Hindawi
Security and Communication Networks
Volume 2021, Article ID 7776193, 10 pages
https://doi.org/10.1155/2021/7776193

Research Article

Edge Computing Assisted an Efficient Privacy Protection Layered Data Aggregation Scheme for IIoT

Rong Ma, Tao Feng, and Junli Fang

School of Computer and Communication, Lanzhou University of Technology, Lanzhou 730050, China

Correspondence should be addressed to Tao Feng; fengt@lut.cn

Received 7 May 2021; Revised 3 August 2021; Accepted 20 August 2021; Published 29 September 2021

Academic Editor: Qing Yang

The emergence of edge computing has improved the real time and efficiency of the Industrial Internet of Things. In order to achieve safe and efficient data collection and application in the Industrial Internet of Things, a lot of computing and bandwidth resources are usually sacrificed. From the perspective of low computing and communication overhead, this paper proposes an efficient privacy protection layered data aggregation scheme for edge computing assisted IIoT by combining the Chinese Remainder Theorem (CRT), improved Paillier homomorphic algorithm, and hash chain technology (edge computing assisted an efficient privacy protection layered data aggregation scheme for IIoT, EE-PPDA). In EE-PPDA, first, a layered aggregation architecture based on edge computing is designed. Edge nodes and cloud are responsible for local aggregation and global aggregation, respectively, which effectively reduces the amount of data transmission. At the same time, EE-PPDA achieves data confidentiality through improved Paillier encryption, ensuring that neither attackers nor semitrusted nodes (e.g., edge nodes and clouds) can know the private data of a single device, and it can resist by simply using hash chains to resist tampering and pollution attacks ensure data integrity. Second, according to the CRT, the cloud can obtain the fine-grained aggregation results of subregions from the global aggregation results, thereby providing fine-grained data services. In addition, the EE-PPDA scheme also supports fault tolerance. Even if some IIoT devices or communication links fail, the cloud can still decrypt incomplete aggregated ciphertexts and obtain the expected aggregation results. Finally, the performance evaluation shows that the proposed EE-PPDA scheme has less calculation and communication costs.

1. Introduction

With the increasing popularity of IoT in the industrial field, IIoT, as an important application of the Internet of Things in the industry, has received more and more attention from researchers. IIoT is dedicated to interconnecting things in industrial scenarios, such as machines, sensors, and actuators [1], as well as sampling, processing, and applying real-time data in industrial environments, which promotes the conversion of traditional industries to smart industries. Since devices and sensors are usually resource-constrained, the traditional IIoT architecture integrates cloud computing models, sending all data collected by local devices to the cloud for processing and storage to reduce the computing and storage costs of local devices [2]. However, with the rapid deployment of IIoT devices, more and more data are frequently sent to remote clouds, which not only causes huge communication costs but also brings huge processing and storage pressure to the cloud. Therefore, it is not practical to rely solely on the cloud computing model for delay-sensitive IIoT applications. In this case, the edge computing model is introduced as a supplement to cloud computing [3] to achieve efficient local data processing in IIoT; that is, user terminals can migrate their computing and storage tasks to the local edge of the network edge node [4], thereby reducing the processing pressure on the cloud, realizing low-latency data processing, and significantly reducing communication overhead.

In IIoT, large amounts of perception data collected by industrial equipment and regularly transmitted to the cloud usually contain sensitive information [5, 6]. Therefore, in recent years, reducing the amount of transmitted data and

protecting the privacy and security of the data have attracted a lot of attention. Data aggregation is seen as an effective method to reduce communication overhead and protect data privacy. For example, edge nodes can perform aggregation operations on the received data and then deliver a single aggregation result to the cloud, thereby significantly reducing the amount of data transmission, and the data privacy of a single device is leaked [7]. Although data aggregation can achieve a great performance improvement, the aggregation operation is usually performed by an untrusted third party, so privacy and security (confidentiality and integrity) are still threatened. For example, curious entities (such as edge nodes and clouds) can observe private content in received data packets.

In order to provide fine-grained data services on the cloud while protecting data privacy, confidentiality, and integrity, this paper proposes an efficient privacy protection layered data aggregation scheme for edge computing assisted IIoT. The main contributions are summarized in the following points:

(1) The first major contribution is the design of a layered aggregation architecture based on edge computing, which enables data aggregation to be implemented on the local edge nodes and the cloud separately, which significantly reduces the amount of data transferred from the edge nodes to the cloud.

(2) The second contribution is that edge nodes use a simple hash chain mechanism to resist tampering and pollution attacks, while also preventing the leakage of individual device privacy information at semitrusted nodes and resisting eavesdropping attacks on all communication links in the IIoT.

(3) The third contribution is that the cloud can recover the aggregate results of all subregions and the entire region from a single global aggregated ciphertext to support fine-grained data services. At the same time, when the IIoT device or transmission channel fails, the cloud can still decrypt the aggregated ciphertext smoothly; that is, the proposed scheme supports fault tolerance.

This remainder of the article is organized as follows: Section 2 covers the work of the edge computing and data aggregation scheme for IIoT. The system model and adversary model of the proposed privacy protection data aggregation scheme are described in Section 3. In Section 4, we describe the efficient privacy protection layered data aggregation scheme. Section 5 analyzes the proposed program in terms of safety and performance, respectively. Section 6 summarizes the full text.

2. Related Work

Recently, many methods to protect cloud/edge system data security have been proposed, such as certificateless signature [8] and blockchain [9]. There are also many schemes that use homomorphic encryption to achieve secure data aggregation [10]. For example, Lu et al. [11] designed an efficient and privacy-protected aggregation scheme in the smart grid. The scheme uses a super-increasing sequence to integrate multidimensional data into a one-dimensional form and then uses the Paillier algorithm to aggregate the encrypted data. This reduction significantly improves communication efficiency and better meets the real-time requirements of communication. Chen et al. [12] introduced a novel multifunctional data aggregation scheme that allows the gateway to perform multifunctional aggregation, and the control center can calculate various statistical information (variance, one-way analysis of variance, etc.) in a privacy-protected manner and be flexible and provide diversified services locally. At the same time, by increasing the acceptable noise to resist the differential attack [13], Li et al. [14] constructed an effective privacy protection demand response scheme. By combining homomorphic encryption and key update technology, the solution can provide privacy protection, confidentiality, and key update functions. In addition, Li et al. [15] proposed a privacy protection dual-function aggregation scheme based on lattice encryption technology. The data control center in the smart grid can calculate the mean and variance of all users' power consumption and protect user privacy to prevent eavesdropping. Wang et al. [16] designed an anonymous aggregation scheme for edge-assisted cloud computing systems. This scheme reduces bandwidth consumption by using intermediate fog nodes to perform homomorphic aggregation and protects identity privacy through anonymity mechanisms. However, the above solutions can only achieve privacy protection against external attackers and cannot prevent privacy leakage caused by internal threats. For example, a semitrusted or compromised cloud control center can obtain individual device data.

In order to overcome the above shortcomings, in literature [17], the authors designed a privacy-protected data aggregation scheme based on untrusted aggregators, which enables each user to encrypt data with different keys to prevent the aggregator from infringing on data privacy. In addition, the scheme also uses differential privacy technology to resist differential attacks. Ni et al. [18] proposed a security-enhanced data aggregation scheme based on Paillier encryption, in which a trapdoor hash function is used to implement data authentication to protect the confidentiality and integrity of data and prevent malicious aggregation. In addition, Chen et al. [19] designed a fault-tolerant data aggregation scheme using homomorphic Paillier encryption. This solution can protect personal user data from attacks from gateways, control centers, and powerful attackers that can destroy the control center, while supporting fault tolerance. Kamil et al. [20] designed a privacy aggregation scheme suitable for smart grids based on the elliptic curve encryption algorithm, which can not only safely resist internal attacks but also solve a series of security challenges. Zhang et al. [21] proposed a novel space-time aggregation scheme, in which the time dimension aggregation is performed on the user side, and the gateway is responsible for the spatial aggregation of the entire community. This scheme realizes privacy protection by resisting internal and external collusion attacks. However, the above solutions can only

provide a global aggregation result for the control center and cannot meet the more fine-grained requirements of the cloud. For example, the cloud needs to know the aggregation results of multiple specific subregions.

In order to solve the above problems, Lu et al. [22] proposed a novel privacy protection subset aggregation scheme to meet the needs of the control center to obtain more fine-grained aggregation results. This scheme divides the entire user residence into two subsets according to the set threshold and then obtains the total energy consumption and the number of users in each subset by using the composite order group. At the same time, the data privacy of individual users is protected at the curious gateway and control center. Lu et al. extended the work in [22] to support data integrity authentication and proposed a subset aggregation scheme based on data integrity [23]. This scheme is based on a novel hash chain construction mechanism to complete the verification of the integrity of the aggregated data. Literature [24] proposed a privacy-protected multi-subset data aggregation scheme, which can protect the privacy of users while calculating the number of users and summarizing the total power consumption of each subset. However, this scheme lacks a verification mechanism to ensure the integrity of the received data and does not support fault tolerance. In addition, Knirsch et al. [25] proposed a fault-tolerant and efficient scheme to aggregate data on different groups. The solution is based on CRT, Shamir's secret sharing, and Paillier algorithm to formulate a novel aggregation protocol to support efficient and fault-tolerant group aggregation with privacy protection, as well as the dynamic joining and leaving of households. However, this solution is not fault-tolerant. When any smart meter fails, it will not be able to recover the global aggregation result. At the same time, both literatures [25] lack a data integrity authentication scheme.

The above schemes can all produce certain privacy protection data aggregation effects, but there still remain the following unresolved problems: (1) Data aggregation operations are usually performed by untrusted third parties, so there are privacy and security risks. While resisting external attackers, we also need to guard against internal attackers. (2) The cloud can recover the aggregate results of all subregions and the entire region from a single global aggregated ciphertext to support fine-grained data services. (3) When the IIoT device or transmission channel fails, the cloud can still decrypt the aggregated ciphertext smoothly; that is, the proposed scheme supports fault tolerance.

3. Problem Description

3.1. System Model. In the IIoT network scenario based on edge computing, a layered data aggregation system model is constructed. The model includes three layers, sensing layer, edge layer, and cloud layer, and mainly includes five entities: IIoT device, edge node (EN), Industrial Cloud (IC), Trusted Management Authority (TMA), and user. The detailed relationship between these entities is shown in Figure 1.

In the sensing layer, IIoT devices are divided into multiple subareas based on geographical distribution. Each IIoT device has sensing, processing, and communication functions and is regarded as a data source. Their main responsibility is to collect sensing data in designated areas in real time and periodically forward their encrypted data to the industrial cloud through edge nodes. The purpose is to monitor specific areas and protect the privacy of sensitive data at the same time.

In edge layer, each subarea is managed by an adjacent edge node, and the edge node is an intermediate device between the IIoT device and the cloud. The edge node is mainly responsible for two tasks. The first task is data authentication: when the edge node receives sensing data from the IIoT device, in order to ensure the authenticity and integrity of the data, the edge node will perform authentication operations on the received data. If the received data has not been tampered with or is not contaminated data injected by an active attacker, the edge node will accept the data; otherwise, it will be deleted. The second task is data aggregation: after the edge node authenticates all the received data, it aggregates all the valid encrypted data into a number and generates a local report to send to the industrial cloud, which greatly reduces the amount of communication between edge nodes and the cloud while reducing the processing burden on the cloud.

The cloud layer contains an IC as the data management center of the system. IC is responsible for collecting data of all IIoT devices forwarded through edge nodes and performing global aggregation operations on the received local aggregated data to track aggregate statistics at any time. At the same time, IC can provide fine-grained services, that is, provide users with statistical information of designated subregions or global regions when they receive their requests.

As regards the user, for legitimate users, if they need to know the statistics of a specific subregion or global region, they can send a request to the cloud. Subsequently, according to the requirements in the user's request, the cloud returns the corresponding statistical information to the user.

Regarding TMA, in EE-PPDA, it is assumed that a fully trusted TMA only participates in the system initialization phase, and its responsibility is to initialize system parameters and keys and publish public parameters and key distribution to IIoT devices, edge nodes, and ICs.

3.2. Adversary Model. This article is mainly concerned with the security, integrity, and privacy protection in the process of data generation and transmission. Assuming that the industrial cloud and edge nodes in the network model are both "honest and curious" entities, this means that they honestly implement security protocols but at the same time remain curious about the device's sensing data.

In our adversary model, we consider a strong attacker A whose goal is to perceive as much of the user's personal privacy data as possible. "Strong" means that attacker A not only can listen to all the communication data in our system model but also can initiate the following attacks:

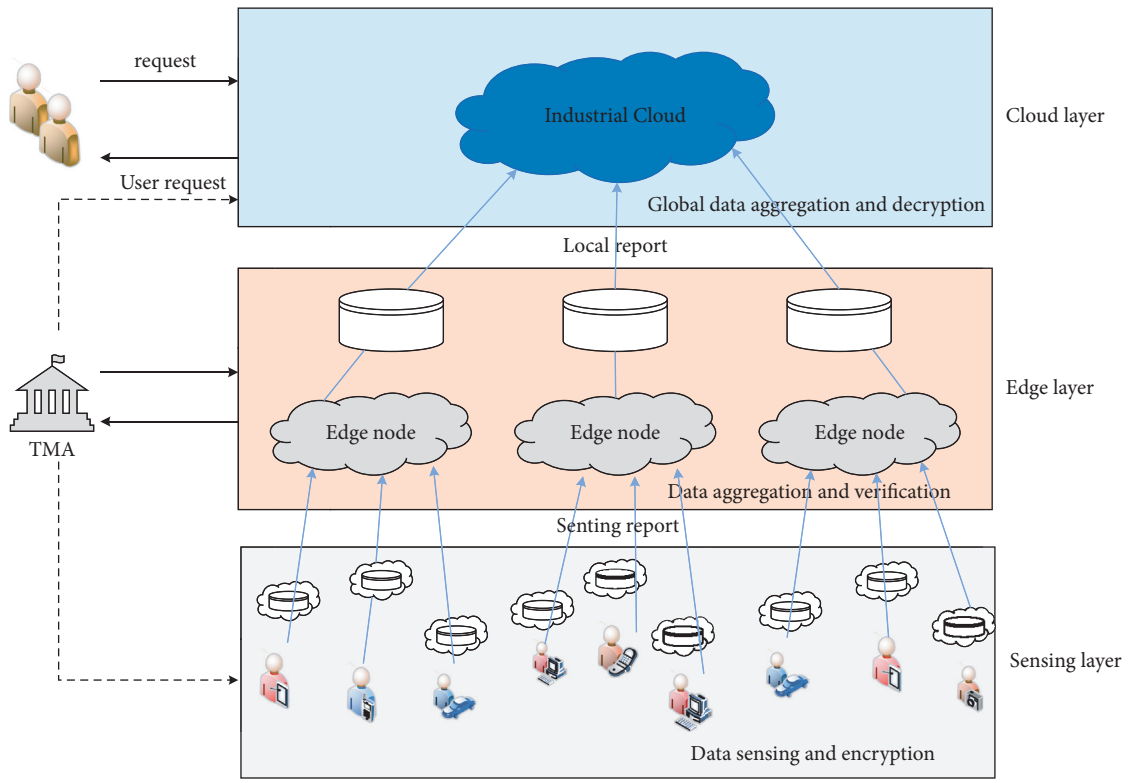

Figure 1: System model.

A may tamper with the transmitted data for malicious purposes or directly inject contaminated data. Therefore, the intermediate edge node should have the ability to detect and delete erroneous data locally.

A can eavesdrop on all communication channels to steal the transmitted perception data, which will lead to the leakage of private information.

In addition, a practical application scenario is also considered; that is, there is an IIoT device or a communication channel failure, which may cause the cloud to fail to decrypt the received aggregated ciphertext.

4. Efficient Privacy Protection Layered Data Aggregation Scheme

This section proposes an efficient privacy protection layered data aggregation scheme for IIoT. This scheme integrates the concept of layered aggregation, improved Paillier encryption, the Chinese remainder theorem, and hash chain technology to achieve efficient and fine-grained aggregation statistics decryption without exposing personal privacy and low-cost integrity authentication. The scheme mainly includes four parts: system initialization, data collection and encryption, local data processing, and global data aggregation and decryption. The details are as follows.

4.1. System Initialization. First, set two security parameters (μ, l) in the IIoT system, and then TMA randomly selects two large prime numbers Q_1 and Q_2; $|Q_1| = |Q_2| = \mu$. At the same time, calculate the public and private keys of homomorphic Paillier encryption ($N = Q_1Q_2$, $g = 1 + N$), and define a function as $L(x) = x - 1/N$. Assuming that there are k subregions in the sensing layer and n sensing devices in each subregion, TMA selects k relatively prime positive integers $p_1, p_2, \ldots, p_k, |p_i| = l$ to calculate coefficient a_i of each subregion. The process is as follows:

$$\begin{cases} P = \prod_{i=1}^{k} p_i, \\ P_i = \dfrac{P}{p_i}, \\ T_i \equiv P_i^{-1} \bmod p_i, \\ a_i = T_i \cdot P_i. \end{cases} \tag{1}$$

Subsequently, TMA uses a pseudorandom number generator to generate kn uncorrelated random numbers $\{s_{11}, \ldots s_{1n}, \ldots, s_{k1}, \ldots, s_{kn}\}$, which are assigned to corresponding sensing devices as private keys. At the same time, the private key s_0 of the industrial cloud (IC) is calculated according to the following equation and sent to the IC:

$$s_0 + \sum_{i=1}^{k} \sum_{j=1}^{n} s_{ij} \equiv 0 \bmod \lambda. \tag{2}$$

In addition, generate a set of pseudorandom numbers $\{I_{11}, \ldots, I_{1n}, \ldots, I_{k1}, \ldots, I_{kn}\}$ to construct a set of hash

chain heads $\{H_{11,0}, \ldots, H_{1n,0}, \ldots, H_{k1,0}, \ldots, H_{kn,0}\}$, and each hash head is attached with a TMA signature σ. Then it is sent to the corresponding IIoT sensing devices and edge nodes. In addition, TMA selects a cyclic group G and two secure encryption hash functions: h: $\{0,1\}^* \longrightarrow Z_N^*$ and H: $\{0,1\}^* \longrightarrow G$. Finally, TMA chooses a random number k_i as the shared key between the edge node edge_i and IC and publishes the system public parameters $\{G, L(x), N, a_i, p_i: i = 1, 2, \ldots k, h, H\}$.

4.2. Data Collection and Encryption.

(1) Collection of industrial data: Each IIoT sensing device continuously collects real-time sensing data and periodically sends the collected data to the IC through the edge node. Suppose that there are k subregions A_i in the sensing layer, satisfying the condition $A_i \cap A_j = \varnothing, \quad i = 1, 2, \ldots k, i \neq j$. Each subarea A_i is governed by an adjacent edge node edge_i and contains n IIoT sensing devices $IID_{ij}, \quad j = 1, 2, \ldots n$. At the same time, we assume that the reporting period of the IIoT sensing device is$\Gamma = \{t_1, t_1, \ldots t_{\text{MAX}}\}$, and the raw perception data collected by IID_{ij} at time $t_\tau \in \Gamma$ is denoted as $m_{ij,\tau} \in Z_N$.

(2) Sensing data encryption: Because the data collected by each IIoT sensing device always contains sensitive and private information, and there are active attackers and eavesdroppers in the communication channel between the sensing device and edge nodes, in order to prevent the privacy data of individual sensing devices from being contaminated or eavesdropped by attackers, each sensing device IID_{ij} needs to perform the following encryption operations to obtain its ciphertext $c_{ij,\tau}$ before forwarding its data $m_{ij,\tau}$ to the upper edge node:

$$h(t_\tau)^{S_{ij}} \in Z_N^*,$$
$$m'_{ij,\tau} = m_{ij,\tau} \cdot a_i,$$
$$c_{ij,\tau} = g^{mij,\tau'} \cdot r^N \bmod N^2 = (1+N)^{mij,\tau'} \cdot h(t_\tau)^{s_{ij} \cdot N} \bmod N^2 \longrightarrow (1+N)^m = 1 + \sum_{i=1}^{m} \binom{m}{i} N^i = (1 + N \cdot mij,\tau') \cdot h(t_\tau)^{s_{ij} \cdot N} \bmod N^2. \tag{3}$$

In addition, in order to provide evidence of the integrity of the received data at the edge node to ensure that the data has not been tampered with or contaminated by an attacker, a hash chain with one-way characteristics is used to calculate the current hash chain value $H_{ij,\tau}$ of the ciphertext $c_{ij,\tau}$:

$$H_{ij,\tau} = H(c_{ij,\tau}) \oplus H_{ij,\tau-1}. \tag{4}$$

Finally, the encrypted sensing report $(c_{ij,\tau}, H_{ij,\tau})$ is sent to the upper edge node edge_i, waiting for further aggregation processing.

4.3. Local Data Processing.

(1) When the edge node edge_i receives the encrypted sensing report $(c_{ij,\tau}, H_{ij,\tau})$ sent by all the sensing devices in the subarea under its jurisdiction in the time slot t_τ, it first passes the hash chain value $H_{ij,\tau}$ in the inspection report. The correctness of hash chain value verifies the integrity of all received data in turn. The specific process is as follows: edge_i calculates the hash chain value $H'_{ij,\tau} = H(c_{ij,\tau}) \oplus H_{ij,\tau-1}$ for verification based on the ciphertext $c_{ij,\tau}$ and checks whether the equation $H'_{ij,\tau} = H_{ij,\tau}$ holds. If it is true, the verification is passed, and edge_i receives $c_{ij,\tau}$ and stores $H_{ij,\tau}$ for the next integrity verification.

(2) When all verified ciphertexts $c_{ij,\tau}, \quad j = 1, 2, \ldots n$, are obtained, edge_i uses the additive homomorphism of Paillier encrypted ciphertexts to aggregate all ciphertexts without decryption. Get the aggregation result $C_{i,\tau}$ of subregion A_i under jurisdiction:

$$C_{i,\tau} = \prod_{j=1}^{n} c_{ij,\tau} = \left(1 + N \cdot \sum_{j=1}^{n} m'_{ij,\tau}\right) \cdot h(t_\tau)^{N \cdot \sum_{j=1}^{n} s_{ij}} \bmod N^2. \tag{5}$$

(3) In order to ensure the integrity of the aggregated ciphertext $C_{i,\tau}$ of the subarea, edge_i calculates the verification code $H_{i,\tau} = H(C_{i,\tau} \| k_i)$ through the shared secret key k_i with the IC and provides verification evidence for the IC. Finally, edge_i sends its local report $(C_{i,\tau}, H_{i,\tau})$ to the IC.

4.4. Global Data Aggregation and Decryption.

(1) After the cloud center receives the local reports $(C_{i,\tau}, H_{i,\tau}), 1 \le i \le k$, of k edge nodes, it first verifies the integrity of the aggregated ciphertext $C_{i,\tau}$ of all subregions in turn. The specific process is as follows: IC based on the previous one Hash chain value $H_{i,\tau-1}$ calculates $H'_{i,\tau} = H(C_{i,\tau}) \oplus H_{i,\tau-1}$ to verify whether the equation $H'_{i,\tau} = H_{i,\tau}$ is correct. If the equation is

correct, the verification is passed and the IC accepts $C_{i,\tau}$.

(2) In order to simplify the key management of the IC while enhancing the privacy protection of the individual perception device data, the system only allocates a unique key s_0 to the IC, so that the IC cannot directly decrypt the aggregated ciphertext of each subarea. In order to restore the aggregated statistical values of the desired subregion, IC must first aggregate all subregions aggregated ciphertext through the following calculation to obtain a global aggregation result C_τ:

$$C_\tau = \prod_{i=1}^{k} c_{i,\tau} = \left(1 + N \cdot \sum_{i=1}^{k}\sum_{j=1}^{n} m'_{ij,\tau}\right) \cdot h(t_\tau)^{N\cdot\left(\sum_{i=1}^{k}\sum_{j=1}^{n} s_{ij}\right)} \bmod N^2. \tag{6}$$

Next, IC can decrypt and obtain the statistical value of each subarea and the global statistical value (e.g., the sum and the average value) by performing the following steps.

Step 1: IC uses its key s_0 to eliminate the term containing $h(t_\tau)$ in the expression of C_τ and obtain value B after simplification:

$$\begin{aligned} B &= C_\tau \cdot h(t_\tau)^{s_0}, \\ &= (1+N)^{\sum_{i=1}^{k}\sum_{j=1}^{n} m'_{ij,\tau}} \cdot h(t_\tau)^{N\cdot\left(\sum_{i=1}^{k}\sum_{j=1}^{n} s_{ij} + s_0\right)} \bmod N^2 \\ &\qquad s_0 + \sum_{i=1}^{k}\sum_{j=1}^{n} s_{ij} \equiv 0 \bmod \lambda \Rightarrow s_0 + \sum_{i=1}^{k}\sum_{j=1}^{n} s_{ij} = \phi\lambda, \quad \text{where} \phi \varepsilon Z_N^* \\ &\qquad\qquad \longrightarrow \\ &= (1+N)^{\sum_{i=1}^{k}\sum_{j=1}^{n} m'_{ij,\tau}} \cdot h(t_\tau)^{N\cdot\phi\lambda} \bmod N^2 \\ &= (1+N)^{\sum_{i=1}^{k}\sum_{j=1}^{n} m'_{ij,\tau}} \bmod N^2 \\ &= \left(1 + N \cdot \sum_{i=1}^{k}\sum_{j=1}^{n} m'_{ij,\tau}\right) \bmod N^2 \end{aligned} \tag{7}$$

Step 2: According to value B, IC can decrypt to obtain a pseudoglobal aggregate value W:

$$\begin{aligned} W &= \frac{(A-1)}{N \bmod N^2}, \\ &= \sum_{i=1}^{k}\sum_{j=1}^{n} m'_{ij,\tau} m_{ij,\tau} \\ &= \sum_{i=1}^{k} a_i \sum_{j=1}^{n} \end{aligned} \tag{8}$$

Step 3: In order to obtain the total aggregation result of the global area, IC first needs to calculate the aggregation statistics of each subarea. Based on the known system parameters p_i, $i = 1, 2, \ldots, k$, IC can obtain the statistics and $D_{i,\tau}$ of each subarea through the Chinese remainder theorem:

$$\begin{aligned} D'_\tau &= W \bmod P, \\ D_{i,\tau} &= \sum_{j=1}^{n} m_{ij,\tau} = D'_\tau \bmod p_i. \end{aligned} \tag{9}$$

At the same time, the corresponding mean value $E_{i,\tau}$ of each subregion can also be obtained:

$$E_{i,\tau} = \frac{D_{i,\tau}}{n}. \tag{10}$$

Finally, the global statistics sum D_τ and the corresponding mean value E_τ of k subregions can be obtained:

$$\begin{aligned} D_\tau &= \sum_{i=1}^{k} D_{i,\tau}, \\ E_\tau &= \frac{D_\tau}{kn}. \end{aligned} \tag{11}$$

4.5. Fault Tolerance. Consider a practical scenario. Some devices in a subarea fail at a certain point in time, and the edge node cannot receive its report, causing the edge node and the cloud to receive incomplete aggregation results. Since the cloud only has one key s_0, obtaining incomplete aggregated ciphertext will cause the above-mentioned decryption process to fail to be successfully performed, and the cloud will not be able to correctly decrypt the aggregated ciphertext.

Since each edge node holds n hash chains, these hash chains are used to verify the sensing reports of n different devices at different points in time, so edge nodes can find damage by inspecting unverified hash chain devices. Let $A_i' \subset A_i$ denote the collection of faulty equipment, and let C_τ' denote the incomplete aggregation result received by edge_i at time t_τ. In order to obtain information $h'(t_\tau)$ related to the devices in the fault set A_t', edge_i sends a loss report (A_t', t_τ) to the TMA. Since the TMA manages the keys of all devices, the report is received (A_t', t_τ), and TMA can use the private key of the device involved in A_t' to calculate $h'(t_\tau)$:

$$h'(t_\tau) = h(t_\tau)^{\sum_{IID_{ij} \in A_i'} s_{ij}}. \tag{12}$$

The missing information is returned $h'(t_\tau)$ to edge_i. After receiving $h'(t_\tau)$, edge_i combines it with C_τ' to obtain the decryptable ciphertext C_τ through the following calculation:

$$C_\tau = C_\tau' \cdot h'(t_\tau) = \left(1 + N \cdot \sum_{IID_{ij} \in A_i / A_i'} m_{ij,\tau}'\right) \bmod N^2. \tag{13}$$

Then, according to equations (8)–(12), the cloud can still decrypt the incomplete aggregate ciphertext and obtain the expected aggregate statistical value.

5. Security and Performance Evaluation

5.1. Security Analysis. According to the attacker model defined in the problem description, this section will evaluate the privacy, confidentiality, and integrity of the device-sensing data.

5.1.1. Confidentiality and Privacy. For confidentiality, the ciphertext form of the sensing data $m_{ij,\tau}$ of each device IID_{ij} is $c_{ij,\tau} = (1 + N \cdot m_{ij,\tau}') \cdot h(t_\tau)^{s_{ij} \cdot N} \bmod N^2$. If $h(t_\tau)^{s_{ij}}$ is regarded as a random number, the converted ciphertext form $c_{ij,\tau}$ can regard $c_{ij,\tau} = (1 + N \cdot m_{ij,\tau}') \cdot r_{ij}^N \bmod N^2$ as the encryption result of the Paillier algorithm. Similarly, the aggregation result of subarea A_i and global area A is also a valid Paillier encryption result. Since the Paillier encryption algorithm is semantically safe against selective plaintext attacks [26], EE-PPDA can resist eavesdropping attacks and ensure the confidentiality of the original sensing data and aggregated results. At the same time, except that the authorized IC can successfully decrypt the aggregation results of each subarea and the entire area, other unauthorized entities (such as edge nodes) cannot obtain the plaintext of the aggregation results.

For privacy, neither semitrusted aggregators (edge nodes and cloud) nor eavesdroppers can obtain the perception data of a single device. When a semitrusted edge node receives all perception reports from its subarea, it will not be possible for the edge node to recover any perception data of any IIoT device because it cannot obtain the decryption private key. After all the ciphertexts are aggregated, because the aggregated result is semantically secure, the edge node still cannot infer any real information from the encrypted aggregated result. For a semitrusted IC, although it can use its private key s_0 to decrypt and read the aggregated plaintext of each subarea, it cannot observe the sensing data of a single device from the aggregated plaintext. In addition, based on the above confidentiality analysis, even if an eavesdropper can obtain the ciphertext transmitted on all communication links, it still cannot infer the original sensing data of a single IIoT device. Summarizing the above analysis results, it can be concluded that the proposed EE-PPDA scheme protects the privacy of the original data of a single IIoT device.

5.1.2. Integrity. In the transmission link between the IIoT device and the edge node, an attacker may tamper with the transmitted data or directly inject polluted data. In order to ensure the validity of the data received in the edge node, the hash chain technology is used on the edge node to achieve integrity authentication. At each transmission time point, the sensing report $(c_{ij,\tau}, H_{ij,\tau})$ of each IIoT device contains a new hash chain value, which can be calculated, where it is the previous hash chain value. Based on the one-way characteristic of the hash chain, it is not feasible for an attacker to obtain from it, so it is difficult for an attacker to launch a successful tampering attack. When the edge node receives the sensing report, if it is verified in the previous time period, it can effectively verify the integrity of the data through calculation. If it is equal, it means that it has not been tampered with or is not the tainted data injected during the communication. Therefore, EE-PPDA can effectively protect data integrity to resist malicious attacks by attackers.

5.2. Performance Evaluation. This section will evaluate the proposed EE-PPDA scheme from two aspects: the computing overhead of IIoT devices, edge nodes, and ICs and the amount of data transmission. IT is compared with three other schemes: the SEDA scheme proposed in [18], the LPDA-EC scheme in [27], and the APPA scheme in [28]. These three schemes all use the standard Paillier algorithm, and the ciphertext form is $c = g^m \cdot r^N \bmod N^2$. The simulation experiment runs on a computer configured with Intel Core i5-8250U@1.60 GHZ CPU, 8 G RAM.

5.2.1. Computational Overhead. Let the symbols C_E, C_M, C_H, C_{XOR}, C_e, C_P, and C_m denote an exponential operation on $Z_{N^2}^*$, a $Z_{N^2}^*$ multiplication operation, a hash operation, an XOR operation, and an exponential operation on the cyclic group G bilinear pairing and multiplication on G, respectively. As compared with the time-consuming bilinear pairing C_P operation, the calculation time of C_M, C_H, and

Table 1: Computational complexity comparison.

	EE-PPDA	SEDA [18]	LPDA-EC [27]	APPA [28]
IIoT device	$C_E + C_M + C_H + C_{\text{XOR}}$	$2C_E + 3C_e$	$2C_E + 3C_m$	$2C_E + C_M + C_H$
EN	$(n-1)C_M + nC_{\text{XOR}} + (n+1)C_H$	$(n-1)C_M + (n+1)C_m$	$(n-1)C_M + 4C_e$	$(n+2)(C_M + C_H)$
IC	$C_H + C_E$	$2C_P + 2C_E + (n+2)C_e$	$2C_P + 2C_E$	$C_E + C_M + C_H$

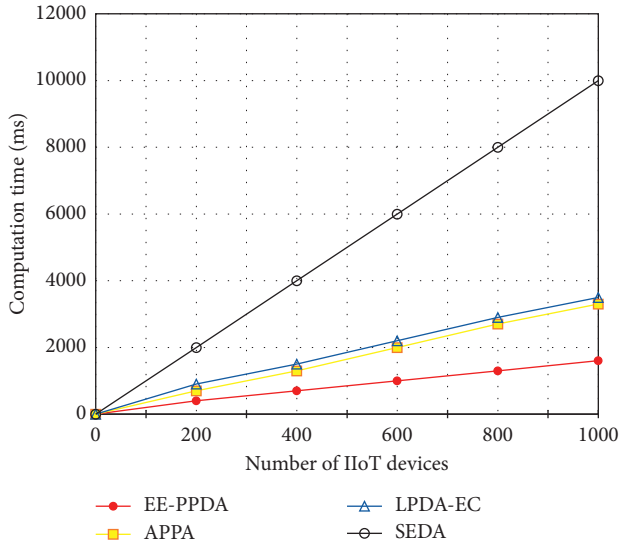

Figure 2: Computational cost comparison.

C_{XOR} and the operation time after decryption are negligible, so the computational overhead caused by these operations can be ignored. Based on the MIRACL and PBC libraries, an experiment was carried out to estimate the time cost of each operation, in which parameters μ and G were set to 512 *bits* and 160 *bits*, respectively. From the perspective of computational complexity, bilinear pairing operations have the highest computational complexity among these operations, followed by exponentiation and multiplication. Our experimental results also confirm this conclusion. The final experimental results show that the calculation time of C_P is close to 18.0 *ms*, C_e and C_E are about 1.70 *ms* and 1.60 *ms*, respectively, and smallest C_m is close to 0.07 *ms*.

In Table 1, the computational overheads of the four schemes at IIoT devices, edge nodes, and IC are listed in detail. In EE-PPDA, the calculation required for an IIoT device IID_{ij} to generate a perception report $(c_{ij,\tau}, H_{ij,\tau})$ is $C_E + C_M + C_H + C_{XOR}$, and C_E occupies the largest computational cost. Therefore, compared to the amount of calculation required by the other three schemes, $2C_E + 3C_e$, $2C_E + 3C_m$, and $2C_E + C_M + C_H$, EE-PPDA reduces the computational overhead by nearly half on the device side.

At edge nodes, if low-calculation operations (such as authentication of a single ciphertext) are ignored, the EE-PPDA, SEDA, and LDPA-EC schemes only need to perform $(n-1)C_M$ operations with a small amount of calculation. It can aggregate n ciphertexts, and the APPA scheme requires $(n+1)$ times. Due to the low time-consuming operation of C_M, it can be said that the computational costs of these four schemes at edge nodes are almost the same. At the IC, the EE-PPDA scheme only needs $C_H + C_E$ operations to verify the received reports and decrypt the aggregation results, which is slightly less than the $C_E + C_M + C_H$ operations required in the APPA scheme. However, the SEDA and LDPA-EC schemes require $2C_P + 2C_E + (n+2)C_e$ and $2C_P + 2C_E$ operations, respectively, both of which include time-consuming C_P operations. As we all know, the computational cost of C_P is significantly higher than operating C_E. Therefore, the EE-PPDA scheme greatly reduces the computational cost of the IC. Combining the above analysis results, it can be concluded that the proposed EE-PPDA scheme achieves lightweight security and privacy protection.

In order to compare the calculation cost more intuitively, the execution time of the above mechanism is calculated, and the curve of the total calculation time as the number of IIoT devices increases is depicted in Figure 2. Obviously, compared with the other three schemes, the proposed EE-PPDA scheme significantly reduces the calculation time. Especially when more IIoT devices are added, more calculations will be saved by the EE-PPDA scheme.

5.2.2. Data Transfer Volume. In the EE-PPDA scheme, data transmission includes two parts: device-to-edge communication (device-to-EN) and edge-to-IC (EN-to-IC) communication. In the device-to-EN phase, the IIoT device sends its sensing report $(c_{ij,\tau}, H_{ij,\tau})$ to the upper edge node edge_i, and the size of the report is $S_{ij} = 2048 + 160\text{bits}$. Therefore, the total amount of data transmission during device-to-edge communication is $S_{DF} = n \cdot S_{ij}$ bits. Next, in

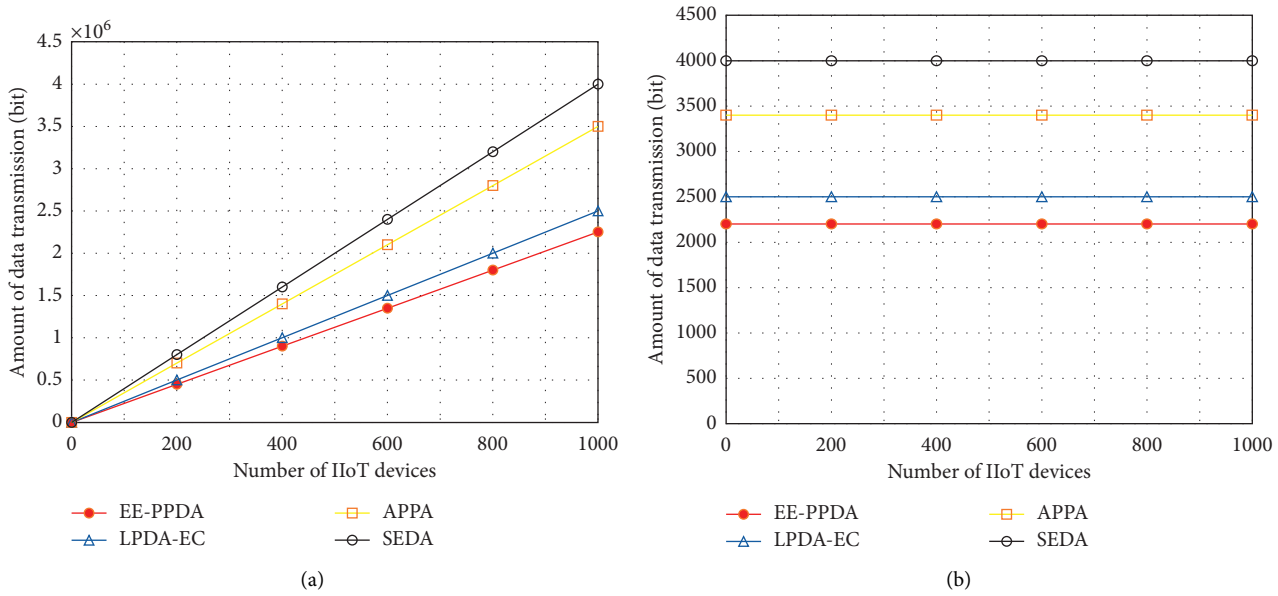

Figure 3: (a) Device-to-EN data transfer volume comparison. (b) EN-to-IC data transfer volume comparison.

the local data processing stage, since each edge node aggregates n ciphertexts into one and generates an aggregate report $(C_{i,\tau}, H_{i,\tau})$ and sends it to the IC, the amount of data transmission from edge node to IC is significantly reduced. Specifically, the amount of data transfer in the EN-to-IC phase is reduced from $(2048 + 160) \cdot n$bits to $S_{FC} = 2048 + 160$bits. Figure 3(a) shows the comparison results of the data transmission volume of the four schemes in the device-to-edge phase. It is obvious that the proposed EE-PPDA scheme achieves the slowest growth rate, and among the four schemes keep the data transfer volume to a minimum. This shows that the EE-PPDA scheme effectively reduces the amount of data communication in the device-to-edge process. From Figure 3(b), it can be found that the increase in the number of IIoT devices will not lead to an increase in the data transmission volume in the EN-to-IC phase, which is attributed to the aggregation operation of the edge nodes. At the same time, the EE-PPDA scheme still achieves the least amount of data transfer among the four schemes in the EN-to-IC phase. Combining Figures 3(a) and 3(b), it can be seen that EE-PPDA can significantly reduce communication overhead and bandwidth consumption.

From the above security and performance analysis results, it can be seen that the proposed EE-PPDA scheme is an efficient and secure data aggregation scheme. These security and performance advantages are very suitable for actual IIoT scenarios.

6. Conclusions

This paper proposes a hierarchical data aggregation scheme with efficient privacy protection in edge computing assisted IIoT, referred to as EE-PPDA. By adopting an improved homomorphic Paillier algorithm and a simple hash chain mechanism, EE-PPDA can provide effective protection for data privacy, confidentiality, and integrity at the same time. In particular, the data privacy of a single device is also protected in semitrusted edge nodes and the cloud. At the same time, the CRT-based hierarchical aggregation design enables the cloud to provide fine-grained data services by obtaining aggregation results in smaller subregions. Finally, the experimental results further prove the advantages of the scheme in terms of calculation and communication costs. In future work, consider integrating data space-time compression, network resource optimization theory, and machine learning methods into the solution in this paper to build a more efficient and smarter data aggregation solution. At the same time, the hierarchical aggregation scheme proposed in this paper provides a fault-tolerant mechanism for data loss to ensure the normal operation of the system. However, data loss will affect the final data analysis results. How to reconstruct the lost data can be considered as a future research direction.

Data Availability

The experimental data used to support the results of this study can be obtained from the corresponding authors upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grants nos. 62162039 and 61762060) and Foundation for the Key Research and Development Program of Gansu Province, China (Grant no.20YF3GA016).

References

[1] P. Borovska and M. Gugutkov, "The intersection of IoT ecosystem security and blockchain technology in the context of industry 4.0," *THERMOPHYSICAL BASIS OF ENERGY TECHNOLOGIES (TBET 2020)*, pp. 10–14, 2021.

[2] 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.

[3] K. Sha, T. A. Yang, W. Wei, and S. Davari, "A survey of edge computing-based designs for IoT security," *Digital Communications and Networks*, vol. 6, no. 2, pp. 195–202, 2020.

[4] G. Alandjani, "Leveraging vulnerabilities in sensor based IOT edge computing networks[J]," *International Journal of Future Generation Communication and Networking*, vol. 14, no. 1, pp. 11–20, 2021.

[5] B. Zhao, X. Liu, W.-N. Chen, W. Liang, X. Zhang, and R. H. Deng, "PRICE: privacy and reliability-aware real-time incentive system for crowdsensing," *IEEE Internet of Things Journal*, no. 99, p. 1, 2021.

[6] 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.

[7] 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, 2021.

[8] Y. Zhang, R. Deng, D. Zheng, J. Li, P. Wu, and J. Cao, "Efficient and robust certificateless signature for data crowdsensing in cloud-assisted industrial IoT," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 9, pp. 5099–5108, 2019.

[9] K. P. Yu, L. Tan, M. Aloqaily, H. Yang, and Y. Jararweh, "Blockchain-enhanced data sharing with traceable and direct revocation in IIoT," *IEEE Transactions on Industrial Informatics*, vol. 17, p. 11, 2021.

[10] 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, 2020.

[11] R. Lu, X. Liang, X. Lin, and X. Shen, "EPPA: an efficient and privacy-preserving aggregation scheme for secure smart grid communications," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1621–1631, 2012.

[12] L. Chen, R. Lu, Z. Cao, K. AlHarbi, and X. Lin, "MuDA: multifunctional data aggregation in privacy-preserving smart grid communications," *Peer-to-Peer Networking and Applications*, vol. 8, no. 5, pp. 777–792, 2015.

[13] B. Yang, X. Cao, X. Li, Q. Zhang, and L. Qian, "Mobile-edge-computing-based hierarchical machine learning tasks distribution for IIoT," *IEEE Internet of Things Journal*, vol. 7, no. 3, pp. 2169–2180, 2019.

[14] H. Li, X. Lin, H. Yang, X. Liang, R. Lu, and X. Shen, "EPPDR: an efficient privacy-preserving demand response scheme with adaptive key evolution in smart grid," *IEEE Transactions on Parallel and Distributed Systems*, vol. 25, no. 8, pp. 2053–2064, 2013.

[15] C. Li, R. Lu, H. Li, L. Chen, and J. Chen, "PDA: a privacy-preserving dual-functional aggregation scheme for smart grid communications," *Security and Communication Networks*, vol. 8, no. 15, pp. 2494–2506, 2015.

[16] H. Wang, Z. Wang, and J. Domingo-Ferrer, "Anonymous and secure aggregation scheme in fog-based public cloud computing," *Future Generation Computer Systems*, vol. 78, pp. 712–719, 2018.

[17] E. Shi, H. T. H. Chan, E. Rieffel, R. Chow, and D. Song, "Privacy-preserving aggregation of time-series data," *Network and Distributed System Security Symposium (NDSS)*, vol. 2, pp. 1–17, 2011.

[18] J. Ni, K. Alharbi, X. Lin, and X. Shen, "Security-enhanced data aggregation against malicious gateways in smart grid," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM)*, pp. 1–6, San Diego, CA, USA, December 2015.

[19] L. Chen, R. Lu, and Z. Cao, "PDAFT: a privacy-preserving data aggregation scheme with fault tolerance for smart grid communications," *Peer-to-Peer Networking and Applications*, vol. 8, no. 6, pp. 1122–1132, 2015.

[20] I. A. Kamil, S. O. Sunday, and O. Ogundoyin, "A privacy-aware data aggregation scheme for smart grid based on elliptic curve cryptography with provable security against internal attacks," *International Journal of Information Security and Privacy*, vol. 13, no. 4, pp. 109–138, 2019.

[21] L. Zhang, J. Zhang, and Y. H. Hu, "A privacy-preserving distributed smart metering temporal and spatial aggregation scheme," *IEEE Access*, vol. 7, pp. 28372–28382, 2019.

[22] R. Lu, K. Alharbi, X. Lin, and C. Huang, "A novel privacy-preserving set aggregation scheme for smart grid communications," in *Proceedings of the IEEE global communications conference (GLOBECOM)*, pp. 1–6, San Diego, CA, USA, December 2015.

[23] M. Tahir, A. Khan, A. Hameed, M. Alam, M. K. Khan, and F. Jabeen, "Towards a set aggregation-based data integrity scheme for smart grids," *Annals of Telecommunications*, vol. 72, no. 9-10, pp. 551–561, 2017.

[24] S. Li, K. Xue, Q. Yang, and P. Hong, "PPMA: privacy-preserving multisubset data aggregation in smart grid," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 2, pp. 462–471, 2017.

[25] F. Knirsch, D. Engel, and Z. Erkin, "A fault-tolerant and efficient scheme for data aggregation over groups in the smart grid," in *Proceedings of the IEEE Workshop on Information Forensics and Security (WIFS)*, pp. 1–6, 2017.

[26] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," *International Conference on the Theory and Applications of Cryptographic Techniques*, pp. 223–238, 1999.

[27] J. Zhang, Y. Zhao, J. Wu, and B. Chen, "LPDA-EC: a lightweight privacy-preserving data aggregation scheme for edge computing," in *Proceedings of the IEEE International Conference on Mobile Ad Hoc and Sensor Systems (MASS)*, pp. 98–106, Chengdu, China, October 2018.

[28] Z. Guan, Y. Zhang, L. Wu et al., "APPA: an anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT," *Journal of Network and Computer Applications*, vol. 125, pp. 82–92, 2019.