ corresponding to task $l_i$.
```

Algorithm 3 employs the ECOD algorithm for data quality assessment and subsequent updates to user reputation scores. Upon the submission of data by collectors, the smart contract employs the ECOD algorithm to assess data quality. This evaluation leads to the updating of reputation scores and the calculation of rewards, considering both the reputation score and the cost incurred by the data collector. Noteworthy is the algorithm's robust scalability, demanding minimal data for fitting and training. Task demanders or a subset of users with the highest reputation scores can provide training data. The anomaly score, indicating the likelihood of data being anomalous, increases with higher values. Anomalous data is denoted as 1, while normal data is marked as 0. User reputation scores, initially set at 0.5, span from 0 to 1 and undergo changes based on the proportion α of anomalous data in their submissions. The reputation score adjustment is determined by the following formula:

$$\Delta rep = \begin{cases} +0.05,\ 0 \le \alpha \le 0.03 \\ -0.05 \times (1+\alpha),\ 0.03 \le \alpha < 0.1 \\ -0.05 \times (2+\alpha),\ \alpha \ge 0.1 \end{cases}$$

Algorithm 3: Employs the ECOD algorithm for data quality assessment and subsequent updates to user reputation scores

```
**Input:** data $d_{i,j}$ collected for each task $l_i \in D_j$.
**Output:** user's reputation score $rep_{u_j}$.
for each $d_{i,j} \in D_j$ do
    calculate the proportion $\alpha$ of abnormal data $d_{i,j}$ through ECOD($d_{i,j}$).
    if $0 \le \alpha \le 0.03$ then
        $\Delta\ rep = +0.05$.
    else if $0.03 \le \alpha < 0.1$
        $\Delta\ rep = -0.05 \times (1+\alpha)$.
    else
        $\Delta\ rep = -0.05 \times (2+\alpha)$.
    end if
    $rep_{u_j} += \Delta\ rep$.
end for
Return $rep_{u_j}$.
```

The guidelines governing user conduct and rewards based on their reputation scores are as follows. *(a) User Reputation Below 0:* Users with a reputation score falling below 0 are ineligible to accept tasks. To restore their reputation score, a fine must be paid, resetting it to 0.3. *(b) User Reputation Below 0.3:* When a user's reputation score is below 0.3, they receive a reward equivalent to their submitted cost for task completion, denoted as $p_{uj} = b_{i,j} \times (1 + rep)$. In essence, they are compensated for their efforts, with the reward capped to cover their costs. *(c) User Reputation Above 0.3:* Users boasting a reputation score surpassing 0.3 are entitled to rewards $p_{uj} = b_{i,j} \times (2 + rep)$ for task completion. Their reputation score mirrors their trustworthiness and reliability in the system, influencing their earnings positively. These regulations establish an incentive framework motivating users to maintain a reputation score exceeding 0.3 by submitting high-quality and reliable data. Users with lower reputation scores face penalties and restricted rewards, while those with higher reputation scores enjoy increased trust and enhanced compensation for their contributions. This structured approach fosters responsible and ethical user behavior, contributing to the overall quality and reliability of crowdsourced data within the platform.

Algorithm 4 verifies whether the user's reputation score falls below 0, proceeding to the subsequent step if it does. Upon confirming that the user's reputation score is below 0, a conditional check assesses whether the user opts to pay a fine. The algorithm assumes that remitting the fine will partially restore the user's reputation score. In the scenario where the user's reputation score is below 0.3, and they choose not to pay a fine, it signifies that the user's reputation remains below the acceptable threshold. Consequently, the user is granted a reward equivalent to their submitted cost for task completion. This mechanism ensures compensation for users with lower reputation scores, albeit constrained to covering their costs. When the user's reputation score attains or exceeds 0.3, signifying their good standing, they become eligible for rewards.

Algorithm 4: Incentive mechanism with user reputation scores

```
Input: User's Reputation Score $rep_{u_j}$.
Output: User's Reward $p_{u_j}$.
if $rep_{u_j} \leq 0$ then
    if user $u_j$ pays fine then
        $rep_{u_j} = 0.3$.
    end if
else if $0 < \alpha \leq 0.3$
    $p_{u_j} = b_{i,j} \times (1 + rep)$.
else
    $p_{u_j} = b_{i,j} \times (2 + rep)$.
end if
Return $p_{u_j}$.
```

4 Evaluation Results

In the realm of the reputation management scheme, the precision of the reputation score update is heavily contingent on the identification of anomalous data; thus, the reliability of the ECOD algorithm assumes paramount importance. Consequently, experiments have been undertaken to juxtapose the ROC curves, precision, and runtime of the ECOD algorithm. The experimentation transpired on a Windows laptop equipped with an Intel(R) Core(TM) i7-10875H CPU @ 2.30 GHz and 16 GB of RAM. We employed 11 datasets in the.mat format, extracted from the ODDS report. Table 2 furnishes comprehensive insights into dataset sizes, dimensions, and anomaly statistics.

Table 2: Dataset Description

Dataset	Number of samples (n)	Number of dimensions (d)	Outlier (%)
Arrhythmia (mat)	452	274	14.601
Breastw (mat)	683	9	34.992
Cardio (mat)	1831	21	9.612
Ionosphere (mat)	351	33	35.897
Optdigits (mat)	5216	64	2.875
Pima (mat)	768	8	34.895
Satellite (mat)	6435	36	31.639
Satimage-2 (mat)	5803	36	1.223
Shuttle (mat)	10,000	9	7.120
Wbc (mat)	378	30	5.555
Wine (mat)	129	13	7.751

The ensuing discussion revolves around a comparative experiment involving 11 datasets in the.mat format, wherein the ECOD algorithm was pitted against IForest, k-nearest neighbors (KNN), and local outlier factor (LOF). The assessment centered on the receiver operating characteristic (ROC) and average precision (AP). The ROC and AP results for the four algorithms across the 11 datasets are elucidated in Tables 3 and 4. As delineated in Table 3, ECOD and IForest consistently clinched the top spot across almost all datasets. Particularly noteworthy is ECOD's average ROC of 0.863, securing its position as the foremost algorithm among the quartet. Table 4 accentuates the dominance of ECOD and IForest in terms of Average Precision (AP), with both algorithms consistently leading in nearly all datasets. ECOD attains an average AP of 0.652, firmly establishing itself as the preeminent performer among the four.

Table 3: Receiver operating characteristic (ROC) results

Dataset	IForest	KNN	LOF	ECOD
Arrhythmia (mat)	0.802 (1)	0.786 (3)	0.779 (4)	0.802 (1)
Breastw (mat)	0.987 (2)	0.976 (3)	0.470 (4)	0.994 (1)
Cardio (mat)	0.923 (1)	0.724 (3)	0.574 (4)	0.897 (2)
Ionosphere (mat)	0.847 (3)	0.927 (1)	0.875 (2)	0.831 (4)
Optdigits (mat)	0.721 (2)	0.371 (4)	0.450 (3)	0.733 (1)
Pima (mat)	0.678 (2)	0.708 (1)	0.627 (4)	0.664 (3)
Satellite (mat)	0.702 (1)	0.684 (2)	0.557 (4)	0.661 (3)
Satimage-2 (mat)	0.995 (1)	0.954 (3)	0.458 (4)	0.985 (2)
Shuttle (mat)	0.998 (1)	0.738 (3)	0.524 (4)	0.998 (1)
Wbc (mat)	0.931 (4)	0.938 (2)	0.935 (3)	0.975 (1)
Wine (mat)	0.816 (3)	0.518 (4)	0.905 (2)	0.949 (1)
AVG	0.837 (2)	0.757 (3)	0.650 (4)	0.863 (1)

Table 4: Average precision (AP) results

Dataset	IForest	KNN	LOF	ECOD
Arrhythmia (mat)	0.506 (1)	0.397 (3)	0.374 (4)	0.473 (2)
Breastw (mat)	0.972 (2)	0.927 (3)	0.322 (4)	0.988 (1)
Cardio (mat)	0.576 (2)	0.345 (3)	0.163 (4)	0.579 (1)
Ionosphere (mat)	0.789 (3)	0.924 (1)	0.821 (2)	0.719 (4)
Optdigits (mat)	0.055 (1)	0.022 (4)	0.029 (3)	0.053 (2)
Pima (mat)	0.503 (3)	0.515 (2)	0.430 (4)	0.541 (1)
Satellite (mat)	0.654 (1)	0.543 (3)	0.390 (4)	0.585 (2)
Satimage-2 (mat)	0.929 (1)	0.419 (3)	0.027 (4)	0.860 (2)
Shuttle (mat)	0.986 (1)	0.204 (3)	0.142 (4)	0.981 (2)
Wbc (mat)	0.590 (2)	0.529 (4)	0.558 (3)	0.783 (1)
Wine (mat)	0.279 (3)	0.095 (4)	0.361 (2)	0.608 (1)
AVG	0.622 (2)	0.447 (3)	0.329 (4)	0.652 (1)

Figs. 3 and 4 visually illustrate that among a range of unsupervised learning techniques, ECOD and IForest demonstrate high accuracy across 11 datasets. Notably, ECOD consistently delivers outstanding performance even in scenarios where IForest exhibits subpar results.

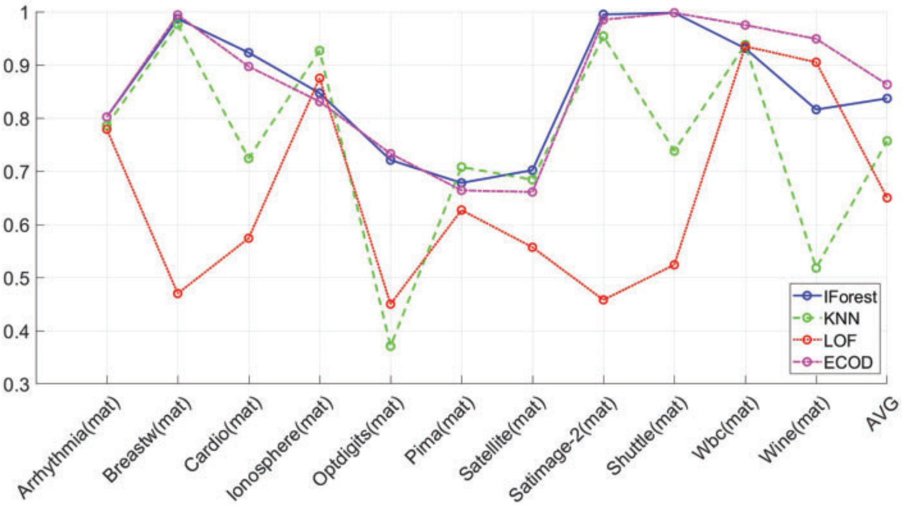

Figure 3: Receiver operating characteristic (ROC) results

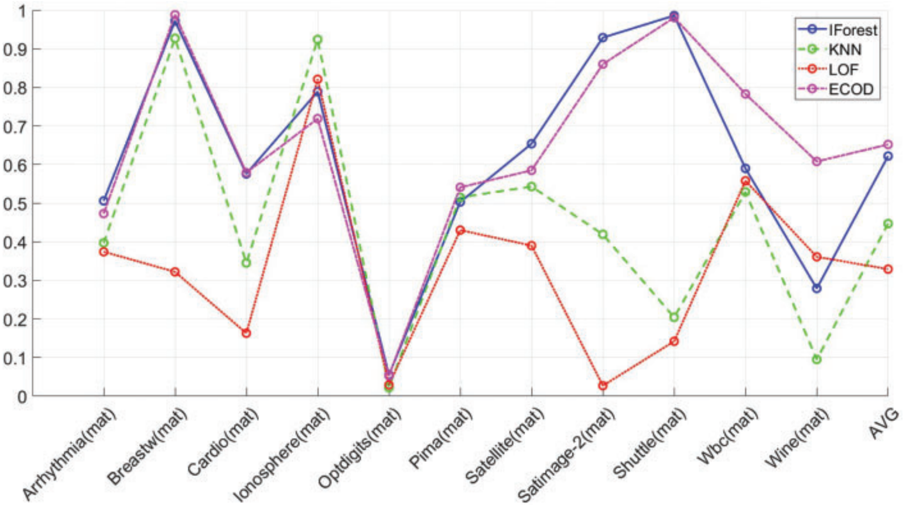

Figure 4: Average precision (AP) results

We infer that the time complexity of ECOD is $O(n \cdot d)$, where n represents the sample size, and d stands for the dimension. Table 5 and Fig. 5 elucidate alterations in ECOD's runtime concerning the augmentation of both sample size and dimension on the experimental computer. The experiments underscore ECOD's commendable performance, particularly in scenarios involving substantial sample sizes and elevated dimensions.

Table 5: ECOD running time

	d = 10	d = 100	d = 1000	d = 10,000
n = 1000	0.068	0.173	1.163	11.460
n = 10,000	0.171	0.468	5.244	55.190
n = 100,000	0.640	7.185	70.541	567.105
n = 1,000,000	11.403	130.974	694.405	5376.593

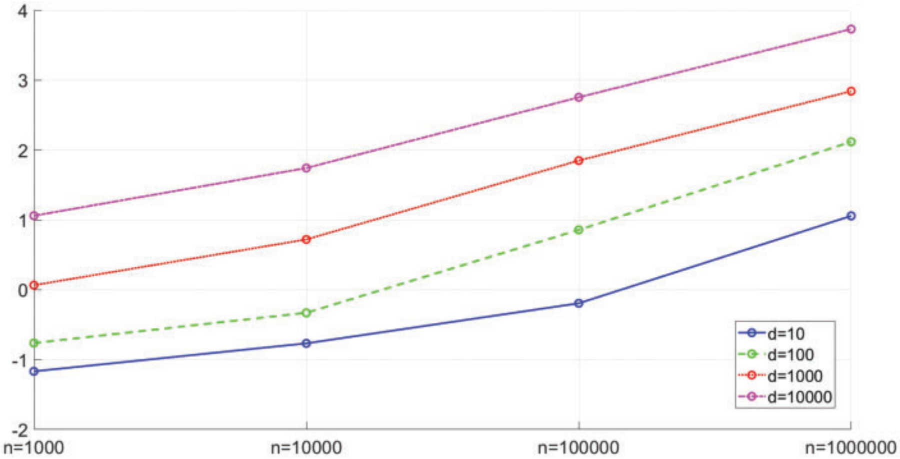

Figure 5: ECOD running time

The paper highlights a beneficial aspect of ECOD, which is that it does not require retraining when dealing with new sample points with a larger sample size, assuming that data shifts do not occur. This property makes ECOD advantageous for real-time detection. The accuracy and runtime efficiency of ECOD contribute significantly to the smooth operation of both reputation updates and the entire crowdsourced sensing system.

Additionally, a preliminary evaluation was also carried out on the CKKS encryption algorithm, comparing its performance with that of Brakerski/Fan-Vercauteren (BFV) [28] and fast fully homomorphic encryption scheme over the torus (TFHE) [29].

BFV: The BFV cryptosystem operates on exact calculus, albeit working with integers necessitates dealing with rounding errors. Consequently, the precision is contingent on the degree of rounding applied to each value. To ensure successful decryption, the resultant value must remain sufficiently low, limiting the initial precision level. To strike a balance between precision and output size, input instances and weights were scaled to fit within 8 bits. With BFV, a negligible loss of 0.01 percentage points in precision was observed, indicating only one additional correct guess out of 10,000 samples in clear domain computation.

CKKS: Designed for approximate calculations, the CKKS cryptosystem assumes that if the maximum output value is significantly distinct from other values, the encrypted network should yield identical results. This assumption stems from the network's nature as a low-degree polynomial function. In practice, the precision of CKKS matched that of the clear values at 96.34.

TFHE: Utilizing the same scaling approach as BFV, ensuring each input instance and weight fits within 8 bits, resulted in a comparable 0.01 percentage point loss in precision. The scaling decision aimed to strike a balance between precision and computational efficiency.

As show in Table 6, the substantial speed contrast between BFV and CKKS can be attributed to parameter discrepancies, with BFV ciphertexts being roughly twice the size of CKKS ciphertexts. This size disparity translates into slower operations in the BFV context, underscoring the imperative of parameter optimization prior to computational tasks.

Table 6: Encryption algorithm running time comparison

	BFV	CKKS	TFHE
KeyGen	4.5 s	1.96 s	0.16 s
Features encryption	0.16 s	0.29 s	0.08 s
Weights encryption	0.24 s	0.35 s	16 s
Network evaluation	2.16 s	0.56 s	>3 days
Total time	7.06 s	3.16 s	>3 days

5 Conclusions

The paper delves into the contemporary landscape of crowdsourced sensing frameworks, underscoring the prevalent inadequacies in existing solutions, particularly the dearth of effective mechanisms for reputation updates and privacy protection during the user incentive process. Many of these frameworks hinge on task publishers' evaluative feedback, lacking a robust reputation update methodology. Within this paper, a novel reputation update scheme leveraging the ECOD anomaly detection is proposed. Empirical findings showcase ECOD's prowess in anomaly detection, manifesting in high accuracy, swift runtime, and commendable scalability. The algorithm adeptly updates user reputation scores, ensuring the seamless functioning of the crowdsourced sensing system. However, an apprehension lingers regarding potential privacy leaks of user identities during transactions. Consequently, the text advocates for future research to delve deeper into developing privacy protection mechanisms for the incentive process.

Acknowledgement: This logical organization and content quality of this paper have been enhanced, so the authors thank anonymous reviewers and journal editors for assistance.

Funding Statement: This work is supported by National Natural Science Foundation of China (Nos. U21A20463, 62172117, 61802383), Research Project of Pazhou Lab for Excellent Young Scholars (No. PZL2021KF0024), and Guangzhou Basic and Applied Basic Research Foundation (Nos. 202201010330, 202201020162, 202201020221).

Author Contributions: Study conception and design: T. Wang and K. Chen; analysis and interpretation of results: T. Wang; draft manuscript preparation: T. Wang and K. Chen. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The 11 experimental datasets have been sourced from the open-source community available at http://odds.cs.stonybrook.edu.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Vahdat-Nejad H, Asani E, Mahmoodian Z, Mohseni MH. Context-aware computing for mobile crowd sensing: a survey. Future Gener Comput Syst. 2019;99:321–32. doi:10.1016/j.future.2019.04.052.
2. Gao Y, Dong W, Guo K, Liu X, Chen Y, Liu X, et al. Mosaic: a low-cost mobile sensing system for urban air quality monitoring. In: IEEE INFOCOM 2016–The 35th Annual IEEE International Conference on Computer Communications; 2016; San Francisco, CA, USA, IEEE. p. 1–9.
3. Abbondati F, Biancardo SA, Veropalumbo R, Dell'Acqua G. Surface monitoring of road pavements using mobile crowdsensing technology. Measurement. 2021;171:108763. doi:10.1016/j.measurement.2020.108763.
4. Thiagarajan A, Ravindranath L, LaCurts K, Madden S, Balakrishnan H, Toledo S, et al. Vtrack: accurate, energy-aware road traffic delay estimation using mobile phones. In: Proceedings of the 7th ACM Conference on Embedded Networked Sensor Systems; 2009; Berkeley, California, USA. p. 85–98.
5. Capponi A, Fiandrino C, Kantarci B, Foschini L, Kliazovich D, Bouvry P. A survey on mobile crowdsensing systems: challenges, solutions, and opportunities. IEEE Commun Surv Tutor. 2019;21(3):2419–65. doi:10.1109/COMST.9739.
6. Gong X, Shroff NB. Truthful mobile crowdsensing for strategic users with private data quality. IEEE/ACM Trans Netw. 2019;27(5):1959–72. doi:10.1109/TNET.90.
7. Jin H, Su L, Xiao H, Nahrstedt K. Incentive mechanism for privacy-aware data aggregation in mobile crowd sensing systems. IEEE/ACM Trans Netw. 2018;26(5):2019–32. doi:10.1109/TNET.2018.2840098.
8. Fu Y, Huang B, Liu X, Chen J, Lu S. Privacy-preserving mobile crowd sensing task assignment with Stackelberg game. Comput Netw. 2023;234:109917. doi:10.1016/j.comnet.2023.109917.
9. Jiang X, Ying C, Li L, Wu H, Luo Y, Düdder B. Incentive mechanism for uncertain tasks under differential privacy. arXiv preprint arXiv:230516793. 2023.
10. Huang Z, de Villafranca AEM, Sipetas C, Quach T. Crowd-sensing commuting patterns using multi-source wireless data: a case of Helsinki commuter trains. arXiv preprint arXiv:230202661. 2023.
11. Lin J, Yang D, Wu K, Tang J, Xue G. A sybil-resistant truth discovery framework for mobile crowdsensing. In: 2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS); 2019; Dallas, TX, USA, IEEE. p. 871–80.
12. Zhou J, Shen H, Lin Z, Cao Z, Dong X. Research advances on privacy preserving in edge computing. J Comput Res Dev. 2020;57(10):2027–51.
13. Luu L, Narayanan V, Zheng C, Baweja K, Gilbert S, Saxena P. A secure sharding protocol for open blockchains. In: Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security; 2016; Hofburg Palace, Vienna, Austria. p. 17–30.
14. Ying C, Xia F, Li J, Si X, Luo Y. Incentive mechanism based on truth estimation of private data for blockchain-based mobile crowdsensing. J Comput Res Dev. 2022;59(10):2212–32.
15. Zhang P, Zhou M. Security and trust in blockchains: architecture, key technologies, and open issues. IEEE Trans Comput Soc Syst. 2020;7(3):790–801. doi:10.1109/TCSS.6570650.
16. Zheng P, Xu Q, Zheng Z, Zhou Z, Yan Y, Zhang H. Meepo: sharded consortium blockchain. In: 2021 IEEE 37th International Conference on Data Engineering (ICDE); 2021; Chania, Greece, IEEE. p. 1847–52.
17. Cheon JH, Kim A, Kim M, Song Y. Homomorphic encryption for arithmetic of approximate numbers. In: Advances in Cryptology–ASIACRYPT 2017: 23rd International Conference on the Theory and Applications of Cryptology and Information Security; 2017 Dec 3–7; Hong Kong, China, Springer. p. 409–37.
18. Li Z, Zhao Y, Hu X, Botta N, Ionescu C, Chen G. Ecod: unsupervised outlier detection using empirical cumulative distribution functions. IEEE Trans Knowl Data Eng. 2023;35(12):12181–93. doi:10.1109/TKDE.2022.3159580.

19. Tian Y, Yuan J, Song H. Secure and reliable decentralized truth discovery using blockchain. In: 2019 IEEE Conference on Communications and Network Security (CNS); 2019; Washington DC, USA, IEEE. p. 1–8.
20. Wu H, Düdder B, Wang L, Sun S, Xue G. Blockchain-based reliable and privacy-aware crowdsourcing with truth and fairness assurance. IEEE Internet Things J. 2021;9(5):3586–98.
21. Huang Z, Zheng J, Xiao M. Privacy-enhanced crowdsourcing data trading based on blockchain and stackelberg game. In: 2021 IEEE 18th International Conference on Mobile Ad Hoc and Smart Systems (MASS); 2021; Denver, CO, USA, IEEE. p. 621–6.
22. Zhang C, Zhu L, Xu C, Sharif K. PRVB: achieving privacy-preserving and reliable vehicular crowdsensing via blockchain oracle. IEEE Trans Veh Technol. 2020;70(1):831–43.
23. Xie L, Su Z, Chen N, Xu Q. Secure data sharing in UAV-assisted crowdsensing: integration of blockchain and reputation incentive. In: 2021 IEEE Global Communications Conference (GLOBECOM); 2021; Madrid, Spain, IEEE. p. 1–6.
24. Sun C. Task diffusion of mobile crowdsensing in social network. Comput Syst Appl. 2023;32(10):166–74.
25. Tutsoy O, Tanrikulu C. A machine learning-based 10 years ahead prediction of departing foreign visitors by reasons: a case on Türkiye. Appl Sci. 2022;12(21):11163. doi:10.3390/app122111163.
26. Tutsoy O. COVID-19 epidemic and opening of the schools: artificial intelligence-based long-term adaptive policy making to control the pandemic diseases. IEEE Access. 2021;9:68461–68471. doi:10.1109/ACCESS.2021.3078080.
27. Díaz-Morales P, Corrochano A, López-Martín M, Le Clainche S. Deep learning combined with singular value decomposition to reconstruct databases in fluid dynamics. Expert Syst Appl. 2024;238:121924. doi:10.1016/j.eswa.2023.121924.
28. Halevi S, Polyakov Y, Shoup V. An improved RNS variant of the BFV homomorphic encryption scheme. In: Topics in Cryptology–CT-RSA 2019: The Cryptographers' Track at the RSA Conference 2019; 2019 Mar 4–8; San Francisco, CA, USA, Springer. p. 83–105.
29. Chillotti I, Gama N, Georgieva M, Izabachène M. TFHE: fast fully homomorphic encryption over the torus. J Cryptol. 2020;33(1):34–91. doi:10.1007/s00145-019-09319-x.