Hindawi
Security and Communication Networks
Volume 2021, Article ID 7354316, 12 pages
https://doi.org/10.1155/2021/7354316

Research Article

Private Data Aggregation Based on Fog-Assisted Authentication for Mobile Crowd Sensing

Ruyan Wang,[1,2,3] Shiqi Zhang,[1,2,3] Zhigang Yang,[1,2,3] Puning Zhang,[1,2,3] Dapeng Wu,[1,2,3] Yongling Lu,[4] and Alexander Fedotov[5]

[1]*School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China*
[2]*Advanced Network and Intelligent Connection Technology Key Laboratory of Chongqing Education Commission of China, Chongqing 400065, China*
[3]*Chongqing Key Laboratory of Ubiquitous Sensing and Networking, Chongqing 400065, China*
[4]*State Grid Jiangsu Electric Power Company Ltd. Research Institute, Nanjing 211103, China*
[5]*Peter the Great St. Petersburg Polytechnic University, Polytechnicheskaya, 29, St.Petersburg 195251, Russia*

Correspondence should be addressed to Shiqi Zhang; 17780734752@163.com

Received 8 May 2021; Accepted 28 August 2021; Published 22 September 2021

Academic Editor: James Ying

Copyright © 2021 Ruyan Wang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In mobile crowd sensing (MCS), the cloud as a single sensing platform undertakes a large number of communication tasks, leading to the reduction of sensing task execution efficiency and the risk of loss and leakage of users' private data. In this paper, we propose a spatial ciphertext aggregation scheme with collaborative verification of fog nodes. Firstly, the cloud and fog collaboration architecture is constructed. Fog nodes are introduced for data validation and slices transmission, reducing computing cost on the sensing platform. Secondly, a multipath transmission method of slice data is proposed, in which the user identity and data are transmitted anonymously by the secret sharing method, and the data integrity is guaranteed by hash chain authentication. Finally, a spatial data aggregation method based on privacy protection is presented. The ciphertext aggregation calculation of the sensing platform is realized through Paillier homomorphic encryption, and the problem of insufficient data coverage in the sensing region is solved by the position-based weight interpolation method. The security analysis demonstrates that the scheme can achieve the expected security goal. The simulation results show the feasibility and effectiveness of the proposed scheme.

1. Introduction

With the rapid development of mobile communication technology and the popularity of various wearable mobile devices, mobile users can collect various data anytime and anywhere. Mobile crowd sensing (MCS) is an emerging perception model. Mobile users collect sensing data for specific tasks through sensors (e.g., cameras and temperature sensors) that are embedded in the phone or wearable device. Then, the data is uploaded to sensing platforms by wireless sensing technologies (e.g., wireless networks and Bluetooth). After the task is completed, mobile users get paid from the platform [1, 2]. While receiving the sensing data, the sensing platform is responsible for evaluating and aggregating sensing data. Data aggregation often mines the raw data for more useful information. For example, the average air quality index obtained by aggregation can reflect the local air quality condition more intuitively; the average travel speed of public transportation on a road can reflect the congestion of that road. After processing the uploaded data, the platform transmits the uploaded data to the task initiator and completes the sensing task. With low deployment cost and large coverage area, MCS can be applied in areas such as traffic congestion prediction [3, 4], industrial IoT [5–7], traffic detection [8, 9], smart medical [10, 11], environmental detection [12], and social networking [13, 14].

However, MCS faces some serious problems in privacy, security, and communication in the above applications.

Firstly, the sensing data collected by MCS often involves the user's location data that contains abundant personal information. If an attacker obtains the user's geographic location from the perceived data, the user's activity range can be inferred [15, 16]. To protect sensitive information of mobile users, most studies encrypt or add noise to the sensing data, such as local differential privacy [17, 18]. However, the sensing platform cannot aggregate the encrypted data, which reduces the usability of the sensing data. Secondly, when transmitting sensing data through wireless networks, the sensing data is easily exposed to channel monitors, making it more easily attacked, stolen, and tampered with. Existing studies mostly carry out tamper-proof authentication of perceived data by generating hash abstract or hash chain [19, 20] or provide an identity authentication system [21] to prevent attackers from malicious submission of false data. However, there is still a risk that the generated hash value will be intercepted by the attacker. In addition, when the number of sensing terminals is too large, the frequent data verification by the sensing platform will bring huge communication and computing costs and reduce the efficiency of the sensing platform. Finally, mobile users are randomly distributed in various locations in the city, and the sensing data collected and uploaded are discrete. These discrete distributions of sensing data are not conducive to the overall evaluation of the sensing area, so to obtain the sample values of unknown locations, they are generally obtained by interpolation algorithms related to the location of the sensing data, but they often reveal the specific location of the mobile users and leak user privacy.

Targeting at the above problems, this paper proposes a spatial ciphertext aggregation scheme with collaborative verification of fog nodes. Inspired by the significant advantages of fog nodes [22, 23], we use fog nodes for data validation and slice transmission to alleviate the communication and computation costs of the sensing platform. Shamir secret sharing is used to transmit the sensing data and user identity information to the fog nodes in the form of slices, which ensures the integrity of the sensing data and the privacy security of the user identity and then combines the one-way hash function to complete data authentication, and finally, the sensing platform recovers the encrypted data and user identity information to complete other operations. The scheme also ensures the aggregated computation of the sensed data in encrypted form, while the prediction of the sample values of unknown locations is realized in combination with the geographic interpolation algorithm, which enables the overall data evaluation of the sensing area. The main contributions of this paper are as follows:

(1) A novel cloud and fog collaboration architecture is constructed. Fog nodes are introduced to assist the sensing platform considering its characteristics of low delay, multiple distribution, and certain computing capacity, realizing data verification and slice reception, and reducing the communication and computing costs of the sensing platform.

(2) A multipath transmission method of slice data is put forward. Sensing data and user identity information are sliced and transmitted through Shamir secret sharing. Then, a reasonable secret threshold t is set according to the number of fog nodes to realize anonymous transmission of user identity, and hash chain authentication is adopted to achieve a trade-off between privacy protection and data integrity.

(3) A spatial data aggregation method based on privacy protection is advanced. The ciphertext aggregation calculation of the sensing platform is realized through Paillier homomorphic encryption, and the problem of insufficient data coverage in the sensing region is solved by the position-based weight interpolation method.

The remainder of this paper is organized as follows. The related works are introduced in Section 2. Section 3 describes the preliminary knowledge of Paillier encryption protocol, secret sharing, and inverse distance weighted. The system model is introduced in Section 4. Then, Section 5 introduces the spatial secret aggregation scheme with collaborative verification of fog nodes. And, the security analysis and simulation results are described in detail in Section 6. Finally, Section 7 summarizes the paper.

2. Related Work

The privacy protection issues in the MCS system mainly focus on privacy task allocation, data collection, and data aggregation. Relevant researchers have published the following research results on these issues.

Based on fog-assisted computing, a Privacy-Aware Task Allocation and Data Aggregation (PTAA) scheme was proposed by using bilinear pairing and homomorphic encryption technology in literature [24]. The scheme took advantage of the fog nodes to assist the sensing platform to assign tasks and used the transport independent protocol and the secure two-party aggregation protocol to realize the privacy task assignment and data aggregation, reducing the burden of the sensing platform. Ni et al. [25] proposed a Fog-Assisted Secure Data Deduplication (Fo-SDD) scheme. By designing a BLS-oblivious pseudorandom function, it enabled fog nodes to delete deduplicated data, while protecting privacy, ensuring data confidentiality, and improving communication efficiency. The scheme also achieved anonymization of user identity during data collection by further extending Fo-SDD. Basudan et al. [26] proposed a Certificateless Aggregate Signcryption (CLASC) scheme to enhance security in data transmission of vehicular crowd sensing based on the road surface condition monitoring system with fog computing, which ensured data privacy security using lower computation cost. However, the above scheme does not consider the risk of interception of sensing data during transmission, and a malicious attacker may intercept the transmission data in the open transmission network, resulting in the loss of sensing data and affecting the sensing task to be performed.

Concerning data collection and aggregation, Chen et al. [27] put forward a data privacy protection method for untrusted servers. The collected data was divided into multiple slices based on the number of adjacent participants, and then, the data slices were forwarded to the adjacent participants. When the number of slices reached a threshold, all slice carriers sent data slices directly to the server. However, this method simply distributed the data slices randomly to the neighboring nodes. When data slices were transmitted, attackers can easily collect data slices, leading to an increased probability of data leakage. In literature [28], a privacy-preserving data aggregation scheme was designed using data slicing and blending techniques, which supports additive aggregation. Data slices were distributed to neighboring participants; thus, the participants' sensing data was hidden. Li and Cao [29] presented a new mobile sensing protocol to obtain the sum of time-series data, which uses homomorphic encryption and a novel key management scheme based on efficient HMAC to achieve additive ciphertext aggregation of sensed data. However, the protocol required additional communication to handle dynamic user access. But the above literature did not consider the case where the participants collude with the server to leak privacy. Fan et al. [30] came up with a novel privacy-aware and trustworthy sum aggregation protocol for mobile sensing, which protected the data privacy of benign users even when multiple users conspire against each other, but there was still a risk of losing the submitted data.

In other studies in the area of MCS security, Agir et al. [31] proposed a user-adaptive location privacy protection scheme, which generated multiple noises by setting a personal privacy threshold and a user-defined privacy protection level. Then, the user's privacy security was guaranteed combined with the spatial steganography unit. However, this solution was computationally expensive and lacked effective privacy level criteria. Gisdakis et al. [32] used Security Assertion Markup Language (SAML) and Transport Layer Security (TLS) protocols to establish trust between entities, and then, Private Information Retrieval (PIR) techniques were adopted to ensure privacy in communication. Based on the Merkle tree, the privacy protection mechanism in literature [33] was presented, which can authenticate participants anonymously without the trusted third party. However, the above schemes did not consider the case that malicious attackers submit false data, which may interfere with the final results.

3. Preliminaries

3.1. Paillier Encryption Protocol. The Paillier Cryptosystem is a modular, public-key encryption scheme, created by Pascal Paillier [34]. The security of this homomorphic encryption scheme is based on determining the nth-order residue class problem. In the following, we will review the specific process of the program:

3.1.1. Key Generation. To construct the key, one must choose two large primes p and q, and then, compute $n = pq$, $\lambda = \mathrm{lcm}[(p-1)(q-1)]$, where lcm (p, q) is calculated as the least common multiple of p and q. Then, select a semi-random, nonzero value $g \in Z^*_{n^2}$ such that $k = L(g^\lambda \bmod n^2)$, where $L(u) = u - 1/n$. It is said that g is semi-random since k generated by g needs to satisfy $\gcd(k, n) = 1$, and then, calculate $\mu = k^{01} \bmod n$.

The public key Pk is (n, g), and the private key Sk is (λ, μ).

3.1.2. Encryption. For the plaintext m, select the random parameter $r \in Z^*_n$. Then, the ciphertext

$$\begin{aligned} c &= E(m) \\ &= g^m \cdot r^n \bmod n^2. \end{aligned} \tag{1}$$

3.1.3. Decryption. The Paillier decryption function:

$$m = L\left(c^\lambda \bmod n^2\right) \cdot \mu \bmod n. \tag{2}$$

3.1.4. Homomorphic Properties. An encryption function with the homomorphic property is an encryption function where two plaintexts m_1 and m_2 satisfy $C(E(m_1), E(m_2)) = E(m_1 \oplus m_2)$, where C is an operation on the ciphertext domain. When $\oplus$ represents addition, the encryption is said to be additive homomorphic encryption; when $\oplus$ represents multiplication, the encryption is said to be multiplicative homomorphic encryption. Homomorphic properties of the Paillier encryption algorithm:

$$D\left(E(m_1) \cdot E(m_2) \bmod n^2\right) \equiv m_1 + m_2 \bmod n. \tag{3}$$

3.2. Shamir Secret Sharing Algorithm. The secret sharing algorithm was proposed by Shamir in 1979 based on Lagrange interpolation, which allows n participants to share a secret value s, but the secret value s can be recovered by any t participants, and less than t participants cannot get any information about s. The above t is called the threshold, and a secret sharing with n participants and a threshold of t is denoted as (t, n)-secret sharing. The formal definition of Shamir secret sharing is as follows.

3.2.1. Related Parameters. The finite domain F_q is chosen, the secret value $s \in F_q$, t is the threshold, the set of participants is $U = \{u_1, u_2, \ldots, u_n\}$, the identity of each participant is u_i, and $u_i \in F_q$ is not equal to zero.

3.2.2. Slicing and Distribution. Randomly choose a t-1 degree polynomial $f(x)$ on F_q; $f(x)$ is shown below:

$$f(x) = s + a_1 x^1 + a_2 x^2 + \cdots + a_{t-1} x^{t-1} \bmod q, \tag{4}$$

where $a_1, a_2, \ldots, a_{t-1} \in F_q$ in $f(x)$. Then, all secret slices are calculated based on participant identity:

$$y_i = f(u_i). \tag{5}$$

Finally, the computed slices are secretly distributed to the corresponding participant u_i.

3.2.3. Secret Recovery. When there are no less than t participants providing secret slices, one can use u_i and y_i to recover $f(x)$, and hence the $t-1$ degree polynomial $f(x)$ can be easily obtained by using the equation as follows::

$$f(x) = \sum_{i=1}^{t} y_i \prod_{j=1, j\neq i}^{t} \frac{x - u_j}{u_i - u_j} \bmod q. \tag{6}$$

After that, the secret value s is recovered by substituting $x = 0$ into $f(x)$.

3.3. Inverse Distance Weighted. Inverse distance weighted (IDW) is a weighted average interpolation method that can be interpolated in an exact or smooth manner. It uses the distance between the interpolation point and the sample point as the weight for the weighted average, and the closer the sample point is to the interpolation point, the greater the weight given to it. Suppose that the predicted location is (x_0, y_0), the predicted value is z, the perceived user location is (x_i, y_i), the perceived data is m_i, and the number of participating users is n. Calculate z according to the following steps:

(1) Calculate the Euclidean distance for each point:

$$(x_i - x_0)^2 + (y_i - y_0)^2 = d_i. \tag{7}$$

(2) Calculate the distance weights for each point:

$$w_i = \frac{d_i^{-1}}{\sum_{i=1}^{n} d_i^{-1}}. \tag{8}$$

(3) Calculate the value of the unknown point:

$$\begin{aligned} z &= \sum_{i=1}^{n} w_i m_i \\ &= \frac{d_1^{-1} m_1 + d_2^{-1} m_2 + \cdots + d_n^{-1} m_n}{\sum_{i=1}^{n} d_i^{-1}}. \end{aligned} \tag{9}$$

4. System Model

4.1. System Model. As shown in Figure 1, the spatial ciphertext aggregation system with collaborative verification of fog nodes consist of sensing platform, task initiator, fog nodes, mobile users, and authority center.

4.1.1. Task Initiator. Task initiators are users of the MCS services. The task initiator is responsible for issuing a specific task, and each task has the clear data type requirement. A task initiator could be an individual or organization that lacks an ability to perform a certain computing or data collection task.

4.1.2. Sensing Platform. The sensing platform could be played by an organization or a corporation that provides a platform for crowdsourcing. It accepts service requests from task initiator, deals with the requests, selects proper mobile users, and assigns relevant tasks to them.

4.1.3. Fog Nodes. The fog nodes act as a relay between the sensing platform and the mobile user, undertaking data verification and the reception and distribution of data slices.

4.1.4. Mobile Users. Referring to mobile users with sensing devices, mobile users collect data and calculate spatially relevant statistical information as required by the task. After encrypting the data, the sensing data and identity data are sliced according to the number of fog nodes deployed. Finally, the slices are sent to the fog nodes along with the authenticated hash digest value.

4.1.5. Authority Center. It is responsible for generating and distributing key materials to data requestors and MCS servers. In this system, the authority center distributes the generated public key and the parameters required for data slicing to mobile users for data encryption and slicing and distributes the private key to task initiator so that they can download the aggregated encrypted data from the sensing platform and get the specified task data.

4.2. Security Model. In the architecture of this paper, we assume that the authority center is fully trusted and that the authority center cannot be attacked by any attackers and that it manages the distribution of keys and other parameters. Task initiator, sensing platform, fog nodes, and mobile users are all honest but curious, and each part will follow the rules to perform its own task, but will also infer information about others based on the data it holds. And, external security threats come from malicious attackers; in general, attackers may listen to communication channels and intercept encrypted sensing data, spatial data, etc.

4.3. Design Objective. Based on the above security model and system architecture, we propose the following design goals:

4.3.1. Privacy. During the task execution, the specific location and sensing data of the mobile user are encrypted, and the fog nodes and sensing platform do not know the specific location and sensing data of the mobile user. In the data aggregation phase, the aggregated data is still stored in the encrypted form in the sensing platform, and only the task initiator can access it through the private key.

4.3.2. Security. The encrypted sensing data and user identity information are distributed to the fog nodes in a slicing manner so that an attacker cannot obtain the specific sensing data and user identity information even if he intercepts part

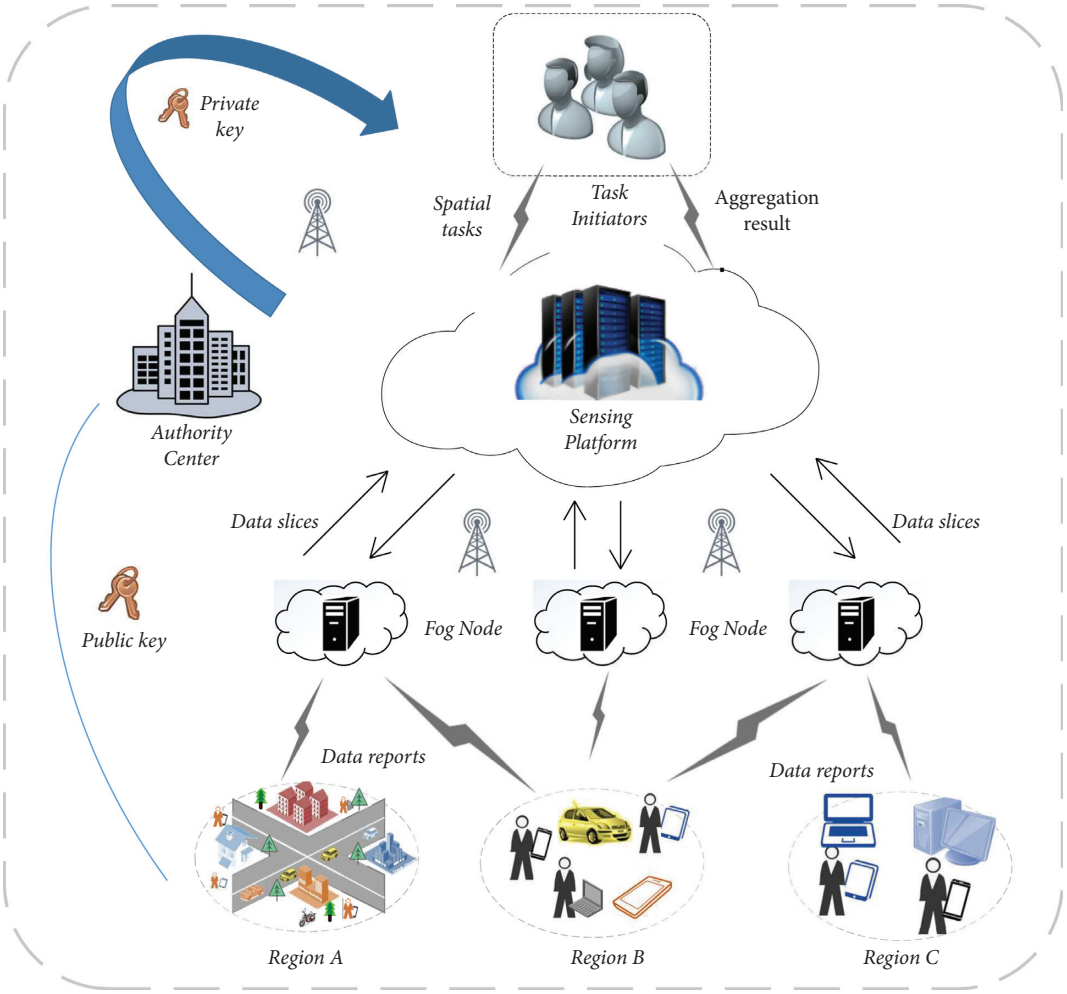

Figure 1: System architecture.

of the data slices. And, the data slices come with a hash digest, so an attacker cannot interfere with sensing data recovery by tampering with some of the data slices. For the internal perception system, the fog nodes only undertake the function of receiving and forwarding in pieces, and the user identity information can only be obtained after secret recovery by the sensing platform, which ensures the privacy and security of the user identity.

4.3.3. Efficiency. Fog nodes take on the verification of sensing data, reducing the communication and computation cost of the sensing platform.

5. Spatial Ciphertext Aggregation Scheme with Collaborative Verification of Fog Nodes

In this section, we propose a spatial secret aggregation scheme with collaborative verification of fog nodes, which consists of five phases: system initialization, mobile user data report generation, data validation and slices reception, secret recovery and data aggregation, and data decryption and result acquisition.

5.1. Overview. Task initiator initiates spatially relevant task requests to obtain overall sensing data for a region. After receiving the task request, the sensing platform assigns the task to the mobile users. Then, the authority center configures the system parameters, distributing the public key and fog nodes identity to the mobile users and the private key to the task initiator. Mobile users collect data according to the requirements of task. Because the specific locations of mobile users within the sensing area are discrete, the uploaded sensing data has limited coverage. And, mobile users need to calculate spatially relevant statistical information to get sample values of some unknown locations in combination with geographic interpolation that make the uploaded data in the area more holistic.

This paper focuses on describing the computation of sample values for unknown locations by data aggregation using homomorphic encryption and geographic interpolation. In this process, in order to hide the mobile users' location data and identity information and to protect the privacy of the sensing data, mobile users encrypt data with public keys, slice the data and identity information based on the number of fog nodes, and then use one-way hash functions to generate hash chain for data authentication.

Mobile users distribute data, identity information slices, and authentication information to the corresponding fog nodes. Afterward, the fog nodes verify its data integrity and transmit the data and identity information slices to the sensing platform after the verification is completed. The sensing platform receives the data slices and performs secret recovery to get the mobile users' encrypted sensing data and the users' original identity information. The sensing platform completes the incentive or other operations based on the identity information and then performs ciphertext data aggregation. After aggregation is completed, the task initiator downloads the aggregated data via the private key to obtain the aggregated results.

5.2. System Initialization. In our system model, consider mobile users as $P = \{p_1, p_2, p_3, \ldots, p_n\}$, mobile user location as $\{x_i, y_i\}$, sensing data as m_i, identity information as p_i, spatially relevant statistical information as D_i, unknown locations as (x_o, y_o), fog nodes as $U = \{u_1, u_2, u_3, \ldots, u_k\}$, each fog node identity as u_j, and hash function as h. At the beginning of the sensing task, the authority center randomly selects two large prime numbers p and q, calculates $n = pq$ according to the predefined calculation principle, and satisfies $\gcd[L(g^{\lambda} \bmod n^2), n] = 1$. The public key (n, g) is transmitted to the mobile users, and the secret sharing-related parameters and the fog node identity u_j are also sent to the mobile users together. Then, the authority center computes $\lambda = \mathrm{lcm}[(p-1), (q-1)]$ and $\mu = L(g^{\lambda} \bmod n^2)^{01} \bmod n$ and transfers the private key (μ, λ) to the task initiator.

5.3. Location-Aware Inverse Distance Weighted Ciphertext Aggregation Protocol. As shown in Figure 2, m_i represents the sensing data collected by mobile user p_i at its location, and d_i represents the Euclidean distance between the mobile user and the unknown location. At the beginning of the sensing task, the sensing platform broadcasts the coordinates of the unknown location and the mobile user computes the Euclidean distance d_i between itself and the unknown location. Then, the mobile user encrypts $d_i^{-1} m_i$ and d_i^{-1} to get C_{i1} and C_{i2}. The sensing platform receives encrypted data from n mobile users and uses homomorphic encryption properties to obtain sensing data aggregation results with the ciphertext form. Then, the task initiator uses the private key transmitted by AC to decrypt and finally gets the aggregated result with plaintext form $d_1^{-1} m_1 + d_2^{-1} m_2 + d_3^{-1} m_3 + \cdots + d_n^{-1} m_n$ and $d_1^{-1} + d_2^{-1} + d_3^{-1} + \cdots d_n^{-1}$. Based on the knowledge in the Preliminaries section, the sample value z for the unknown location can be calculated.

5.4. Mobile User Data Report Generation. This phase is divided into three main steps: sensing data acquisition and spatial data calculation, data encryption, and data transmission.

Step 1. Sensing data acquisition and spatial data calculation: each mobile user p_i collects sensing data m_i as required by the task and calculates spatial data based on its own location:

$$\frac{1}{(x_i - x_0)^2 + (y_i - y_0)^2} = d_i. \tag{10}$$

Due to the properties of Paillier homomorphic encryption, data transformation of d_i is required to obtain spatially relevant statistical information for encryption:

$$D_i = \left[d_i \cdot 10^k \right]. \tag{11}$$

where k varies with the sensing area range to ensure that D_i is an integer and [] is the rounding symbol.

Step 2. Data encryption: for each mobile user p_i, after sensing data collection and computing spatially relevant statistical information are performed, data encryption is performed using the received public key (n, g):

$$\begin{aligned} c_{i1} &= E(D_i m_i) \\ &= g^{D_i m_i} \cdot r^n \bmod n^2, \\ c_{i2} &= E(D_i) \\ &= g^{D_i} \cdot r^n \bmod n^2. \end{aligned} \tag{12}$$

where c_{i1} and c_{i2} denote the ciphertext information obtained by the user after encrypting $D_i m_i$ and D_i.

Step 3. Data transmission: before performing data forwarding, authority center (AC) counts the number of working fog nodes in the current sensing area, sets a maximum number of slices $M_{\max}$, and queries the historical data forwarding success rate of fog nodes in the area. After that, AC makes a trade-off between privacy of the transmitted data and efficiency of the sensing task completion. If this sensing task requires higher privacy of the transmitted data, AC selects the threshold t based on the maximum number of slices $M_{\max}$. On the contrary, if the sensing task needs to be completed efficiently and the privacy requirement of the transmitted data is lower, AC prioritizes the fog nodes with a high success rate of historical forwarded data and generates a threshold t based on the number of these fog nodes. After that, the AC sends the fog node identity, threshold t, and other data slicing related parameters to the mobile user and the sensing platform. Mobile user p_i splits two copies of data c_{i1} and c_{i2} and own identity information p_i into k slices according to the number of fog nodes, while setting a suitable threshold value t. Mobile user p_i slices data and identity information according to the fog nodes' identity $U = \{u_1, u_2, u_3, \ldots, u_k\}$ distributed by the authority center:

$$\begin{aligned} f_1^i(x) &= c_{i1} + a_1 x^1 + a_2 x^2 + \cdots + a_{t-1} x^{t-1} \bmod q, \\ f_2^i(x) &= x_{i2} + a_1 x^1 + a_2 x^2 + \cdots + a_{t-1} x^{t-1} \bmod q, \\ f_3^i(x) &= p_i + a_1 x^1 + a_2 x^2 + \cdots + a_{t-1} x^{t-1} \bmod q. \end{aligned} \tag{13}$$

The mobile user p_i gets the data and identity information slices generated by the identity identifiers of the

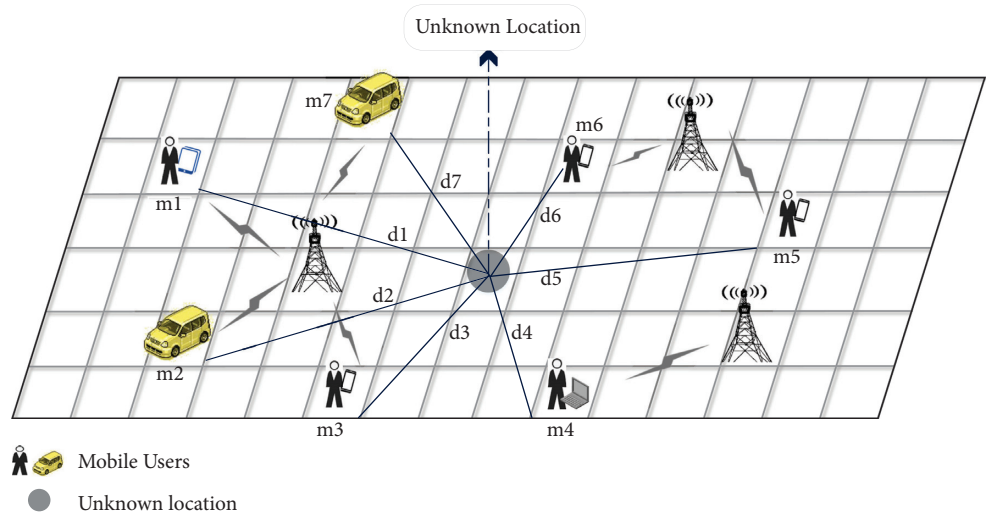

Figure 2: Location-aware inverse distance weighted ciphertext aggregation protocol.

k fog nodes, respectively. $f^i(u_j)$ denotes the slice obtained by the mobile user p_i through the fog node identity u_j, and n and k are the number of mobile users and fog nodes, respectively, and the following are the slices generated by the data c_{i1} and c_{i2} and identity information p_i of user p_i, respectively:

$$\begin{aligned} &f_1^i(u_1), f_1^i(u_2), f_1^i(u_3), \ldots, f_1^i(u_k), \\ &f_2^i(u_1), f_2^i(u_2), f_2^i(u_3), \ldots, f_2^i(u_k), \\ &f_3^i(u_1) f_3^i(u_2), f_3^i(u_3), \ldots, f_3^i(u_k). \end{aligned} \tag{14}$$

As shown in Figure 3, the mobile user p_i generates data slices, connects the data slice $f^i(u_j)$ with the hash digest value h_{j01}^i generated by the previous data slice $f^i(u_{j01})$ to generate a new hash digest value h_j^i, and points to the next data slice $f^i(u_{j+1})$ until the final generation of the end of the hash chain h_k^i.

Finally, the mobile user p_i sends the k data slices $f_1^i(u_j)$ and $f_2^i(u_j)(1 \le j \le k)$ generated from data c_{i1} and c_{i2} along with the corresponding hash digest values and k identity information slices $f_3^i(u_j)(1 \le j \le k)$ to the k corresponding fog nodes.

5.5. Data Validation and Slices' Reception. In this phase, mobile users send their encrypted data slices with authentication information and identity information slices to the fog nodes. Then, fog nodes will first verify the integrity of the encrypted data. As shown in Figure 4, after receiving the data slice $f^i(u_j)$ corresponding to mobile user p_i, fog node u_j uses the hash digest $h_{j-1}^{i'}$ sent by the previous fog node u_{j01}, connects to generate $h_j^{i'}$, and transmits it to the next fog node u_{j+1}. Finally, the last fog node u_k compares the two generated hash chain tails $h_{k1}^{i'}$ and $h_{k2}^{i'}$ with the received h_{k1}^i and h_{k2}^i, and if the results are consistent, the verification is successful. In the above process, there is a certain probability that the data slices are stolen by the attacker, and the fog nodes whose data

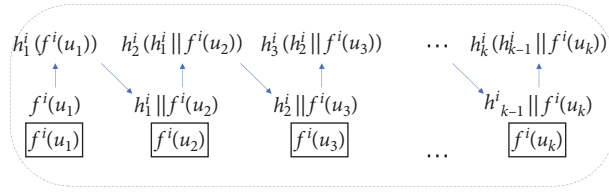

Figure 3: Hash chain.

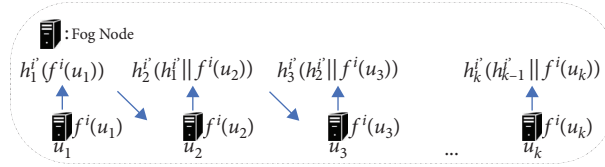

Figure 4: Fog nodes' collaborative data validation.

slices are intercepted by the attacker cannot compute the hash digest to complete collaborative authentication. At this time, if the number of remaining adjacent fog nodes are greater than t, the data slicing can still be collaboratively verified to ensure the integrity and authenticity of the transmitted data. If collaborative verification fails, fog node u_j compares the hash digest h_j received by itself with the computed h_j' to locate the location of the fog node with the wrong data slice. As for the users' identity information slicing, the fog nodes undertake the function of relaying and forwarding to ensure the anonymous transmission of users identity information. The k identity information slices $f_3^i(u_j)(1 \le j \le k)$ of user p_i are stored on the corresponding k fog nodes and transmitted to the sensing platform together after the encrypted data slices are successfully verified.

5.6. Secret Recovery and Data Aggregation. The fog nodes send the received users' identity information slices and the verified data slices to the sensing platform, which first performs secret recovery:

$$f_c^i(x) = \sum_{j=1}^{t} f_c^i(u_j) \prod_{L=1, L\neq j}^{t} \frac{x - u_L}{u_j - u_L} \bmod q \, (c = 1, 2, 3). \quad (15)$$

Substituting $x = 0$ into the above equation, we get

$$\begin{aligned} f_1^i(0) &= c_{i1} \\ &= E(D_i m_i), \\ f_2^i(0) &= c_{i2} \\ &= E(D_i), \\ f_3^i(0) &= p_i. \end{aligned} \quad (16)$$

The sensing platform recovers the encrypted data c_{i1} and c_{i2} of the user p_i and the identity information p_i. Then, the sensing platform uses the received identity information to achieve the incentive mechanism or performs other necessary system operations. Afterward, using the homomorphic encryption property of Paillier, the sensing platform starts ciphertext aggregation of the received encrypted data from all users:

$$\begin{aligned} D_1 m_1 + D_2 m_2 + \cdots + D_n m_n \leftarrow F_1 &= E(D_1 m_1) E(D_2 m_2) \cdots E(D_n m_n) \bmod n^2, \\ D_1 + D_2 + \cdots + D_n \leftarrow F_2 &= E(D_1) E(D_2) \cdots E(D_n) \bmod n^2. \end{aligned} \quad (17)$$

5.7. Data Decryption and Result Acquisition. The task initiator decrypts the aggregation result using the received private key (μ, λ) and then computes $z = Z_1/Z_2$ to obtain the sample value z of the unknown location:

$$\begin{aligned} D_1 m_1 + D_2 m_2 + \cdots + D_n m_n &= L\left(F_1^{\lambda} \bmod n^2\right) \cdot \mu \bmod n = Z_1, \\ D_1 + D_2 + \cdots + D_n &= L\left(F_2^{\lambda} \bmod n^2\right) \cdot \mu \bmod n = Z_2. \end{aligned} \quad (18)$$

6. Performance Evaluation

In this section, we first analyze how the spatial ciphertext aggregation scheme with collaborative verification of fog nodes achieves the given design goals and then experimentally demonstrate the performance of this scheme in terms of communication efficiency and computation cost.

6.1. Security Analysis

6.1.1. Data Privacy and Security. In the data collection phase, the mobile user encrypts the sensing data and spatial data using the public key sent by the authority center, and the encrypted data is transmitted to the fog nodes in the form of data slices. Data verification phase, fog nodes, or other malicious attackers who intercept the data are unable to infer the plaintext message m_i from the ciphertext C_i. In the data aggregation phase, the data slices received by the sensing platform are recovered in ciphertext, and the sensing platform performs data aggregation on the received ciphertext data. After data aggregation, the aggregated results are still stored in the sensing platform in ciphertext, which only the task initiator can get by decrypting with private key. And, the sensing platform cannot get the plaintext data in the aggregation process. In general, only the task initiator can get the final result in plaintext during the above process, while the fog nodes or the sensing platform can only process the ciphertext. The security of Paillier homomorphic encryption technology ensures that the sensing data can withstand internal and external privacy threats of the MCS system.

6.1.2. Data Integrity and Identity Privacy Security. For mobile users, the identity information and encrypted sensing data are divided into k slices based on the number of fog nodes. Each slice is generated based on the corresponding fog node identity, and a suitable recovery threshold t is set. When the data slice is sent to the corresponding fog node, the mobile user generates the corresponding hash chain according to the method in Section 5 and sends it to the corresponding fog node together with the data slices. Therefore, even if a malicious attacker intercepts a part of the data slices, according to the secret sharing feature in Section 3, as long as the number of remaining slices is greater than t, the sensing platform is still able to recover the encrypted data. Although some malicious attackers intercept the data slices and re-send forged messages pretending to be legitimate participants, all fog nodes will collaboratively authenticate based on the received hash chain, which guarantees the accuracy of the data source. The users' identity information are also stored in the form of slices on the fog nodes, and a single fog node cannot know the real identity of the user, less than t fog nodes also cannot collude to launch the real identity of the user, and only the sensing platform can recover to get the users' identity, to achieve the user identity anonymous transmission. After the sensing platform recovers the identity information, it completes the incentive or other system operations according to the user's identity. In this scheme, Shamir secret sharing guarantees the anonymous transmission of user identity, and combining with hash chain message authentication guarantees the integrity of data.

6.2. Experiment. We performed the simulation in Python 3.8, and the scenarios and related configuration parameters involved are as follows.

In the simulations, we consider a scenario in which the task initiator requests the overall air index in a region. We set the number of mobile users to 10 ~ 100 with a growth step of 10 and the number of tasks participated by each mobile user to 10 ~ 50 with a growth step of 10. Mobile user p_i randomly generates sensing data distributed in [100, 1000], and the coordinates of the location of each mobile user are set to $[(x_i, y_i) \| 0 \le x_i \le 100, 0 \le y_i \le 100]$. The number of fog nodes is set to 10 ~ 100, and the growth step is 30. For Paillier homomorphic encryption, we set the number of key bits to 32 ~ 256 bits to meet the security requirements of different data lengths, respectively, but it will bring some computation cost accordingly. All system simulations are simulated on a PC (CPU: Core i5-9400F @ 2.90 GHz and RAM: 8 GB).

The performance metrics include the computation cost of data encryption, data slicing, data recovery and aggregation, and data decryption. Then, we evaluate the impact of the number of mobile users, the number of fog nodes, the secret threshold t, the number of tasks per user, and the key length on the above parts.

6.2.1. Costs of Data Encryption. The computation cost per mobile user in the encryption phase as the number of tasks grows is given in Figure 5 to demonstrate the efficiency of data submission by mobile users. Since mobile users are located in a lightweight computing scenario, the key length of 32 ~ 256 bits can fully fulfill the data encryption requirements in this scenario, and this scheme can fulfill the privacy protection requirements of mobile users with a small increase in computing cost.

To simulate the encryption environment with different data lengths, we also give the computation cost with different key lengths. From the figure, we can see that the computation cost increases as the number of tasks per mobile user grows, which is because mobile users cannot process multiple tasks in parallel, and when the number of tasks is too large, mobile users consume a lot of computation time. At the same time, with the same number of tasks, the encryption cost varies greatly with different key lengths, so it is necessary to choose the appropriate number of key bits according to different encryption environments to fulfill the security requirements in different scenarios.

6.2.2. Cost of Validation and Aggregation. The computation cost of the fog nodes and the sensing platform is demonstrated in Figure 6. From the figure, it can be seen that the fog nodes undertake part of the computation tasks of the sensing platform and reduce the computation cost of the sensing platform, which is consistent with the design goal of this scheme.

In Figure 6, the fog nodes take on the task of data verification, and since each fog node receives data slices generated by each mobile user based on the identity of that fog node, the number of slices processed by each fog node increases as the number of mobile users grows, and

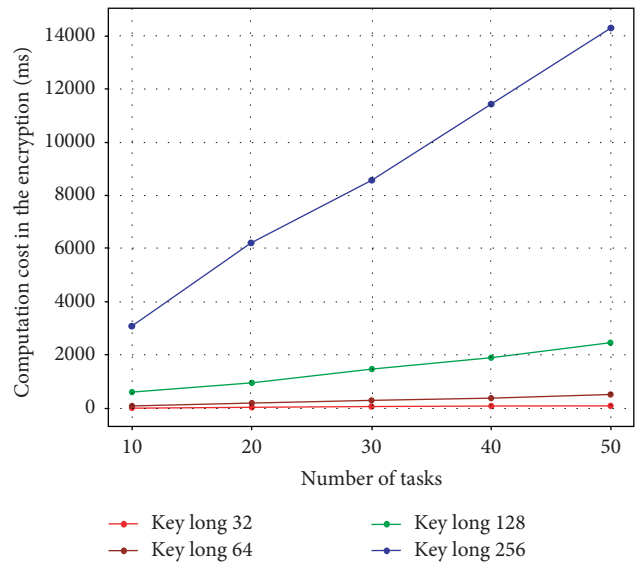

Figure 5: The computation cost in encryption.

the computation cost increases. And, the growth of the number of fog nodes will lead to a longer hash chain, increasing the time for collaborative verification. But the corresponding secret sharing threshold can also be increased, which can improve the security of sensing data transmission. We assume that the data is divided into n slices and the threshold is $t \le n$, which means that the attacker can recover the sensing data by stealing t data slices, and if n is increased and t is increased accordingly, the data slices that the attacker needs to steal will increase accordingly, and the difficulty of stealing will also increase, reducing the risk of sensing data being stolen. Since the sensing platform takes on the task of data slicing recovery and ciphertext aggregation, the computation cost will be higher than fog nodes that only perform authentication. While increasing the secret recovery threshold t affects the data recovery time, the number of mobile users affects the ciphertext aggregation time, and from the four subplots in Figure 6, we can find that the computation cost of the sensing platform increases with the number of mobile users and the threshold.

6.2.3. Data Accuracy. Since this paper combines homomorphic encryption with IDW, the additive homomorphic property is used to compute the sample value of the unknown location. The inverse of the distance between each mobile user and the unknown location is rounded, which leads to a difference between the calculated results and those calculated using IDW. This is the main reason for the error. So, we use the relative error to express the difference between the sample values of unknown locations obtained using this scheme and the real sample values of unknown locations. The relative error can well reflect the degree of data reliability, where Z_t denotes the sample value of the unknown location obtained after the tth encryption and aggregation using this scheme, while $Z_{t'}$ denotes the sample value of the unknown location obtained by the tth direct aggregation

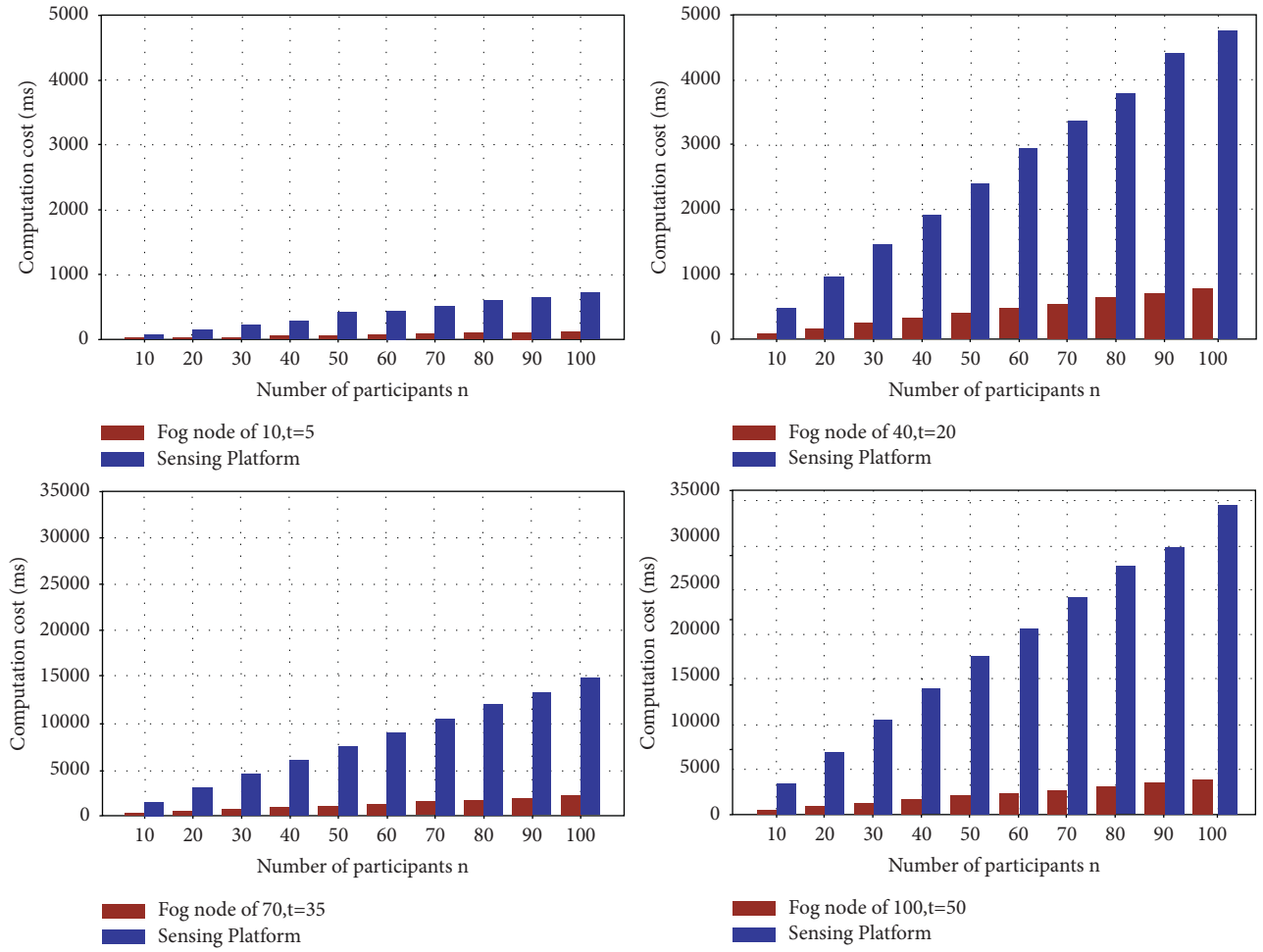

Figure 6: The computation cost in validation and aggregation.

without encryption, δ denotes the relative error, Δ denotes the accuracy, and the scheme will be run 1000 times to get the average relative error. The error in this scheme comes from the data error caused by rounding the data due to encryption when the mobile user calculates the spatially relevant statistical information D_i related to its location:

$$\delta = \frac{1}{n}\sum_{t=1}^{n}\left|\frac{Z_t - Z_t'}{Z_t'}\right|, \qquad \Delta = (10\delta) \times 100\%. \tag{19}$$

We represent in Figure 7 the accuracy of the data obtained when different numbers of mobile users are involved in the task. The figure shows that the results obtained using our scheme are in general agreement with the real values and that our scheme is able to trade-off privacy security in data transmission and encrypted data aggregation with a fairly small loss of accuracy.

6.2.4. Cost of Data Decryption. Figure 8 shows the computation cost of the task initiator to obtain the sensed data. Since the task initiator decrypts the data directly at the sensing platform using the private key, the key length is the main factor affecting the decryption time.

Overall, the computation cost paid by mobile users and task initiators in this scheme is much lower than that of fog nodes and sensing platform, and mobile users only need to pay a small computation cost to fulfill their own requirements for privacy protection. Therefore, this scheme can fulfill the requirements of mobile users and task requestors with limited computation power and achieve lightweight task participation.

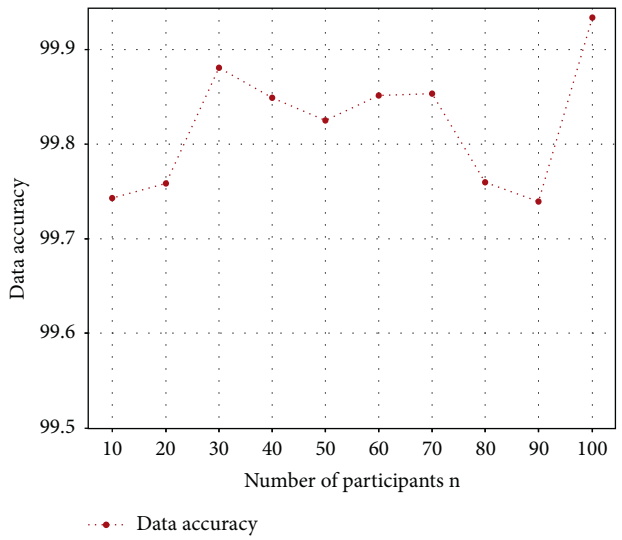

Figure 7: Data accuracy.

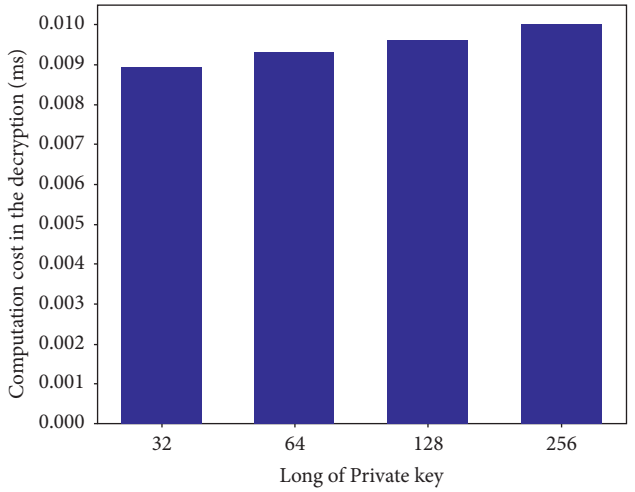

Figure 8: The computation cost in decryption.

7. Conclusion

In this paper, we propose a spatial ciphertext aggregation scheme with collaborative verification of fog nodes. Firstly, a cloud and fog collaboration architecture is constructed, where fog nodes are introduced to undertake the functions of data verification and slice reception, which reduces the computational cost of the sensing platform. Secondly, a multipath transmission method of slice data is advanced to realize the anonymous transmission of user identities. Then, combined with hash chain authentication, the integrity and authenticity of the sensing data are ensured. Finally, a privacy-protected spatial data aggregation method is presented. The interpolation method is adopted to predict the sample values of unknown locations in the sensing area, and the Paillier homomorphic encryption is used to ensure the privacy of the perceived data in this process. Security analysis and simulation results show that the solution can protect user privacy and security and reduce the computational cost of the sensing platform.

Data Availability

The data used to support the findings of the study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by National Natural Science Foundation of China (61901071, 61871062, 61771082, and U20A20157), General Project of Natural Science Foundation of Chongqing (cstc2019jcyj-msxmX0303), Science and Natural Science Foundation of Chongqing, China (cstc2020jcyj-zdxmX0024), University Innovation Research Group of Chongqing (CXQT20017), and Program for Innovation Team Building at Institutions of Higher Education in Chongqing (CXTDX201601020).

References

[1] W. Feng, Z. Yan, H. Zhang, K. Zeng, Y. Xiao, and Y. T. Hou, "A survey on security, privacy, and trust in mobile crowdsourcing," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2971–2992, 2017.

[2] J. Xiong, X. Chen, Q. Yang, L. Chen, and Z. Yao, "A task-oriented user selection incentive mechanism in edge-aided mobile crowdsensing," *IEEE Transactions on Network Science and Engineering*, vol. 7, 2019.

[3] L. Xiao, T. Chen, C. Xie, H. Dai, and H. V. Poor, "Mobile crowdsensing games in vehicular networks," *IEEE Transactions on Vehicular Technology*, vol. 67, no. 2, pp. 1535–1545, 2017.

[4] J. Xiong, R. Bi, M. Zhao, J. Guo, and Q. Yang, "Edge-assisted privacy-preserving raw data sharing framework for connected autonomous vehicles," *IEEE Wireless Communications*, vol. 27, no. 3, pp. 24–30, 2020.

[5] J. Xiong, R. Ma, L. Chen et al., "A personalized privacy protection framework for mobile crowdsensing in iiot," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 6, pp. 4231–4241, 2019.

[6] J. Xiong, M. Zhao, M. Z. A. Bhuiyan, L. Chen, and Y. Tian, "An ai-enabled three-party game framework for guaranteed data privacy in mobile edge crowdsensing of iot," *IEEE Transactions on Industrial Informatics*, vol. 17, no. 2, pp. 922–933, 2019.

[7] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5g wireless communications: a deep learning approach," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 227–236, 2018.

[8] A. Thiagarajan, L. Ravindranath, K. LaCurts 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*, pp. 85–98, Berkeley CA, USA, November 2009.

[9] Z. Zhang, P. Zhang, D. Liu, and S. Sun, "Srsm-based adaptive relay selection for d2d communications," *IEEE Internet of Things Journal*, vol. 5, no. 4, pp. 2323–2332, 2017.

[10] C. De Capua, A. Meduri, and R. Morello, "A smart ecg measurement system based on web-service-oriented architecture for telemedicine applications," *IEEE Transactions on*

Instrumentation and Measurement, vol. 59, no. 10, pp. 2530–2538, 2010.

[11] R. Morello, C. De Capua, and A. Meduri, "A wireless measurement system for estimation of human exposure to vibration during the use of handheld percussion machines," *IEEE Transactions on Instrumentation and Measurement*, vol. 59, no. 10, pp. 2513–2521, 2010.

[12] M. A. Alsheikh, Y. Jiao, D. Niyato, P. Wang, D. Leong, and Z. Han, "The accuracy-privacy trade-off of mobile crowdsensing," *IEEE Communications Magazine*, vol. 55, no. 6, pp. 132–139, 2017.

[13] H. Amintoosi and S. S. Kanhere, "A reputation framework for social participatory sensing systems," *Mobile Networks and Applications*, vol. 19, no. 1, pp. 88–100, 2014.

[14] C. Luo, S. Guo, S. Guo, L. T. Yang, G. Min, and X. Xie, "Green communication in energy renewable wireless mesh networks: routing, rate control, and power allocation," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 12, pp. 3211–3220, 2014.

[15] G. Wang, B. Wang, T. Wang, A. Nika, H. Zheng, and B. Y. Zhao, "Defending against sybil devices in crowdsourced mapping services," in *Proceedings Of the 14th Annual International Conference On Mobile Systems, Applications, and Services*, pp. 179–191, Singapore, June 2016.

[16] Z. Yang, R. Wang, D. Wu, and D. Luo, "Utm: a trajectory privacy evaluating model for online health monitoring," *Digital Communications and Networks*, vol. 7, 2020.

[17] H. Sun, B. Dong, H. Wang, T. Yu, and Z. Qin, "Truth inference on sparse crowdsourcing data with local differential privacy," in *Proceedings of the 2018 IEEE International Conference on Big Data (Big Data)*, pp. 488–497, IEEE, Seattle, WA, USA, December 2018.

[18] N. Wang, X. Xiao, Y. Yang et al., "Collecting and analyzing multidimensional data with local differential privacy," in *Proceedings of the 2019 IEEE 35th International Conference on Data Engineering (ICDE)*, pp. 638–649, IEEE, Macao, China, April 2019.

[19] B. Zhao, S. Tang, X. Liu, X. Zhang, and W.-N. Chen, "Ironm: privacy-preserving reliability estimation of heterogeneous data for mobile crowdsensing," *IEEE Internet of Things Journal*, vol. 7, no. 6, pp. 5159–5170, 2020.

[20] Y. Tian, Z. Wang, J. Xiong, and J. Ma, "A blockchain-based secure key management scheme with trustworthiness in dwsns," *IEEE Transactions on Industrial Informatics*, vol. 16, no. 9, pp. 6193–6202, 2020.

[21] D. Wu, Z. Yang, B. Yang, R. Wang, and P. Zhang, "From centralized management to edge collaboration: a privacy-preserving task assignment framework for mobile crowd sensing," *IEEE Internet of Things Journal*, vol. 8, 2020.

[22] D. Wu, R. Bao, Z. Li, H. Wang, H. Zhang, and R. Wang, "Edge-cloud collaboration enabled video service enhancement: a hybrid human-artificial intelligence scheme," *IEEE Transactions on Multimedia*, vol. 23, 2021.

[23] D. Wu, X. Han, Z. Yang, and R. Wang, "Exploiting transfer learning for emotion recognition under cloud-edge-client collaborations," *IEEE Journal on Selected Areas in Communications*, vol. 39, 2020.

[24] H. Wu, L. Wang, and G. Xue, "Privacy-aware task allocation and data aggregation in fog-assisted spatial crowdsourcing," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 1, pp. 589–602, 2019.

[25] J. Ni, K. Zhang, Y. Yu, X. Lin, and X. S. Shen, "Providing task allocation and secure deduplication for mobile crowdsensing via fog computing," *IEEE Transactions on Dependable and Secure Computing*, vol. 17, no. 3, pp. 581–594, 2018.

[26] S. Basudan, X. Lin, and K. Sankaranarayanan, "A privacy-preserving vehicular crowdsensing-based road surface condition monitoring system using fog computing," *IEEE Internet of Things Journal*, vol. 4, no. 3, pp. 772–782, 2017.

[27] J. Chen, H. Ma, and D. Zhao, "Private data aggregation with integrity assurance and fault tolerance for mobile crowdsensing," *Wireless Networks*, vol. 23, no. 1, pp. 131–144, 2017.

[28] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, "Prisense: privacy-preserving data aggregation in people-centric urban sensing systems," in *Proceedings of the 2010 Proceedings IEEE INFOCOM*, pp. 1–9, IEEE, San Diego, CA, USA, March 2010.

[29] Q. Li and G. Cao, "Efficient and privacy-preserving data aggregation in mobile sensing," in *Proceedings of the 2012 20th IEEE International Conference On Network Protocols (ICNP)*, pp. 1–10, IEEE, Austin, TX, USA, November 2012.

[30] J. Fan, Q. Li, and G. Cao, "Privacy-aware and trustworthy data aggregation in mobile sensing," in *Proceedings of the 2015 IEEE Conference on Communications and Network Security (CNS)*, pp. 31–39, IEEE, Florence, Italy, September 2015.

[31] B. Agir, T. G. Papaioannou, R. Narendula, K. Aberer, and J.-P. Hubaux, "User-side adaptive protection of location privacy in participatory sensing," *GeoInformatica*, vol. 18, no. 1, pp. 165–191, 2014.

[32] S. Gisdakis, T. Giannetsos, and P. Papadimitratos, "Security, privacy, and incentive provision for mobile crowd sensing systems," *IEEE Internet of Things Journal*, vol. 3, no. 5, pp. 839–853, 2016.

[33] Q. Li and G. Cao, "Providing efficient privacy-aware incentives for mobile sensing," in *Proceedings of the 2014 IEEE 34th International Conference On Distributed Computing Systems*, pp. 208–217, IEEE, Madrid, Spain, July 2014.

[34] M. O'Keeffe, "The paillier cryptosystem," *Mathematics Department*, vol. 18, pp. 1–16, 2008.