A Novel Privacy-Preserving Selective Data Aggregation with Revocation for Fog-Assisted IoT

Jianhong Zhang[1(✉)], Luo Ran[1], Dequan Xu[3], Jing Wang[2], Pei Liu[2], and Changgen Peng[3]

[1] School of Electrical and Computer Engineering, North China University of Technology, Beijing 100144, China
[2] Finance and Tax Innovation Department of JD Group, Beijing 100176, China
[3] Guizhou Provincial Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, Guizhou, China

Abstract. Internet-of-Things (IoT) can provide more convenient and intelligent services for our daily life by IoT devices collecting data. Fog computing enables ubiquitous perception, seamless connectivity, and real-time processing for B5G cellular IoT applications by making use of the advantages which fog nodes are deployed on the edge of the network, closer to data sources. Collecting sensor data by combining fog computing and the Internet of Things can enhance the security and efficiency of the B5G network in a low-cost way, which is very important for building a stable B5G network. Most existing data aggregation systems cannot support the aggregation of specific data types, which means that existing data aggregation systems have limitations in real-world applications. To solve these problems, in this work, we propose a novel privacy-preserving Selective Data Aggregation scheme with revocation for the fog-assisted IoT to address selective aggregation of privacy-preserving data and revocation of the application *App* by using homomorphic encryption and searchable encryption technique. The proposed scheme achieves not only privacy protection of data content, but also indistinguishability of data types. In the meanwhile, it enables that application *App* can simultaneously extract different types of data. Finally, security analysis show that the proposed scheme can achieve the corresponding security goals.

Keywords: Fog computing · Internet of Things · Data privacy · Data integrity · Selective data aggregation

1 Introduction

The appearance of the Internet of Things is one of the more remarkable phenomena of recent years. The Internet of Things refers to entities interconnected between heterogeneous entities, which may be sensors, devices, people or anything that requests or provides services. The IoT is changing the daily lives.

J. Xiong et al. (Eds.): MobiMedia 2021, LNICST 394, pp. 291–303, 2021.
https://doi.org/10.1007/978-3-030-89814-4_22

Through IoT devices which collect sensing data, IoT can provide real-time intelligent decision to better traffic conditions and forecast the weather. Up now, the rapid deployment of commercial 5G cellular networks offers a range of benefits to the Internet of Things that 4G or other technologies cannot offer.

The ultra-reliability and low latency of 5G will make self-driving cars, smart energy grids, factory automation and other demanding applications a reality. However, to realize these applications, a massive number of connected IoT devices need to be deployed, which will generate a large amount of data. It is reported that there will be 41 billion IoT devices by 2027. The data growth rate has been explosive due to consumer adoption and demand. If IoT devices with insufficient security design are connected to the 5G network, such massive growth in data traffic and connected IoT devices means more vulnerabilities, threats, and attacks resulting in catastrophic damages on financial markets and people's daily life. In addition, the diversity of the deployed nodes and access mechanisms at the edge of networks may result in some novel security challenges since the 5G networks have moved from centralized, hardware-based switching to distributed, software-defined digital routing. These problems pose an important challenge how to securely store, communicate, and compute these volumes of data.

To address these problems, fog computing paradigm [1] is proposed to enhance the IoT applications and to satisfy ultra-low delays requirement in 5G networks [2,3]. Due to being closer to where data is created and acted upon, fog computing makes some real-time and heterogeneous IoT applications feasible and practical [3]. Although fog computing overcomes the limitations of IoT devices and enables us to design a more capable architecture, it still unavoidably faces many security and privacy issues. As a non-trivial extension of cloud, some security and privacy issues in the context of cloud computing [4–7,10–28], still exist in the fog computing. Compared with the traditional Internet of things, the fog-assisted IoT confronts more complex network environment and network architecture. In addition to the traditional gateways for one application and fixed data sources, fog nodes need to collect data from multiple data sources and provide new aggregation services (selective data aggregation) for different applications, with different data types as intermediaries. For example, fog nodes (such as cellular base stations and roadside units) collect both the patient's physical condition data (such as heart rate and pulse rate) and road condition and traffic data (such as speed and traffic flow) to support disease monitoring applications and traffic sensing applications. It can not only provide real-time medical services to individual or community, but also improve the ability of healthcare organizations to monitor, track and control certain diseases on some regions. For a fog node, in order to conduct these different types of data, it first needs to distinguish different data types, and then performs data processing on the same type of data. In data processing procedure, both the data type and data content should be protected since the data type also leaks the privacy in an implied way [8,9], especially in the situations where the data sources come from the electronic bracelet. Obviously, traditional data aggregation techniques do

not satisfy the kind of selective data processing. However, this kind of selective data aggregation is one of the most important operations in the statistics of data aggregation and data analysis. It is able to be used to analyze the difference of data traffic among different time slots in a certain App application. Thus, it is significant to study this kind of data aggregation technology.

To achieve selective data aggregation construction, in this work, we proposed a novel privacy-preserving selective data aggregation scheme with revocation for fog-assisted IoT by combining homomorphic encryption and signature. It can not only cope with both the data privacy and the data integrity, but also achieve the revocation of application Apps. And then we also analyze the security of the proposed scheme, the result show that our proposed scheme can achieve data privacy and data integrity. Finally, the proposed scheme can achieve better performance by experimental testing.

The rest of this paper is organized as follows. Section 2 give related background knowledge. Section 3 gives the detailed construction of the proposed scheme. Then, Sect. 4 presents security analysis of the proposed scheme. Section 5 analyzes the simulation results. Finally, Sect. 5 concludes this paper.

2 Preliminaries

In this part, we first give our network architecture, threat model, and identify our design goals. And then some related basic primitives are introduced.

2.1 Network Architecture

Take full advantage of cloud computing and fog computing, our IoT network architecture is a three-tier architecture, cloud layer, fog node layer and terminals layer. It is composed of four types of entities: a trust authority, a group of heterogeneous IoT devices, the deployed fog nodes at the network edges, and some application App which is run on cloud platform. Their detailed architecture is shown in Fig. 1.

1. **Trust authority (TA):** It is a trusted third party, and responsible to initialize the system and generate key materials for the other entities.
2. **IoT device:** They are terminal devices with embedded sensors and communication module, and can periodically gather and submit their sensing data to application App via the fog node. In general, IoT devices may be some fixed monitoring sensors, electronic bracelet, moving vehicles and so on, and provide various sensing data according to the detailed application requirements.
3. **Fog nodes:** They are deployed at the local network edges and serve as the middle-ware between IoT devices and application App in cloud. supporting latency data response) and storage capability (e.g. storing some data for data process).
4. **Application App:** The application App is some kinds of softwares. It can gather the sensing data from IoT devices via the fog nodes, and conduct data analysis according to specific requirements. And then some decision can be made or some system performance can be improved by analytic results.

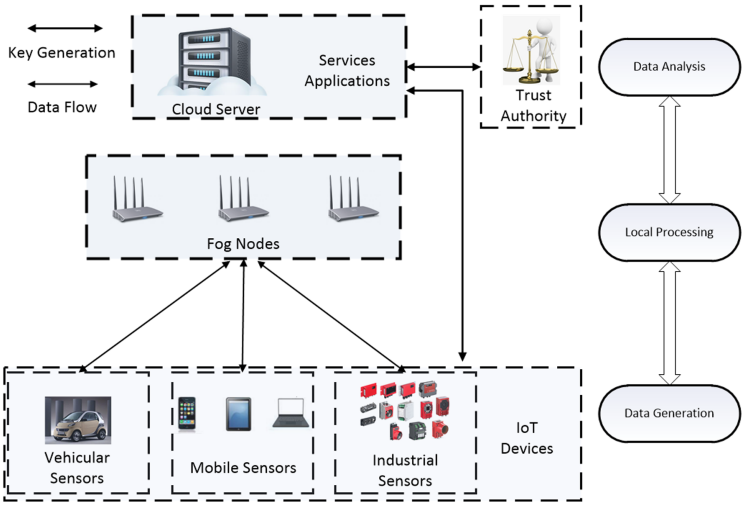

Fig. 1. Three-tried architecture of cloud and fog-based IoT network

2.2 Threat Model and Security Requeriments

In our network architecture, the trusted authority (TA) is a fully trusted-entity, the IoT devices and application Apps are honest and do not collude with the fog nodes, fog nodes are the honest-but-curious entity which follows protocol but will try to learn as much information as possible, without actively "cheating". And they should satisfy the following security requirements.

1. **Data Privacy:** The data privacy involves two aspects: data type privacy and sensing data privacy. Data type privacy means that, given sensing data, fog nodes can not determine its data type. Sensing data privacy indicates that sensing data and the aggregated result should satisfy confidentiality for fog node, and cannot be leaked in the data aggregation process.
2. **Data Integrity:** It means that the attackers can not forge and tamper the sensing data and the aggregated result.
3. **Indistinguishability:** It means that fog node cannot distinguish Whether $w_i = w_j$ under the condition that the ciphertexts $c_i = E(w_i)$ and $c_j = E(w_j)$ of data types w_i and w_j are given.

2.3 Design Goals

To construct privacy-preserving selective data aggregation scheme for fog-assist IoT networks, our goals are given as follows:

1. Security: The proposed scheme should satisfy data privacy and data integrity. Data privacy ensures that the data type and data content are confidential; and data integrity ensure that the sensing data and the aggregated result can not be tampered and modified by the attackers.

2. Efficiency: It means that all entities should each entity produces as low computational cost and communication cost as possible. And entities interact with each other as little as possible. Complex computation should be offloaded to fog node.
3. Easy Deployment: The scheme should ensure each entity to be easily deployed. Namely, the fog-assist network architecture should provide different applications, and the resource-limited IoT devices can expediently perform their key management.
4. Revocation: When an *App* application is revoked, fog node can recognize the service request from the IoT devices and delete the corresponding ciphertext.

2.4 Modified Paillier Homomorphic Cryptosystem

Paillier encryption is a kind of public key encryption scheme based on composite residuos classes. The security of the scheme is based on the difficulty to factor a big composite number N. The detail is given as follows: Let $(N, g, h = g^{\theta} \mod N^2)$ be public key, where $N = p \cdot q$, $p = 2p' + 1, q = 2p' + 1$, p', q' are two prime numbers, $g = -a^{2N} \mod N^2$, $a \in Z_{N^2}, \theta \in [1, N^2/2]$ are two random numbers. And the order of g is $2p'q'$. Let m be a encrypted plaintext, the ciphertext C is computed as follows: randomly select a number $r \in [1, N/4]$ to compute

$$(C_1 = g^r \mod N^2, C_2 = h^r(1 + m \cdot N) \mod N^2$$

To decrypt the ciphertext C, the plaintext m can be recovered by the key θ:

$$m = L(C_2/T_1^{\theta} \mod N^2)$$

where $L(x) = \frac{x-1 \mod N^2}{N}$. Additionally, if p, q are known, then $\lambda = 2p'q'$ can be obtained, thus,

$$C_2^{\lambda} = g^{\lambda \cdot r}(1 + m\lambda N) = (1 + m\lambda N)$$

if $gcd(N, \lambda) = 1$, then $m = L(C_2^{\lambda}) \cdot \lambda^{-1}$.

3 The Proposed Scheme

In this section, we proposed a novel privacy-preserving selective data aggregation scheme with revocation by combining Modified Paillier cryptosystem, search encryption and Lagrange interpolating polynomial technique. It is comprised of five phases: *System setup*, *data generation*, *fog-assisted selective aggregation*, *data-reading and verification*, and *revocation*. For the sake of easy explanation, the data format of the report which is generated by IoT is (id_i, τ_i, m_i, T_i), where id_i is the identifier of IoT device D_i, m_i is the data content which IoT device generates, τ_i is data type of m_i and T is the time slot of the reported data.

3.1 System Setup

In this phase, TA needs to run **Key Generation** algorithm, **Key Distribution** algorithm and **Selective Aggregation Initialization** algorithm to build system parameters and assign secret key for each App_j and IoT device D_i, respectively.

Key Generation. To bootstrap the entire system parameters, The trust authority (TA) takes a security parameter λ_1 as input and outputs the Modified Paillier crytosystem parameters $(n, g = \theta^2 \mod n^2, h = g^x)$, where $x \in [1, (n^2)/2]$ is private key of TA, and $\theta \in Z_{n^2}$ is a random number. And then TA uses another security parameter λ_2 to produce two cyclic groups $\mathbb{G}_1$ and $\mathbb{G}_2$ with the same order p. Let ρ be a generator of group $\mathbb{G}_1$ and $e : \mathbb{G}_1 \times \mathbb{G}_1 \to \mathbb{G}_2$ be a bilinear map. H is a hash function. And TA randomly chooses $x_1, x_2, f_1, t \in Z_p$ and secretly constructs a polynomial

$$f(x) = x_1 + f_1 x$$

Finally, the public parameters PK is published as follow:

$$PK = (n, g, h, \mathbb{G}_1, \mathbb{G}_2, e, p, \rho, \rho_1, \rho_2, H, EK)$$

where $\rho_1 = \rho^{x_1}, \rho_2 = \rho^{x_2}$ and $EK = \rho^{f(t)/x_1}$.

Key Distribution. For each application service App_i, TA picks a number $t_i \in Z_p$ at random and builds a subset $\Theta_i = t_i \cup t$. And then it makes use of the Lagrange interpolation theorem to calculate App_j's secret key $SK_i = (SK_{i1}, SK_{i2})$ as follows:

$$SK_{i1} = \rho_2^{f(t_i) \cdot \triangle_{t_i, \Theta_i}(0)}, SK_{i2} = {\rho_2}^{x_1 \cdot \Delta_{t, \Theta_i}(0)}$$

And add (ID_i^{App}, t_i, SK_i) to the list $\mathcal{K}$, where ID_i^{App} is the identifier of App_i. And TA selects a large prime $P \in [1, n/4]$ which satisfies that $(1, P, P^2, \cdots, P^d)$ is a super-increasing sequence, and randomly picks $\tau_i \in \lambda(n^2)/8$ to set App_i's public key as $h_i^{App} = g^{\tau_i}$. At last, (ID_i^{App}, h_i^{App}) should be published.

For all IoT devices $D_i, i = 1, \cdots, d$, if they enjoy the services of App_j application, they must register to App_j. Therefore, App_j firstly selects a random number $\beta \in Z_p$ as secret key, and then it distributes β to all IoT devices via a secure channel. Finally, (id_i, ρ_2^{β}) is published, where id_i is the identifier of IoT device D_i.

Selective Aggregation Initialization. For an application service App_j, it is assumed to have κ kinds of data types and $\kappa < d$. Let $\{w_1, \cdots, w_\kappa\}$ denote its data types. For $i = 1$ to κ, App_j randomly picks $\xi_i \in Z_p$ to compute the following ciphertext on data type w_i:

$$T_{i1} = \rho_1^{\xi_i}, T_{i2} = H(ID_j^{App} || w_i)^{\xi_i}, T_{i3} = SK_{j2}^{\xi_i}, T_{i4} = SK_{j1}^{\xi_i}$$

Finally, it sends κ ciphertexts $\{T_{i1}, T_{i2}, T_{i3}, T_{i4}\}_{i=1,\cdots,\kappa}$ and the large prime P to fog nodes for selective aggregation operations.

3.2 Data Generation

To report its sensing data (id_i, m_i, T) to the application App_j at time period T, an IoT device D_i needs to calculate two ciphertexts: one is to produce the cihpertext of data type, the other is to produce the ciphertext of the sensing data.

Data Type Encryption. For each IoT device, if its sensing data falls into data type w_l of application App_j, then the IoT device D_i picks two random numbers $r_{l1}, r_{l2} \in Z_p$ to calculate the ciphertext C_l of $w_l||ID_j^{App}$ as

$$C_{l1} = \rho_2^{r_{l2}} \cdot H(w_l||ID_j^{App})^{r_{l1}}, C_{l2} = \rho_1^{r_{l1}}, C_{l3} = EK^{r_{l2}}, C_{l4} = \rho^{r_{l2}}$$

Sensing Data Encryption. To encrypt the sensing data m_i for App_j, the IoT device D_i chooses two random numbers $r_i', \hat{r}_i \in [1, n/4]$ to compute a Paillier ciphertext

$$\begin{aligned} c_i &= (c_{i1}, c_{i2}) = (g^{r_i'} \mod n^2, (h_j^{App})^{r_i'}(1 + m_i \cdot n) \mod n^2) \\ \sigma_i &= (\sigma_{i1}, \sigma_{i2}) = ((H(ID_j^{App}||T)^{\hat{r}_i} \cdot \rho_1^{m_i})^{\beta}, \rho_2^{\hat{r}_i}) \end{aligned}$$

Finally, the IoT device D_i broadcasts $(id_i, ID_j^{App}, c_i, C_l, \sigma_i)$ to the fog node, where $C_l = (C_{l1}, C_{l2}, C_{l3}, C_{l4})$.

3.3 Fog-Assisted Selective Aggregation

After time period T, fog node receives m reports which are from the IoT devices D_i $(i = 1, \cdots, \pi)$, where π denotes the number of IoT devices. And then it executes the following two sub-phases:

Data Types Selection. To select data type, fog node executes Algorithm 1 to select data types, and produces data aggregation of the same type of all sensing reports. Finally, the aggregated result (s_{j1}, s_{j2}) of each data types $w_i, (i = 1, \cdots, \kappa)$ is returned.

If the sensing report does not satisfy Eq. (1), it means that the sensing data type is matched with all data types of w_i; Then it is dropped.

$$e(C_{l1}, T_{j1}) \stackrel{?}{=} e(C_{l2}, T_{j2}) \cdot e(C_{l3}, T_{j3}) \cdot e(C_{l4}, T_{i4}) \tag{1}$$

Content Aggregation. According to all sensing reports and the above data type selection, fog node can achieves all sensing reports aggregation by the following process:

$$CT = (CT_1, CT_2) = (\prod_{i \in \{1, \cdots, \kappa\}} s_{i1}^{P^i}, \prod_{i \in \{1, \cdots, \kappa\}} s_{i2}^{P^i})$$

For the sensing reports' signatures, they can be aggregated into

$$\sigma = (\sigma_1, \sigma_2) = (\prod_{i\in\{1,\cdots,m\}} \sigma_{i1}, \prod_{i\in\{1,\cdots,m\}} \sigma_{i2})$$

The reason that all signatures can be aggregated is that all IoT devices share a secret key β.

Input: (T_i)=$\{T_{i1}, T_{i2}, T_{i3}, T_{i4}\}, i = 1, \cdots, \kappa$ and $\{C_{i1}, C_{i2}, C_{i3}, C_{i4}\}, i = 1, \cdots, m$
Output: $(s_{j1}, s_{j2}), j = 1, \cdots, \kappa$

```
for (j = 1; j ≤ κ; j++) do
    s_{j1} = 1;
    s_{j2} = 1;
    for (l = 1; l ≤ m; l++) do
        if e(C_{l1}, T_{j1}) ?= e(C_{l2}, T_{j2}) · e(C_{l3}, T_{j3}) · e(C_{l4}, T_{i4}) then
            s_{j1} = s_{j1} · c_{l1}  mod n^2;
            s_{j2} = s_{j2} · c_{l2}  mod n^2;
        end
    end
end
return (s_{j1}, s_{j2}), j = 1, ..., κ;
```

Algorithm 1: Data type selection algorithm

At last, the fog node forwards the selection aggregation results (CT, σ) to application App_j.

3.4 Data Reading and Verification

After receiving the aggregated results at time period T, application App_j can execute data reading and verification by its secret keys.

First, App_j uses its secret key x to decrypt the aggregated results of each data type by the following steps:

1. It computes
$$M = \frac{CT_2}{CT_1^x} = 1 + (\sum_{i=1}^{\kappa} M_i P^i)n \mod n^2$$
where M_i denotes the aggregated result of all sensing reports with data type w_i.
2. Then, it recovers $\sum_{i=1}^{\kappa} M_i P^i = \frac{M-1 \mod n^2}{n}$, And execute Algorithm 2 to recover the aggregated result of each data type.
3. Finally, The signature of aggregated result can also be verified by the following equation
$$e(\sigma_1, \rho_2) = (e(H(Id_j^{App}||T), \sigma_2) e(\rho_1, \sigma_2)^{\sum_{i=1}^{\kappa} M_i})^{\beta}$$

Input: $M = M_1 + M_2P + \cdots + M_{\kappa-1}P^{\kappa-1}$, a super-increasing sequence $(1, P, P^2, \cdots, P^{\kappa-1})$ with $M_i < P - 1$
Output: $(M_1, \cdots, M_\kappa)$
1 set $\Phi_{\kappa-1} = M$;
2 **for** $(j = \kappa; j > 1; j--)$ **do**
3 $\quad \Phi_{j-2} = \Phi_{j-1} \mod P^{j-1}$;
4 $\quad M_j = \frac{\Phi_{j-1} - \Phi_{j-2}}{P^{j-1}}$;
5 **end**
6 $M_1 = \Phi_0$;
7 return $(M_1, M_2, \cdots, M_\kappa)$;

Algorithm 2: Recover the aggregated report of all data types

3.5 Revocation

For achieving revocation, we introduce Revocation List (RL) to design a light-weight mechanism for application revocation. If an application App_j is were taken off the shelves and suspended, TA produces a revocation token $RvT_j = (SK_{j1}^k, SK_{j2}^k)$, where $k \in Z_p$ is a random number, and then it added RvT_j to RL. Note RL is broadcasted to all fog nodes.

In selective aggregation initialization phase, upon receiving $\{T_{i1}, T_{i2}, T_{i3}, T_{i4}\}$, fog node makes use of revocation token RvT_j in the RL and checks the following relation

$$e(T_{i3}, SK_{j1}^k) \stackrel{?}{=} e(SK_{j2}^r, T_{i4}) \tag{2}$$

If there exists a revocation token which make Eq. (2) true, it outputs 1; if Eq. (2) is false for all the revocation tokens in RL, it outputs 0.

4 Security Analysis

In the following, we first discuss the correction of data type selection, then demonstrate the security of the proposed scheme in terms of privacy and confidentiality of sensing data.

Theorem 1. For a sensing report, if $(C_{l1}, C_{l2}, C_{l3}, C_{l4})$ is the ciphertext of its data type w_l, then it must satisfy the verification equation Eq. (1).

Proof. Since the generated cihpertext of w_l by application App_j is $\{T_{l1}, T_{l2}, T_{l3}, T_{l4}\}$, then we have

$$\begin{aligned} & e(C_{l1}, T_{l1}) \\ = & e(\rho_2^{r_2} H(w_l)^{r_1}, \rho_1^{\xi_l}) = e(\rho_2^{r_2}, \rho_1^{\xi_l}) e(H(w_l)^{r_1}, \rho_1^{\xi_l}) \end{aligned}$$

$$e(C_{l2}, T_{l2}) = e(\rho_1^{r_1}, H(w_l)^{\xi_l})$$

$$e(C_{l3}, T_{l3}) = e(\rho^{\frac{f(t)\cdot r_2}{\alpha_1}}, \rho_2^{\alpha_1 \cdot \xi_l \cdot \Delta_{t,\Theta_i}(0)}) = e(\rho, \rho_2)^{r_2 \xi_l f(t) \cdot \Delta_{t,\Theta_i}(0)}$$

$$e(C_{l4}, T_{l4}) = e(\rho^{r_2}, \rho_2^{\xi_l f(t_i) \Delta_{t_i,\Theta_i}(0)}) = e(\rho, \rho_2^{r_2 \xi_l f(t_i) \Delta_{t_i,\Theta_i}(0)})$$

According to Lagrange interpolation formula, we can know

$$\begin{aligned} & e(C_{l3}, T_{l3}) e(C_{l4}, T_{l4}) \\ &= e(\rho, \rho_2)^{r_2 \xi_l (f(t_i)\Delta_{t_i,\Theta_i}(0) + f(t)\Delta_{t,\Theta_i}(0))}) \\ &= e(\rho, \rho_2)^{\alpha_1 r_2 \xi_l} = e(\rho_1, \rho_2)^{r_2 \xi_l} \end{aligned}$$

Thus, we have $e(C_{l1}, T_{l1}) = e(C_{l2}, T_{l2}) e(C_{l3}, T_{l3}) e(C_{l4}, T_{l4})$ □

In the following, we will discuss data privacy and data integrity.

Sensing Data Privacy. In our three-tier architecture, to ensure the privacy of the transmitted data between fog node and IoT device, we adopt Paillier public encryption algorithm to encrypt the sensing report m_i from IoT devices. Because fog node does not know the corresponding private key, it makes that fog node cannot decrypt the corresponding ciphertext. In addition, although fog node can obtain the ciphertext $(g^{r_i'}, h_j^{App} = g^{\tau_i}, (h_j^{App})^{r_i'}(1 + m_i \cdot n) \mod n^2)$, it is impossible to extract m_i from the ciphertext since to extract m_i needs that an adversary must know $(h_j^{App})^{r_i'}$. However, given $(g^{r_i'}, h_j^{App} = g^{\tau_i})$, to obtain $(h_j^{App})^{r_i'}$ is equivalent to solving computational Diffie-Hellman problem. Obviously, it is in contradiction with the difficulty of solving computational Diffie-Hellman assumption. Thus, sensing data's privacy is preserved.

For data type w_i, to ensure the privacy and secure matching of data type, we use secure encryption scheme with keyword search to encrypt data type, and make use of the search ability on the encrypted data to achieve the matching of data type. It appears in the form of the ciphertext $\{C_{l1}, C_{l2}, C_{l3}, C_{l4}\}$. Because the adopted searchable encryption scheme is semantically secure against an adaptive chosen keyword attack, Fog node can obtain nothing about data type w_i from $\{C_{l1}, C_{l2}, C_{l3}, C_{l4}\}$ in data type selection phase. In addition, given two ciphertexts C_l and C_l' of two data types, Fog node cannot also distinguish whether these two ciphertexts correspond to the same data type. For these two ciphertexts, they has the following formats:

$$C_{l1} = \rho_2^{r_{l2}} \cdot H(w_l || ID_j^{App})^{r_{l1}}, C_{l2} = \rho_1^{r_{l1}}, C_{l3} = EK^{r_{l2}}, C_{l4} = \rho^{r_{l2}}$$
$$C_{l1}' = \rho_2^{r_{l2}'} \cdot H(w_l' || ID_j^{App})^{r_{l1}'}, C_{l2}' = \rho_1^{r_{l1}'}, C_{l3}' = EK^{r_{l2}'}, C_{l4}' = \rho^{r_{l2}'}$$

where $r_{l1}, r_{l2}, r_{l1}', r_{l2}'$ are random numbers. To obtain the relation between the ciphertexts C_l' and C_l, an adversary must obtain $\rho_2^{r_{l2} - r_{l2}'}$ since given $\rho_2^{r_{l2} - r_{l2}'}$,

an adversary can determine the relation of C_l and C_l. The reason is that the size of data type space is usually polynomial or low-entropy distribution, and an adversary can check whether the equation

$$e(\frac{C_{l1}/C'_{l1}}{\rho_2^{r_{l2}-r'_{l2}}}, \rho_1) = e(H(w_l||ID_j^{App}), \rho_1^{r_{l2}-r'_{l2}})$$

holds by exhaustive search attack. However, given $(\rho, \rho_2 = \rho^{x_2}, \rho^{r_{l2}-r'_{l2}})$, it is equivalent to solving the Computational Diffie-Hellman problem to obtain $\rho_2^{r_{l2}-r'_{12}}$. Obviously, it is inconsistent with the difficulty of solving the Computational Diffie-Hellman problem.

In summary, our scheme can preserve privacy of the content and achieve indistinguishability of data type.

Data Integrity. In content aggregation phase, fog node can aggregate all ciphertexts $c_i, i = 1, \cdots, m$ into a ciphertext CT by using homomorphism of Paillier encryption scheme, but fog node can not tamper/modify the IoT devices' aggregation results since linear homomorphic signature scheme is required in data verification phase. In our proposed scheme, the improved Paillier encryption only supports Additive homomorphic property, and the adopted homomorphic signature scheme only supports multiplicative homomorphic property. It enable that a tampered sensing report m'_i can not pass the verification of signature in data verification phase. Thus, our proposed scheme can achieve sensing data integrity.

Revocation. After an application App_i is revoked, TA needs to publish the corresponding revocation token $RvT_i = (SK_{i1}^r, SK_{i2}^r)$ and add it in revocation list RL. Fog node deletes all the ciphertext of data types which correspond to the suspended application App_i by checking the relation according to the revocation token in the updated RL.

$$e(T_{i3}, SK_{j1}^k) \stackrel{?}{=} e(SK_{j2}^r, T_{i4}) \tag{3}$$

5 Conclusion

In this paper, we have proposed a privacy preserving selective data aggregation scheme with revocation for fog-assisted IoT. Then, our scheme has been proposed and designed particularly based on the features of fog computing and IoT, to guarantee data privacy and data integrity distributed fog nodes, and multiple application services. In the future work, we focus on selective forwarding attack: the attackers only selectively aggregate part of the data and forward the incomplete results to the application service.

Acknowledgement. This work is supported in part by The Natural Science Foundation of Beijing (No. 4212019), National Natural Science Foundation of China (No. 62172005), Guangxi Key Laboratory of Cryptography and Information Security (No. GCIS201808) and Foundation of Guizhou Provincial Key Laboratory of Public Big Data (No. 2019BDKF JJ012).

References

1. Dastjerdi, A.V., Buyya, R.: Fog computing: helping the internet of things realize its potential. IEEE Comput. **49**(8), 112–116 (2016)
2. Morocho, M., Lee, H., Lim, W.: Machine Learning for 5G/B5G mobile and wireless communications: potential, limitations, and future directions article. IEEE Access **7**, 137184–137206 (2020)
3. Chiang, M., Zhang, T.: Fog and IoT: an overview of research opportunities. IEEE Internet Things J. **3**(6), 854–864 (2016)
4. Takabi, H., Joshi, J.B., Ahn, G.J.: Security and privacy challenges in cloud computing environments. IEEE Secur. Priv. **8**, 271–350 (2010)
5. Huang, C., Liu, D., Shen, S.: Reliable and privacy-preserving selective data aggregation for fog-based IoT. In: 2018 IEEE International Conference on Communications (ICC), pp. 1–6 (2018)
6. Boneh, D., Gentry, C., Lynn, B., Shacham, H.: Aggregate and verifiably encrypted signatures from bilinear maps. In: Biham, E. (ed.) EUROCRYPT 2003. LNCS, vol. 2656, pp. 416–432. Springer, Heidelberg (2003). https://doi.org/10.1007/3-540-39200-9_26
7. Zhang, J., Zhu, J., Zhang, N.: An improved privacy-preserving collaborative filtering recommendation algorithm. In: Proceedings of Asia Conference on Information Systems, pp. 277–288 (2014)
8. Alghamdi, W.Y., Wu, H., Kanhere, S.S.: Reliable and secure end-to-end data aggregation using secret sharing in WSNs. In: IEEE WCNC, pp. 1–6 (2017)
9. Qian, J., Qiu, F., Wu, F., Ruan, N., Chen, G., Tang, S.: Privacy-preserving selective aggregation of online user behavior data. IEEE Trans. Comput. **66**(2), 326–338 (2017)
10. Hu, H., Lu, R., Zhang, Z., Shao, J.: REPLACE: a reliable trust-based platoon service recommendation scheme in VANET. IEEE Trans. Veh. Tech. **66**(2), 1786–1797 (2017)
11. Wang, K., Shao, Y., Shu, L., Zhu, C., Zhang, Y.: Mobile big data fault-tolerant processing for eHealth networks. IEEE Netw. **30**(1), 36–42 (2016)
12. Zhang, H., Qiu, Y., Long, K., Karagiannidis, G.K., Wang, X., Nallanathan, A.: Resource allocation in NOMA based fog radio access networks. IEEE Wirel. Commun. **25**(3), 110–115 (2018)
13. Zhang, J., Zhang, Q., Ji, S.: A fog-assisted privacy-preserving task allocation in crowdsourcing. IEEE Internet Things J. **7**(9), 8331–8342 (2020)
14. Zhang, J., Bai, W., Wang, Y.: Non-interactive ID-based proxy re-signature scheme for IoT based on mobile edge computing. IEEE Access **7**, 37865–37875 (2019)
15. Liu, X., Deng, R.H., Choo, K.K.R., Weng, J.: An efficient privacy-preserving outsourced calculation toolkit with multiple keys. IEEE Trans. Inf. Forensics Secur. **11**(11), 2401–2414 (2016)
16. Lu, R., Heung, K., Lashkari, A.H., Ghorbani, A.A.: A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT. IEEE Access **5**, 3302–3312 (2017)

17. Shen, H., Zhang, M., Shen, J.: Efficient privacy-preserving cube-data aggregation scheme for smart grids. IEEE Trans. Inf. Forensics Secur. **12**(6), 1369–1381 (2017)
18. Choubin, M., Taherpour, A., Rahmani, M.: Collaborative data aggregation using multiple antennas sensors and fusion centre with energy harvesting capability in WSN. IET Commun. **13**(13), 1971–1979 (2019)
19. Rezaeibagha, F., Yi, M., Huang, K., Chen, L.: Secure and efficient data aggregation for IoT monitoring systems. IEEE Internet Things J. (2020). https://doi.org/10.1109/JIOT.2020.3042204
20. Yan, O., Liu, A., Xiong, N., Wang, T.: An effective early message ahead join adaptive data aggregation scheme for sustainable IoT. IEEE Trans. Netw. Sci. Eng. (2020). https://doi.org/10.1109/TNSE.2020.3033938
21. Boneh, D., Goh, E.-J., Nissim, K.: Evaluating 2-DNF formulas on ciphertexts. In: Kilian, J. (ed.) TCC 2005. LNCS, vol. 3378, pp. 325–341. Springer, Heidelberg (2005). https://doi.org/10.1007/978-3-540-30576-7_18
22. Zhou, J., Cao, Z., Dong, X., Lin, X.: Security and privacy in cloud-assisted wireless wearable communications: Challenges, solutions, and future directions. IEEE Wireless Commun. **22**(2), 136–144 (2015)
23. Yi, X., Bouguettaya, A., Georgakopoulos, D., Song, A., Willemson, J.: Privacy protection for wireless medical sensor data. IEEE Trans. Dependable Sec. Comput. **13**(3), 369–380 (2016)
24. Bao, H., Lu, R.: A new differentially private data aggregation with fault tolerance for smart grid communications. IEEE Internet Things J. 2(3), 248–258 (2015)
25. Tang, W., Ren, J., Deng, K., Zhang, Y.: Secure data aggregation of lightweight e-healthcare IoT devices with fair incentives. IEEE Internet Things J. **6**(5), 8714–8726 (2019)
26. Xiong, J., et al.: Enhancing privacy and availability for data clustering in intelligent electrical service of IoT. IEEE Internet Things J. **6**(2), 1530–1540 (2019)
27. Xiong, J., et al.: A personalized privacy protection framework for mobile crowdsensing in IIoT. IEEE Trans. Industr. Inf. **16**(6), 4231–4241 (2020)
28. Xiong, J., Chen, X., Yang, Q., Chen, L., Yao, Z.: A task-oriented user selection incentive mechanism in edge-aided mobile crowdsensing. IEEE Trans. Netw. Sci. Eng. **7**(4), 2347–2360 (2020)