Hindawi
Security and Communication Networks
Volume 2022, Article ID 5626305, 11 pages
https://doi.org/10.1155/2022/5626305

Research Article

Grouping-Based Reliable Privacy Preservation for Blockchain-Assisted Data Aggregation in Mobile Crowdsensing

Yajie Li,[1,2] **Guanghui Wang**,[1,2] **Haochen Yang**,[1,3] **Fang Zuo**,[2,4] **Junyang Yu**,[1,3] **and Xin He**[1,3]

[1]*School of Software, Henan University, Kaifeng 475004, China*
[2]*Henan International Joint Laboratory of Intelligent Network Theory and Key Technology, Henan University, Kaifeng 475004, China*
[3]*Henan Provincial Engineering Research Center of Intelligent Data Processing, Henan University, Kaifeng 475004, China*
[4]*Software Engineering Intelligent Information Processing Innovation Base-Subject Innovation Base of Henan Higher Universities, Henan University, Kaifeng 475004, China*

Correspondence should be addressed to Guanghui Wang; gwang@vip.henu.edu.cn and Xin He; hxsyjkf@foxmail.com

Received 20 March 2022; Revised 12 July 2022; Accepted 5 August 2022; Published 24 August 2022

Academic Editor: Zhili Zhou

Privacy-preserving data aggregation is an important technology for mobile crowdsensing. Blockchain-assisted data aggregation enables the traceability of sensing data to improve the trustworthiness of data aggregation results. However, directly using blockchains for data aggregation may introduce the risk of privacy leakage because all nodes, including malicious nodes, can access the data on blockchains. In this paper, we propose a grouping-based reliable privacy-preserving data aggregation (RPPDA) method using private blockchains for mobile crowdsensing. First, the sensing nodes are divided into multiple groups, and each group maintains a private blockchain to store the data aggregation records, which avoids the leakage of the aggregated results and ensures the traceability of the sensory data. Then, a zero-sum noise-adding mechanism is utilized to not only preserve the private information during aggregation and ensure the correctness of the aggregated results but also improve the efficiency of privacy preservation. Furthermore, we theoretically prove the correctness, privacy, efficiency, and reliability of the proposed RPPDA algorithm. Real-world and simulated experiments demonstrate the effectiveness and advantages of the proposed RPPDA algorithm in terms of correctness, efficiency, and privacy.

1. Introduction

Mobile crowdsensing is a data acquisition paradigm based on crowdsourcing and the sensing capabilities of intelligent devices. Portable mobile devices constitute interactive and participatory intelligence sensing networks. The crowd in the network collaboratively completes the sensing tasks to achieve the purpose of data collection and information sharing [1–3]. Data aggregation is an essential prerequisite for data collection and information sharing in mobile crowdsensing networks. Data aggregation eliminates redundant information and extracts valuable information by processing local sensing data [4]. In smart grid applications, electricity consumption data is the basis for power companies to adjust power supply and demand control in real-time. The electric meter can sense users' electricity consumption data in real-time and aggregate it into regional electricity consumption data [5]. In intelligent transportation applications, road condition information is the basis for public travel and route planning. Mobile vehicles can sense traffic data in real-time and aggregate it to form regional traffic information, such as the user's location and speed [6].

The risk of privacy leakage exists during the data aggregation process in mobile crowdsensing networks [7]. In a crowdsensing network, the data perceived by mobile devices (such as electricity data and location data) is often sensitive [8]. Attackers may use sensitive data to speculate on the

user's living habits and behavioral patterns and then carry out malicious attacks. Therefore, preserving user's data privacy in the data aggregation process is the key to promoting the application of crowdsensing networks [9]. Existing research has carried out in-depth research on privacy-preserving data aggregation. Various privacy-preserving data aggregation schemes have been proposed [10–12]. Based on the homomorphic encryption mechanism, the sensing data can be computed directly on the ciphertext [13, 14]. Based on the random noise-adding mechanism, the sensing data can be computed after hiding the real data, thus achieving privacy-preserving data aggregation [15–17]. However, the existing privacy-preserving data aggregation schemes still face the problem of unreliable data aggregation processes. Crowdsensing nodes are deployed in the public environment, and their capabilities are limited. It is easy for a network attacker to control the nodes in the crowdsensing network and add illegal or fake data during the data aggregation process [18, 19]. Therefore, it is crucial to achieve data traceability and reliability of aggregation results in a privacy-preservation data aggregation process.

As a decentralized network public ledger, blockchain has the characteristics of decentralization, non-tampering, and traceability [20]. The unique trust management method of blockchain provides a promising way to study the methods of reliable privacy-preserving data aggregation [21–23]. In the process of the blockchain-assisted data aggregation, nodes jointly maintain an immutable transaction record of sensing data to realize the traceability of sensing data and improve the reliability of data aggregation results [24–26].

However, direct use of blockchain for privacy-preserving data aggregation may increase the risk of privacy leakage. Blockchain-assisted data aggregation requires nodes to jointly maintain transaction records. All nodes (including malicious nodes) can obtain the information contained in transaction records, which leads to privacy leakage in the process of data aggregation. Furthermore, privacy-preserving data aggregation is necessary to ensure the privacy of sensing data and intermediate calculation results. All nodes in the crowdsensing network should only know their private data and know nothing about the intermediate calculation results. Therefore, it is important to deeply study the blockchain-assisted privacy-preserving data aggregation method for improving reliability.

To solve the above problem, we propose a grouping-based reliable privacy-preserving data aggregation (RPPDA) algorithm. First of all, crowdsensing nodes are grouped to construct a private blockchain during the data aggregation process. The traceability and reliability of aggregation results are ensured. Then, a zero-sum noise-adding mechanism is introduced to preserve the privacy of data aggregation within each group. The communication and computation costs are reduced compared to the encryption mechanism-based data aggregation. The correctness of data aggregation is also ensured since the added random noise is canceled during the aggregation process. Therefore, the RPPDA algorithm ensures that the blockchain is fairly utilized to improve the reliability of data aggregation and avoid the effect of directly utilizing the blockchain on privacy-preserving data aggregation. The zero-sum noise-adding mechanism ensures the correctness, privacy, and efficiency of privacy-preserving data aggregation. The contributions of this paper are summarized as follows:

(1) We explore a new way to address the privacy leakage issue of directly applying blockchains to privacy-preserving data aggregation. A grouping-based, reliable, privacy-preserving data aggregation algorithm, RPPDA, is proposed using private blockchains for mobile crowdsensing.

(2) We conduct theoretical analyses to prove the privacy, correctness, and reliability of the RPPDA algorithm. The communication and computation costs are also analyzed.

(3) The correctness of aggregation results and the efficiency of privacy preservation are verified using experiments on the platform of Hyperledger Fabric. The advantages of RPPDA are also demonstrated through simulations in terms of correctness, execution time, and privacy preservation.

The remainder of the paper is organized as follows: Section 2 reviews the related work on privacy-preserving data aggregation. Section 3 describes the system model and problem setup. Section 4 designs the RPPDA algorithm. Section 5 theoretically analyzes the performance of RPPDA. Section 6 evaluates the performance of RPPDA through experiments and simulations. Section 7 concludes the paper.

2. Related Work

2.1. Privacy-Preserving Data Aggregation. Existing work has proposed various privacy preservation mechanisms to address the problem of privacy leakage during mobile crowdsensing data aggregation [27–29]. Existing privacy preservation mechanisms for data aggregation usually meet the following requirements. (1) Aggregation nodes can obtain aggregated results of data from all sensing nodes. (2) Aggregation nodes cannot obtain the private data of sensing nodes. (3) The sensing node cannot obtain the sensing data of other nodes. Existing work has designed privacy preservation methods for data aggregation from different perspectives based on encryption mechanisms or noise-adding mechanisms. For example, for the multidimensional data aggregation problem in the Internet of Things (IoT) scenario, Peng et al. designed an efficient privacy-preserving data aggregation method based on a homomorphic encryption method to protect data privacy during data aggregation [30]. Shi et al. designed a zero-sum weighted noise to address the problem of location information leakage during distributed localization, achieving privacy preservation and accuracy of localization results [31]. Wang et al. designed a local differential privacy data aggregation method based on differential privacy theory to solve the privacy disclosure problem of user data in crowdsensing discrete distribution estimation [32]. Xie and Chen designed a secure data aggregation framework based on the Shamir secret sharing mechanism to solve the

problem of data privacy disclosure in distributed data aggregation [33].

The above studies designed privacy preservation data aggregation methods by introducing homomorphic encryption mechanisms, zero-sum random noise, differential privacy noise, and the Shamir secret sharing mechanism. However, the reliability of privacy-preserving data aggregation still needs improvements. If some nodes use illegal or false data in the data aggregation process, it is necessary to enable the traceability of the data source. In this paper, the blockchain technology is used to achieve the reliability of the privacy-preserving data aggregation process. The proposed RPPDA algorithm ensures the traceability of the data aggregation process to improve the reliability of the data aggregation results.

2.2. Blockchain-Based Privacy Preservation. Blockchain-assisted data aggregation has been studied to improve the performance of crowdsensing [34]. The security and integrity are improved by utilizing blockchains for data aggregation in crowdsensing. For example, Wang et al. proposed a blockchain-based secure data aggregation strategy, which considers security level-based task classification and energy-efficient task fulfillment against privacy disclosure [24]. Li et al. presented a blockchain-enhanced data aggregation framework for UAV-assisted WSNs [25]. By combining blockchain building and data aggregation, a disaster semantic blockchain based on a data reconstruction-directed consensus mechanism is presented to ensure the security of data transmission. Arulprakash and Jebakumar designed and implemented a real-time privacy-preserving data aggregation distribution scheme for mobile crowdsensing called SMARTEE [26]. The SMARTEE stores and transfers data using a protected blockchain mechanism and a gateway-based authentication scheme to ensure proper data transmission and integrity. Therefore, it is noted that blockchain offers a novel way to enhance privacy preservation during data aggregation in crowdsensing.

Privacy-preserving data aggregation schemes have been investigated using blockchains [35]. Existing work designed privacy-preserving data aggregation from different perspectives to avoid the single point failure and the use of false data in the process of data aggregation. For instance, Zhang et al. designed a privacy-preserving and a reliable sensing scheme using a homomorphic encryption mechanism and blockchain to avoid a single point of failure and achieve reliable data aggregation services [36]. Kong et al. designed a privacy-preserving and verifiable data sharing scheme using homomorphic encryption mechanism and blockchains to achieve privacy protection and verifiability [37]. Lin et al. designed a secure data aggregation strategy by putting node security level indicators into blockchains [38]. The strategy improves the security level of data aggregation in industrial applications. Guan et al. designed a privacy-preserving multi-party computing mechanism based on blockchains to improve the security of multi-party computing [39].

The above work improves the security and reliability of the data aggregation process. However, the above privacy preservation schemes based on blockchains are different from our reliable privacy preservation method. The privacy-preserving data aggregation approach using homomorphic encryption mechanisms may incur a significant communication and computational cost. In this paper, a noise-adding mechanism is utilized to preserve privacy during data aggregation and reduce the communication and computation costs. A grouping-based reliable privacy preservation method is designed to enrich the application of the blockchain technology.

3. System Model and Problem Setup

3.1. System Model. We consider a mobile crowdsensing network consisting of four parties, including a target node, many sensing nodes, several aggregation nodes, and a blockchain network. First, to collect data and acquire information, such as with the smart grid crowdsensing, the target node distributes sensing tasks to suitable sensing nodes. After adding random noise to the sensing results, the sensing nodes send the noise-added sensing results to the designated blockchain network. Then, by calculating the aggregation result based on the noise-added data, the aggregation node sends the aggregation result to the target node. Figure 1 shows the system model of the blockchain-assisted data aggregation in mobile crowdsensing.

(1) *Target node.* The target node is the requestor of the crowdsensing data and distributes the crowdsensing tasks to sensing nodes by crowdsensing platforms or application servers. The target node obtains the aggregation results based on the sensing data.

(2) *Sensing node.* A sensing node is a smart mobile device with the basic functionalities of sensing, computing, and storage. The sensing nodes could be smart phones, wearable devices, and mobile vehicles. Sensing nodes are the basic units of the mobile crowdsensing networks, which are used to collect different types of data, such as power consumption, temperature, location, and speed. Without affecting the aggregation results, zero-sum random noise is usually added to the sensing data to preserve the data privacy of the sensing node.

(3) *Blockchain network.* The blockchain network is a distributed storage ledger that records the noise-added sensing data in sequence and is maintained by distributed sensing nodes. Under a consensus process, the sensing nodes take the noise-added sensing data as a transaction and pack it into the block. The blockchain network automatically invokes a smart contract containing the aggregation algorithm to output the aggregation result to the aggregation node.

(4) *Aggregation node.* The aggregation node is responsible for aggregating the sensing data and sending the aggregation result to the target node. In this model, a sensing node can participate in the data aggregation process so as to be an aggregation node.

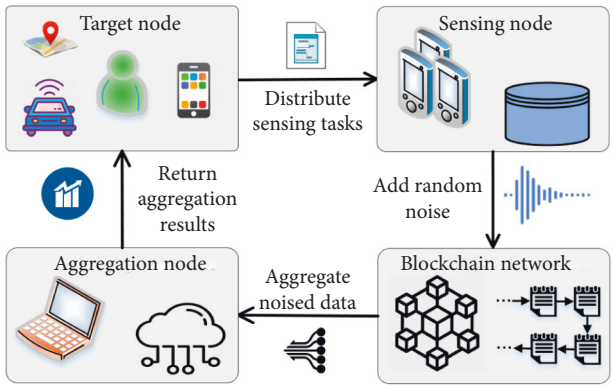

Figure 1: System model of the blockchain-assisted data aggregation in mobile crowdsensing.

In mobile crowdsensing networks, N_0 denotes the target node. The sensing node is denoted as $N_i\,(i = 1, 2, \ldots, m)$, where m is the number of sensing nodes. N_i takes the sensing data as its private data $d_i\,(i = 1, 2, \ldots \infty, m)$. All sensing nodes are divided evenly into n groups. Each group has an aggregation node $N_j'\,(j = 1, 2, \ldots, n)$ responsible for collecting intermediate aggregation results of the sensing data within the group. N_0 is responsible for collecting intermediate aggregation results and calculating final sensing data aggregation results. Therefore, the purpose of privacy-preserving data aggregation is to compute correct-sensing data aggregation results while preserving privacy.

In this paper, a semi-honest model is used to analyze the threat in the process of privacy-preserving data aggregation. Assuming that all nodes in the data aggregation process are honest but curious. Each node performs computation and communication according to the steps specified in the algorithm. But each node is curious whether it can deduce the private information of other nodes from the messages it obtained [40]. Moreover, data summation in crowdsensing is an important data aggregation scenario. Therefore, we focus on how to use blockchain to enhance the reliability of private data and to avoid information leakage during aggregation results due to the use of public blockchain. Furthermore, during the crowdsensing data aggregation process, the sensing data of the nodes need to be aggregated without considering the movement of the nodes. The sensing data of a moving node can be broadcast and uploaded to the blockchain through the nearby base stations [41]. Therefore, it is assumed that the moving nodes have no impact on the privacy-preserving data aggregation.

3.2. Problem Setup. The reliable and privacy-preserving data aggregation algorithm based on a private blockchain is desired to meet the following objectives:

(1) *Correctness of aggregation results.* The data aggregation results of all sensing nodes can be correctly calculated, that is, the algorithm execution results are equal to the direct calculation results without considering privacy by adding random noise.

(2) *Privacy of data.* It is ensured that private data d_i is known only by the node N_i. The final sensing data aggregation result is known only by the target node N_0, and the intra-group aggregation result is known only by intra-group nodes.

(3) *Efficiency.* The designed algorithm does not use any encryption mechanism. Compared with the privacy-preserving data aggregation algorithm based on the encryption mechanism, the designed algorithm has low computation and communication costs.

(4) *Reliability.* Compared with the privacy-preserving data aggregation algorithm using public blockchain, the designed algorithm can calculate reliable results by the grouping process. Each group maintains a private blockchain to store the data aggregation records, which reduces the risk of privacy leakage from directly utilizing the public blockchain.

4. Reliable Privacy Preserving Data Aggregation

4.1. Basic Idea. In the process of mobile crowdsensing data aggregation, RPPDA is designed to realize the correctness, efficiency, privacy preservation, and reliability of aggregation results. Firstly, all sensing nodes are divided into groups to maintain private blockchains and carry out privacy-preserving data aggregation based on the zero-sum noise-adding mechanism. Data transactions during the aggregation are not completed until they are posted to a private blockchain, ensuring that data transactions can be traced back. Secondly, the aggregation node obtains the aggregation result within the group. Then, to preserve the privacy of the intermediate aggregation results, the aggregation nodes again add zero-sum noise to the intermediate aggregation results while waiting for the target node to perform secondary data aggregation. Finally, the target node uses the noise-added intermediate aggregation results to compute the final aggregation result. Due to the zero-sum noises, the effect of noise on the aggregation results can be canceled, which ensures the correctness of the aggregation result.

The idea of the reliable privacy preservation using private blockchains is further discussed. The blockchain construction process is a key step in the implementation of RPPDA. A private blockchain takes a partially centralized approach and is only open to specific individuals. Data reads, data writes, and consensus mechanisms comply with the private blockchain manager, which maximizes the maintenance of sensitive data from illegal access and tampering [42]. In particular, the sensing nodes are divided into multiple groups, and each group maintains a private blockchain. The block has the following components: merkle root, block ID, hash, previous hash, time stamp, and the noise-added sensing data. Furthermore, the private blockchain reduces the risk of the privacy leakage issue of utilizing the public blockchain. When using the public blockchain, the aggregation results are available to any node, which may reduce the privacy preservation level of the data aggregation process. However, by storing the noise-added sensing data on the private blockchain in each group, it ensures the

correctness of data aggregation results, preserves the privacy of sensing data, and improves the reliability of privacy-preserving data aggregation.

4.2. Intra-Group Data Aggregation. In the process of intra-group data aggregation, the sensing nodes add zero-sum noise to realize the privacy preservation for sensing data and the correctness of data aggregation results. At the same time, the key to intra-group data aggregation is maintaining a private blockchain. In the process of aggregating noise-added data, the data transaction can be published on the blockchain only after the consensus algorithm is implemented, which ensures the reliability and traceability of data. Therefore, two steps are considered for the intra-group data aggregation based on blockchain, including privacy-preserving data aggregation based on noise-adding mechanism and consensus mechanism of data transaction on the blockchain during data aggregation.

4.2.1. Noise-Adding based Privacy Preservation. It is crucial to ensure the correctness of aggregation results and the privacy of the data. During data aggregation, zero-sum noise is added to eliminate the effect of noise on the aggregation result and achieve accurate, privacy-preserving data aggregation.

The privacy-preserving data aggregation process based on the zero-sum noise is described below [40]. Assuming N_i has a private matrix $M_i (i = 1, \ldots, m)$, The size of the private matrix M_i is predetermined based on the crowdsensing data aggregation scenario. For example, in the three-dimensional crowdsourced localization scenario, the sensing data denotes the location information, including three rows and one column. N_0 needs to obtain the summation without knowing the private information of N_i $(\alpha = \sum_{i=1}^{m} M_i)$. Random noise matrices are used to preserve privacy information during summation calculation. Firstly, N_i generates m random noise matrices $p_i^k (k = 1, \ldots, m)$. p_i^k has the same number of rows and columns as the matrix M_i and the sum of these matrices is a zero matrix. Secondly, N_i keeps one of the matrices and sends the remaining matrices one by one to the other nodes. Therefore, N_i can add up the random matrices received from other nodes to construct a new random noise matrix, denoted as P_i, where $\sum_{i=1}^{m} P_i = 0$. Finally, N_i sends mixed messages $\alpha_i = M_i + P_i$ to N_0. Therefore, without obtaining the private information of the other nodes, N_0 calculates the summation of all the private information by $\alpha = \sum_{i=1}^{m} \alpha_i$.

4.2.2. Consensus Mechanism. Data privacy of nodes on the blockchain is protected by adding the zero-sum noise, and all nodes are semi-honest. Therefore, the nodes in the group are not anonymous when they publish the block storing personal noise-added information as miners. Assuming that the average block-producing time of the blockchain within the group is consistent.Because the number of nodes in the group is small, the delay of block propagation is ignored. The block is broadcast and immediately received by the other nodes in the group. After a block is added to the blockchain, each node will receive blocks broadcast by other nodes. The nodes will verify the validity of these blocks and select a block to be added to the end of the locally maintained blockchain through a consensus mechanism.

In the process of data aggregation, the consensus mechanism of data transaction polling is described as follows. First, NUM_i denotes the total number of blocks published by miner i. T_i denotes the time between the miner i publishing the last block and the current block. T_i can be obtained by subtracting the timestamp in the block published by the previous miner i from the timestamp in the block packed by the current miner i. Each miner will sort the blocks received in T seconds according to the polling consensus mechanism and select a block to add to the end of the locally maintained blockchain. Then, the consensus mechanism compares the total number of blocks that miners have published on the blockchain. The output is the block packaged by the miner who has published the least number of blocks. When there are multiple blocks with equal NUM values in this step, all of them have a minimum value. The time after the miner who packed these blocks last published a block is compared in the blockchain. The blocks packed by the miner with the longest wait time is the output. Further, if the above steps fail to output a single block, the hash values of these blocks are compared to the output of the block with the lowest hash value.

4.3. Algorithm Design. The aggregation node calculates the intra-group data aggregation result and performs the zero-sum noise adding mechanism to ensure the privacy and correctness of the aggregation results. Therefore, based on the idea of adding zero-sum noise in Section 4.2.1, the intra-group data aggregation results are aggregated again by adding zero-sum noise to achieve privacy-preserving aggregation among groups.

Based on intra-group data aggregation and data aggregation between groups, the RPPDA algorithm is designed in Algorithm 1. The input is private data $d_i (i = 1, 2, \ldots, m)$ for sensing nodes. The output is the final sensing data aggregation result. First, the sensing nodes are grouped to avoid information disclosure of final aggregation results caused by the public blockchain. The m sensing nodes involved in the aggregation calculation are divided equally into n groups, denoted by $G_j (j = 1, 2, \ldots, n)$. Each group has $k (k \geq 3)$ nodes $(N_{j1}, N_{j2}, \ldots, N_{jk})$. One node from each group G_j is selected as the aggregation node N_j' of the group. Second, the zero-sum noise is added to private sensing data. The nodes in group G_j are denoted by $N_{jl} (l = 1, 2, \ldots, k)$. The nodes collaborate to generate the zero-sum noise α_{jl}, where $\sum_{l=1}^{k} \alpha_{jl} = 0$. Then, the noise-added data (i.e. transaction data records) is uploaded to the blockchain by the consensus mechanism. Each group G_j maintains a private blockchain BC_j. N_{jl} adds zero-sum noise α_{jl} to private data d_{jl}, according to $d_{jl}' = d_{jl} + \alpha_{jl}$. The node identification information and noise-added data d_{jl}' are stored in the private blockchain BC_j. Next, the aggregation node N_j' aggregates the noise-added data d_{jl}' in the private blockchain BC_j,

according to $\text{Sum}_j = \sum_{l=1}^{k} d'_{jl} = \sum_{l=1}^{k} (d_{jl} + \alpha_{jl})$. Further, the aggregation nodes N'_j collaborate to generate the zero-sum noise β_j again, where $\sum_{j=1}^{n} \beta_j = 0$. N'_j adds zero-sum noise β_j to the intra-group data aggregation result Sum'_j, according to $\text{Sum}'_j = \text{Sum}_j + \beta_j$. Finally, the target node N_0 calculates the final aggregation result $\text{Sum} = \sum_{j=1}^{n} \text{Sum}'_j = \sum_{j=1}^{n} (\text{Sum}_j + \beta_j)$.

5. Algorithm Analyses

5.1. Correctness. The correctness of RPPDA means that the execution result is equal to the direct aggregation result without adding random noise by considering privacy. It is necessary to prove that the final aggregation result calculated by the target node N_0 is equal to the sum of the private data of all sensing nodes N_i. The following theorem gives the correctness of RPPDA.

Theorem 1. *Consider m nodes $N_i (i = 1, 2, \ldots, m)$ participating in the summation computation. The private data of all sensing nodes is $d_i (i = 1, 2, \ldots, m)$. Then, the following equation holds.*

$$S = \text{Sum}, \tag{1}$$

where S is the sum of the private data of all sensing nodes N_i. Sum *is the final aggregation result calculated by the target node N_0.*

Proof. Firstly, we need to prove the correctness of the intra-group summation calculation results Sum_j of the aggregation node. S_j denotes the sum of the private data of the sensing nodes in G_j. According to $S_j = \sum_{l=1}^{k} d_{jl}$ and $\sum_{l=1}^{k} \alpha_{jl} = 0$, we have

$$\text{Sum}_j = \sum_{l=1}^{k} d'_{jl} = \sum_{l=1}^{k} \left(d_{jl} + \alpha_{jl}\right) = \sum_{l=1}^{k} d_{jl} + \sum_{l=1}^{k} \alpha_{jl} = \sum_{l=1}^{k} d_{jl} = S_j, \tag{2}$$

where α_{jl} is the zero-sum noise of the l-th sensing node in G_j. d'_{jl} is the noise-added data of N_{jl}. Then, it is necessary to prove that the summation results, Sum of the target node is equal to the sum of the intra-group summation calculation results. Due to $\sum_{j=1}^{n} \beta_j = 0$, we have

$$\text{Sum} = \sum_{j=1}^{n} S'_j = \sum_{j=1}^{n} \left(\text{Sum}_j + \beta_j\right) = \sum_{j=1}^{n} \text{Sum}_j + \sum_{j=1}^{n} \beta_j = \sum_{j=1}^{n} \text{Sum}_j, \tag{3}$$

where β_j is the zero-sum noise of the aggregation node in G_j. S'_j is the noise-added summation data of G_j. Combining equations (2) and (3), we know that

$$\text{Sum} = \sum_{j=1}^{n} Sum_j = \sum_{i=1}^{m} d_i = S. \tag{4}$$

Therefore, Theorem 1 is proved. The final aggregation result of RPPDA is correct. □

5.2. Privacy Preservation and Reliability

5.2.1. Privacy Preservation. The privacy preservation of RPPDA means that the private data of all nodes involved in the calculation cannot be deduced by other nodes, and only the target node knows the final aggregation result. The following theorem gives the privacy preservation of the RPPDA.

Theorem 2. *When $m \geq 9$, $n \geq 3, k \geq 3$, RPPDA holds privacy preservation.*

Proof. Two cases need to be proved to support Theorem 2. (1) All nodes cannot directly or indirectly reason to obtain the private data of other nodes; (2) The final summation result can only be computed by the target node.

In the process of data aggregation, each private blockchain stores the noise-added data of sensing nodes. Meanwhile, at least three nodes within each group collaborate to generate zero-sum noise because during the execution of RPPDA, a zero-sum noise addition is performed within each group and between all groups, respectively. Also, the zero-sum noise mechanism needs to ensure that the private information is not inferred by other nodes when three or more nodes are present. For example, when there are only two nodes, a node can reason out the private information of another node based on its zero-sum noise term. Therefore, even nodes within the same group cannot obtain and infer the private data of other nodes from the noise-added data. In addition, the aggregation nodes of each group collaborate to generate the zero-sum noise again and add the noise to the intermediate calculation results. Then, the noise-added data is treated as private data. Therefore, when the target node communicates with the aggregator node, the target node gets the noise-added intermediate computation result. The target node does neither have access to the private data of the aggregator node nor can it infer the private data of the aggregator node from the noise-added intermediate computation results. Condition 1 holds true.

For condition 2, all nodes involved in the computation are divided into groups. The nodes in each group store their private data in a private blockchain after adding zero-sum noise. The summation result of the data for each group can only be calculated by the node within that group, i.e., the aggregation node can get the summation result of each group (the intermediate calculation result). The target node communicates with the aggregation node and obtains the noise-added intermediate results of each group. Then the target node calculates the final summation result. In conclusion, Theorem 2 is proved. □

5.2.2. Reliability. The reliability of RPPDA is to ensure the traceability of original sensing data in the process of data aggregation. Many factors influence sensing data in the mobile crowdsensing network. For example, mobile device failure leads to data storage failure and malicious node attacks lead to data errors. RPPDA is based on private blockchain. All nodes are honest but curious and follow the

Input: The private information $d_i\,(i = 1, 2, \ldots, m)$ of node N_i.
Output: The final aggregation result $\text{Sum} = \sum_{i=1}^{m} d_i$.
(1) $N_i\,(i = 1, 2, \ldots, m)$ are grouped equally into $G_j\,(j = 1, 2, \ldots, n)$.
(2) **For** $j = 1$ to n**do**
(3) $N_j^{'}$ is elected from G_j.
(4) The nodes $\{N_{j1}, N_{j2}^{'}, \ldots, N_{jk}\}$ collaborate to generate the zero-sum noise $\{\alpha_{j1}, \alpha_{j2}, \ldots, \alpha_{jk}\}$.
(5) **For** $l = 1$ to k**do**
(6) $d_{jl}^{'} = d_{jl} + \alpha_{jl}$.
(7) Store the noise-added data $d_{jl}^{'}$ in the private blockchain BC_j.
(8) **End**
(9) $\text{Sum}_j = \sum_{l=1}^{k} d_{jl}^{'}$.
(10) **End**
(11) The nodes $\{N_1^{'}, N_2^{'}, \ldots, N_n^{'}\}$ collaborate to generate the zero-sum noise $\{\beta_1, \beta_2, \ldots, \beta_n\}$.
(12) **For** $j = 1$ to n**do**
(13) $\text{Sum}_j^{'} = \text{Sum}_j + \beta_j$.
(14) **End**
(15) $\text{Sum} = \sum_{j=1}^{n} \text{Sum}_j^{'} = \sum_{j=1}^{n} (\text{Sum}_j + \beta_j)$.

Algorithm 1: Reliable privacy-preserving data aggregation (RPPDA).

consensus mechanism of data transaction polling during the data aggregation. The sensing data in mobile crowdsensing is published to the blockchain after adding random noise. The sensing data can be traced back through blockchain when there are mobile device failures or being attacked. In addition, sensing data cannot be tampered with or deleted because the blockchain is jointly maintained by all nodes. Therefore, RPPDA ensures the traceability of original sensing data and improves the reliability of data aggregation in mobile crowdsensing.

5.3. Communication and Computation Cost

(1) *Communication Cost.* The communication cost of the RPPDA consists of three components, i.e., the communication cost between the sensing nodes, the communication cost between the sensing nodes and the aggregation nodes, and the communication cost between the aggregation nodes and the target node. Because the number of nodes in each group is equal, it is assumed that the process of adding noise and requesting to join the blockchain takes place simultaneously in each group. For convenience, we consider the communication cost of only one group. Firstly, to generate zero-sum noise, each node in the group needs to transmit $(k - 1)$ elements to complete the communication. The communication cost in the private blockchain is T_{bc}. Next, to get the data on the private chain and perform aggregation, the aggregation node needs to communicate with the other nodes in the group by transmitting $(k - 1 + T_{bc})$ elements. Finally, to obtain the intermediate aggregation results and perform the final aggregation, the target node communicates with the aggregation node by transmitting n elements. Assuming that a number is represented by 24 bits. In summary, the communication cost of the aggregation process for RPPDA is roughly $(24 \times (n(k-1) + n(k - 1 + T_{bc}) + n))$.

(2) *Computation Cost.* Assuming that there are m nodes involved in the aggregation. The nodes are divided equally into n groups. Each group has k nodes. There are k addition operations to be performed when each group of nodes adds noise to the private data. To sum up the noise-added data of each group and add the noise to it again, the aggregation node needs to perform $(k - 1 + n)$ addition operations. The target node communicates with the aggregation node to compute the final summation result and perform $(n - 1)$ addition operations. Therefore, the computation cost of the summation calculation process for RPPDA is roughly $(n(k\varphi + C_{bc} + (k - 1 + n)\varphi + (n - 1)\varphi))$, where φ is the cost of addition operations and C_{bc} is the computation cost in the private blockchain.

6. Performance Evaluations

In this section, to evaluate the performance of the RPPDA, the experiments and simulations are designed for the mobile crowdsensing data aggregation scenario.

6.1. Experimental Results. The comparison schemes are the noise-adding based privacy-preserving data aggregation (NPPDA) and homomorphic encryption-based privacy-preserving data aggregation (HPPDA). NPPDA achieves privacy-preserving data aggregation by adding zero-sum noise and does not use blockchain [31]. HPPDA uses the Paillier homomorphic encryption method to achieve privacy-preserving data aggregation [13, 30]. In the experiment, the number of sensing nodes was set from 10 to 100, and 10 nodes were added each time. To reduce the influence of random variables on the results, the results of all experiments are the average of 1,000 independent runs.

The RPPDA algorithm was deployed to the Hyperledger Fabric blockchain platform (version 2.1.1) to evaluate the performance. Experimental tests were carried out regarding the efficiency of privacy preservation. The configuration information for the experimental environment parameters is shown in Table 1. The HPPDA uses the Paillier homomorphic encryption system based on the implementation using the Go language.

The execution time of HPPDA, RPPDA, and NPPDA algorithms is shown in Figure 2. It is seen that as the number of nodes increases, the execution time of the three algorithms increases. Compared with HPPDA, the execution time of RPPDA was reduced by 56.84% on average. Since the HPPDA algorithm performs a homomorphic encryption mechanism, the time overhead is large. However, NPPDA has a small time overhead because it does not use blockchain. Therefore, the experimental results are reasonable and prove that RPPDA can improve efficiency while achieving reliable privacy protection.

6.2. Simulation Results. Simulations are conducted for the RPPDA algorithm using Pycharm software running on a Dell desktop computer with an Inter core i5-9500 CPU @ 3.00 GHz processor and 8.00 GB (7.81 GB available) of RAM. The HPPDA, RPPDA, and NPPDA algorithms are compared in terms of aggregation accuracy, communication and computation cost, and privacy preservation strength. The simulation experiment considers the smart grid crowdsensing scenario, where the electricity consumption of users is collected in an area for a week. About 100 smart meters are distributed in the area. The smart meter is the sensing node, and the user's electricity consumption is the sensing data. The number of sensing nodes was set from 10 to 100. The electricity consumption of users ranges from 0 to 100. To reduce the influence of random variables on the results, the results of all simulations are the average of 1,000 independent runs.

Figure 3 shows the data aggregation results with and without noise to evaluate the correctness of RPPDA. It can be seen that the summation results with noise are always the same as the summation results without noise. The reason is that the zero-sum random noise is added in the RPPDA data aggregation process, which counteracts the effects of noise. Therefore, the RPPDA algorithm is able to correctly aggregate the data, which validates the theoretical results in Theorem 1 of Section 5.1.

The computation and communication costs of HPPDA, NPPDA, and RPPDA are shown in Figures 4 and 5. The comparison shows that the computation and communication costs of RPPDA are on average 94.03% and 99.86% lower than those of HPPDA, respectively. This is because HPPDA uses homomorphic encryption containing long-bit products and exponential operations. The computation and communication costs of RPPDA are higher than NPPDA since RPPDA uses a private blockchain to record data transactions. Therefore, RPPDA ensures the efficiency of the data aggregation process.

The privacy protection strength is evaluated for the RPPDA algorithm. Inspired by the definition of privacy protection strength in existing studies [40, 43, 44], the privacy protection strength is measured by the probability that a node's private data is not accessed by other nodes in the simulations. For example, the privacy protection strength is defined by the chance of activity prediction of users in location-based services (LBS) [43]. A smaller chance of activity prediction means a higher privacy

Table 1: Parameter information for the experiments on hyperledger fabric.

Component	Description
Node	V12.18.0
NPM	V6.14.0
Golang	12.8
Docker compose	go1.15.6 linux/amd64
Docker engine	1.22.0
CPU	Intel(R) Xeon(R) Silver 4210 (2.20 GHz) CPU
Operating systems	CentOS Linux release 7.8.2003
Memory	190G RAM

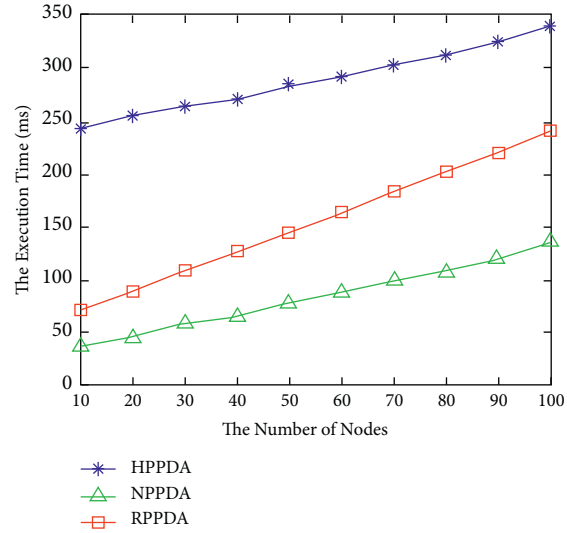

Figure 2: The execution time (ms) comparison among HPPDA, RPPDA, and NPPDA.

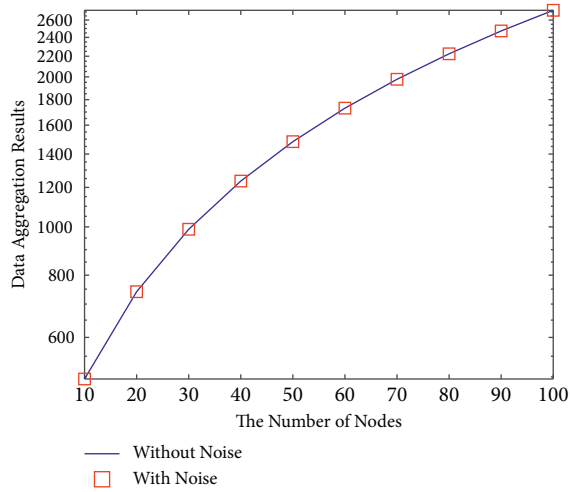

Figure 3: The data aggregation results with and without noise.

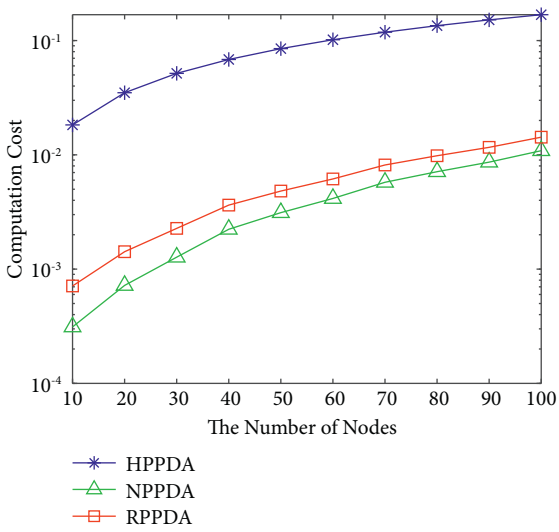

Figure 4: The computation cost of HPPDA, NPPDA, and RPPDA.

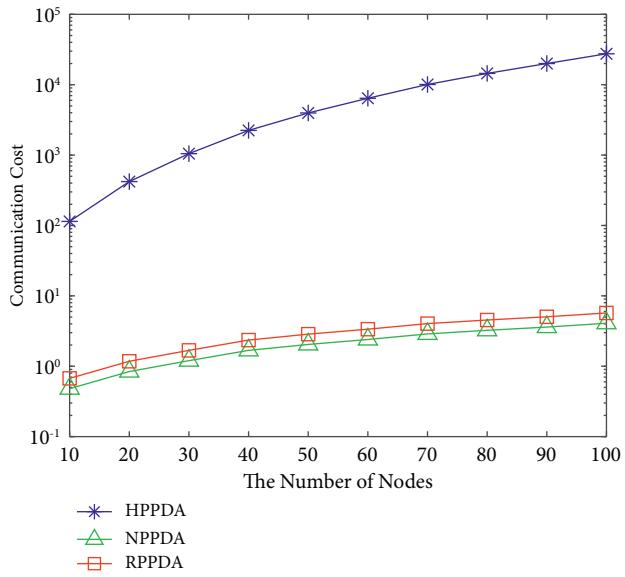

Figure 5: The communication cost of HPPDA, NPPDA, and RPPDA.

protection strength. Similarly, data aggregation is privacy-preserving when the private data of the sensing node cannot be known by other nodes. Due to the public and transparent nature of the blockchain, the data stored in the blockchain can be accessed by other nodes. When m nodes are involved in aggregation and divided into n groups, the number of nodes in each group is m/n, where the privacy protection strength is $1 - (1/mn)$. Under the above definition of privacy protection strength, Figure 6 compares the privacy protection strength under the different number of groupings. It can be found that when the number of sensing nodes involved in aggregation is certain, the more the number of groupings, the higher the privacy protection strength because the greater the number of groupings, the

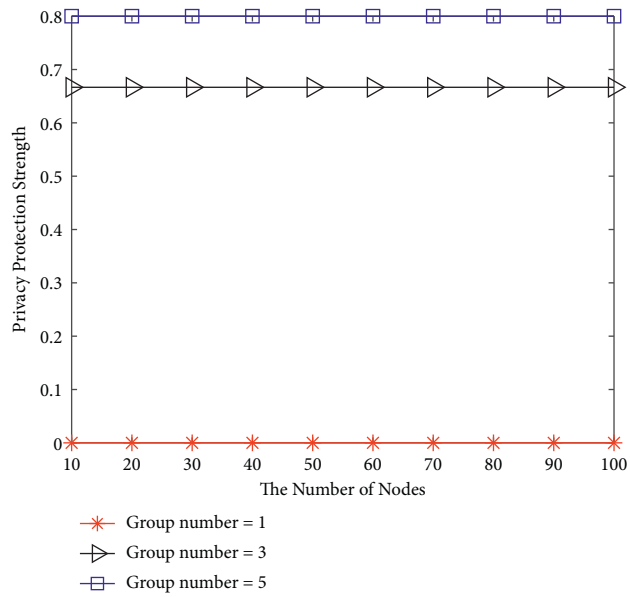

Figure 6: The privacy protection strength in the different number of groupings.

fewer the number of nodes that jointly maintain a blockchain. Therefore, grouping nodes can improve the privacy-preserving strength of the data aggregation process.

According to the evaluation results of the above experiments and simulations, RPPDA can ensure the correctness of data aggregation results. RPPDA has a lower computation and communication cost than HPPDA. RPPDA also improves the privacy preservation strength by dividing nodes into groups to avoid the privacy leakage of directly using blockchain. Therefore, the RPPDA algorithm not only ensures the correctness of the data aggregation results but also has higher efficiency and strong privacy preservation.

7. Conclusions

Privacy-preserving data aggregation play an important role in mobile crowdsensing. Direct use of the public blockchain can improve the reliability of privacy-preserving data aggregation. However, there is the risk of privacy leakage during the process of data aggregation. In this paper, we propose a grouping-based reliable privacy-preserving data aggregation algorithm to avoid the risk of privacy leakage. First, we introduce a private blockchain to ensure traceability during data aggregation. All sensing nodes are divided into groups to maintain private blockchains. Then, the zero-sum noise mechanism is utilized to ensure the correctness of data aggregation results, preserve the privacy of sensing data, and improve the efficiency of data aggregation. Finally, the correctness, privacy-preservation, reliability, and efficiency of RPPDA are theoretically analysed. The effectiveness of RPPDA is demonstrated by experiments and simulations.

Data Availability

No data were used to support the findings of this study.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported in part by the Henan Provincial Major Public Welfare Project (201300210400), the China Postdoctoral Science Foundation (2020M672211 and 2020M672217), the Key Scientific Research Projects of Henan Provincial Colleges and Universities (21A520003), and the Key Technology Research and Development Program of Henan (182102210106, 212102210090, 212102210094, 212102210078, 222102210133, and 222102210055).

References

[1] Y. Wu, J. R Zeng ., H. Peng, H. Chen, and CP. Li, "Survey on incentive mechanisms for crowd sensing," *Journal of Software*, vol. 27, no. 8, pp. 2025–2047, 2016.

[2] A. Capponi, C. Fiandrino, B. Kantarci, L. Foschini, D. Kliazovich, and P. Bouvry, "A survey on mobile crowdsensing systems: challenges, solutions, and opportunities," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2419–2465, 2019.

[3] Y. Liu, L. Kong, and G. Chen, "Data-oriented mobile crowdsensing: a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 3, pp. 2849–2885, 2019.

[4] X. Yan, W. Ng, B. Zeng et al., "Verifiable, reliable, and privacy-preserving data aggregation in fog-assisted mobile crowdsensing," *IEEE Internet of Things Journal*, vol. 8, no. 18, pp. 14127–14140, 2021.

[5] S. Zhao, F. Li, H. Li et al., "Smart and practical privacy-preserving data aggregation for fog-based smart grids," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 521–536, 2021.

[6] Z. Wang, Y. Li, D. Li et al., "Enabling fairness-aware and privacy-preserving for quality evaluation in vehicular crowdsensing: a decentralized approach," *Security and Communication Networks*, vol. 2021, Article ID 9678409, 11 pages, 2021.

[7] Z. Shao, H. Wang, Y. Zou, Z. Gao, and H. Lv, "From centralized protection to distributed edge collaboration: a location difference-based privacy-preserving framework for mobile crowdsensing," *Security and Communication Networks*, vol. 2021, Article ID 5855745, 18 pages, 2021.

[8] T. Zhang, X. Song, L. Zheng, Y. Han, K. Zhang, and Q. Li, "Towards time-sensitive and verifiable data aggregation for mobile crowdsensing," *Security and Communication Networks*, vol. 2021, Article ID 6679157, 14 pages, 2021.

[9] M. Zhang, L. Yang, S. He, M. Li, and J. Zhang, "Privacy-preserving data aggregation for mobile crowdsensing with externality: an auction approach," *IEEE/ACM Transactions on Networking*, vol. 29, no. 3, pp. 1046–1059, 2021.

[10] Y. Liu, F. Liu, H. T. Wu, X. Zhang, B. Zhao, and X. Yan, "PriDPM: privacy-preserving dynamic pricing mechanism for robust crowdsensing," *Computer Networks*, vol. 183, Article ID 107582, 2020.

[11] Y. Liu, W. Guo, C. I. Fan, L. Chang, and C. Cheng, "A practical privacy-preserving data aggregation (3PDA) scheme for smart grid," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 3, pp. 1767–1774, 2019.

[12] J. Li, H. Ye, T. Li et al., "Efficient and secure outsourcing of differentially private data publishing with multiple evaluators," *IEEE Transactions on Dependable and Secure Computing*, vol. 19, no. 1, pp. 67–76, 2022.

[13] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proceedings of the Advances in Cryptology-EUROCRYPT '99, International Conference on the Theory and Application of Cryptographic Techniques*, pp. 223–238, Prague, Czech Republic, May 1999.

[14] J. Li, Y. Huang, Y. Wei et al., "Searchable symmetric encryption with forward search privacy," *IEEE Transactions on Dependable and Secure Computing*, vol. 18, no. 1, pp. 460–474, 2021.

[15] C. Fan, S. Huang, and Y. Lai, "Privacy-enhanced data aggregation scheme against internal attackers in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 1, pp. 666–675, 2014.

[16] J. He, L. Cai, C. Zhao, P. Cheng, and X. Guan, "Privacy-preserving average consensus: privacy analysis and algorithm design," *IEEE Transactions on Signal and Information Processing over Networks*, vol. 5, no. 1, pp. 127–138, 2019.

[17] Z. Zhou, Y. Su, Y. Zhang et al., "Coverless information hiding based on probability graph learning for secure communication in IoT environment," *IEEE Internet of Things Journal*, vol. 9, no. 12, pp. 9332–9341, 2022.

[18] Z. Wang, L. Liao, R. Meng, C. N. Yang, Z. Zhou, and H. Yang, "Verification grid and map slipping based graphical password against shoulder-surfing attacks," *Security and Communication Networks*, vol. 2022, Article ID 6778755, 9 pages, 2022.

[19] G. Wang, Y. Xu, J. He, J. Pan, F. Zuo, and X. He, "Resilient participant selection under vulnerability-induced colluding attacks for crowdsourcing," *IEEE Transactions on Vehicular Technology*, vol. 71, no. 7, pp. 7904–7918, 2022.

[20] Z. Chen, C. Fiandrino, and B. Kantarci, "On blockchain integration into mobile crowdsensing via smart embedded devices: a comprehensive survey," *Journal of Systems Architecture*, vol. 115, no. 102011, pp. 102011–102019, 2021.

[21] M. S. Ali, M. Vecchio, M. Pincheira, K. Dolui, F. Antonelli, and M. H. Rehmani, "Applications of blockchains in the internet of things: a comprehensive survey," *IEEE Communications Surveys & Tutorials*, vol. 21, no. 2, pp. 1676–1717, 2019.

[22] Z. Guo, Y. Zhang, FL. An et al., "Secure computing outsourcing scheme for polynomial with privacy protection based on blockchain," *Journal of Cyber Security*, vol. 6, no. 1, pp. 78–89, 2021.

[23] Z. Zhou, M. Wang, C. N. Yang, Z. Fu, X. Sun, and Q. J. Wu, "Blockchain-based decentralized reputation system in e-commerce environment," *Future Generation Computer Systems*, vol. 124, no. 13, pp. 155–167, 2021.

[24] X. Wang, S. Garg, H. Lin, G. Kaddoum, J. Hu, and M. S. Hossain, "A secure data aggregation strategy in edge computing and blockchain empowered internet of things," *IEEE Internet of Things Journal*, vol. 9, no. 16, pp. 14237–14246, 2022.

[25] G. Li, B. He, Z. Wang, X. Cheng, and J. Chen, "Blockchain-enhanced spatiotemporal data aggregation for UAV-assisted wireless sensor networks," *IEEE Transactions on Industrial Informatics*, vol. 18, no. 7, pp. 4520–4530, 2022.

[26] M. Arulprakash and R. Jebakumar, "People-centric collective intelligence: decentralized and enhanced privacy mobile crowd sensing based on blockchain," *The Journal of Supercomputing*, vol. 77, no. 11, pp. 12582–12608, 2021.

[27] M. S. Yousefpoor, E. Yousefpoor, H. Barati, A. Barati, A. Movaghar, and M. Hosseinzadeh, "Secure data aggregation methods and countermeasures against various attacks in wireless sensor networks: a comprehensive review," *Journal of Network and Computer Applications*, vol. 190, pp. 103118–103142, Article ID 103118, 2021.

[28] K. Sarwar, S. Yongchareon, J. Yu, and S. ur Rehman, "Lightweight, divide-and-conquer privacy-preserving data aggregation in fog computing," *Future Generation Computer Systems*, vol. 119, pp. 188–199, 2021.

[29] T. Wan, S. Yue, and W. Liao, "Privacy-preserving incentive mechanism for mobile crowdsensing," *Security and Communication Networks*, vol. 2021, Article ID 4804758, 17 pages, 2021.

[30] C. Peng, M. Luo, H. Wang, M. Khan, and D. He, "An efficient privacy-preserving aggregation scheme for multi-dimensional data in IoT," *IEEE Internet of Things Journal*, vol. 9, no. 1, pp. 589–600, 2022.

[31] X. Shi, F. Tong, W. A. Zhang, and L. Yu, "Resilient privacy-preserving distributed localization against dishonest nodes in Internet of Things," *IEEE Internet of Things Journal*, vol. 7, no. 9, pp. 9214–9223, 2020.

[32] S. Wang, L. Huang, Y. Nie et al., "Local differential private data aggregation for discrete distribution estimation," *IEEE Transactions on Parallel and Distributed Systems*, vol. 30, no. 9, pp. 2046–2059, 2019.

[33] X. Xie and Y. Chen, "Decentralized data aggregation: a new secure framework based on lightweight cryptographic algorithms," in *Proceedings of the IEEE Conference on Dependable and Secure Computing (DSC)*, Aizuwakamatsu, Japan, January 2021.

[34] X. He, H. Yang, G. Wang, and J. Yu, "Towards trusted node selection using blockchain for crowdsourced abnormal data detection," *Future Generation Computer Systems*, vol. 133, pp. 320–330, 2022.

[35] T. Feng, X. Wang, C. Liu, and J. Fang, "Secure data collaborative computing scheme based on blockchain," *Security and Communication Networks*, vol. 2021, Article ID 6630291, 9 pages, 2021.

[36] C. Zhang, L. Zhu, C. Xu, and K. Sharif, "PRVB: achieving privacy-preserving and reliable vehicular crowdsensing via blockchain oracle," *IEEE Transactions on Vehicular Technology*, vol. 70, no. 1, pp. 831–843, 2021.

[37] Q. Kong, L. Su, and M. Ma, "Achieving privacy-preserving and verifiable data sharing in vehicular fog with blockchain," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 8, pp. 4889–4898, 2021.

[38] H. Lin, S. Garg, J. Hu, G. Kaddoum, M. Peng, and M. Hossain, "A blockchain-based secure data aggregation strategy using sixth generation enabled network-in-box for industrial applications," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 10, pp. 7204–7212, 2021.

[39] Z. Guan, X. Zhou, P. Liu, L. Wu, and W. Yang, "A blockchain based dual side privacy preserving multi party computation scheme for edge enabled smart grid," *IEEE Internet of Things Journal*, vol. 9, no. 16, pp. 14287–14299, 2022.

[40] G. Wang, J. He, X. Shi, J. Pan, and S. Shen, "Analyzing and evaluating efficient privacy-preserving localization for pervasive computing," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2993–3007, 2018.

[41] X. Xu, Q. Liu, X. Zhang, J. Zhang, L. Qi, and W. Dou, "A blockchain-powered crowdsourcing method with privacy preservation in mobile environment," *IEEE Transactions on Computational Social Systems*, vol. 6, no. 6, pp. 1407–1419, 2019.

[42] X. Chen, K. Nguyen, and H. Sekiya, "An experimental study on performance of private blockchain in IoT applications," *Peer-to-Peer Networking and Applications*, vol. 14, no. 5, pp. 3075–3091, 2021.

[43] Y. Huang, Z. Cai, and A. Bourgeois, "Search locations safely and accurately: a location privacy protection algorithm with accurate service," *Journal of Network and Computer Applications*, vol. 2018, no. 103, 156 pages, 2018.

[44] X. Shi and J. Wu, "To hide private position information in localization using time difference of arrival," *IEEE Transactions on Signal Processing*, vol. 66, no. 18, pp. 4946–4956, 2018.