IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON SECURITY AND PRIVACY IN APPLICATIONS AND SERVICES FOR FUTURE INTERNET OF THINGS

Received January 25, 2017, accepted February 19, 2017, date of publication March 2, 2017, date of current version March 28, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2677520

INVITED PAPER

A Lightweight Privacy-Preserving Data Aggregation Scheme for Fog Computing-Enhanced IoT

RONGXING LU, (Senior Member, IEEE), KEVIN HEUNG, ARASH HABIBI LASHKARI, (Member, IEEE), AND ALI A. GHORBANI, (Senior Member, IEEE)

Faculty of Computer Science, University of New Brunswick, Fredericton, NB E3B 5A3, Canada

Corresponding author: R. Lu (rlu1@unb.ca)

ABSTRACT Fog computing-enhanced Internet of Things (IoT) has recently received considerable attention, as the fog devices deployed at the network edge can not only provide low latency, location awareness but also improve real-time and quality of services in IoT application scenarios. Privacy-preserving data aggregation is one of typical fog computing applications in IoT, and many privacy-preserving data aggregation schemes have been proposed in the past years. However, most of them only support data aggregation for homogeneous IoT devices, and cannot aggregate hybrid IoT devices' data into one in some real IoT applications. To address this challenge, in this paper, we present a lightweight privacy-preserving data aggregation scheme, called Lightweight Privacy-preserving Data Aggregation, for fog computing-enhanced IoT. The proposed LPDA is characterized by employing the homomorphic Paillier encryption, Chinese Remainder Theorem, and one-way hash chain techniques to not only aggregate hybrid IoT devices' data into one, but also *early* filter injected false data at the network edge. Detailed security analysis shows LPDA is really secure and privacy-enhanced with differential privacy techniques. In addition, extensive performance evaluations are conducted, and the results indicate LPDA is really lightweight in fog computing-enhanced IoT.

INDEX TERMS Internet of Things, fog computing, privacy-preserving aggregation, lightweight, differential privacy.

I. INTRODUCTION

The advancement and wide deployment of Internet of Things (IoT) have revolutionized our lifestyle greatly by providing the most convenience and flexibility in our various daily applications. The typical applications of IoT include smart grid [1], smart healthcare [2], smart home [3], smart city [4], and even smart nation [5]. Although these IoT applications have their unique characteristics in their respective fields, their essences are the same, i.e., any IoT application is an interconnected network formed by a number of IoT devices, which can not only collect nearly real-time data but also exchange them in time for achieving better and intelligent decisions. For example, in smart grid application, smart meters are deployed at all homes in a residential area, each smart meter can collect user's electricity use data, and periodically (e.g., every 15 minutes [6]) report to the control center, and the control center can base the reported data to make real-time data analytics and take the corresponding actions to guarantee the health of power system.

Clearly, IoT will be of great benefit to our daily lives. However, in order to fully take advantage of IoT, we have to address some challenges lying ahead in IoT [7]. Many IoT applications generate huge volumes of data for real-time analytics, as a result, IoT is a big data problem [8]. In order to manage and analyze huge volumes of data and derive potential values from IoT, we need to consider more suitable real-time big data mining and machine learning techniques for IoT insights. At the same time, IoT is not just about big data analytics, but also about the connected IoT devices and the data transmission from IoT devices to the control center.

2169-3536 © 2017 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

The more the real-time data report from IoT devices, the better the decision can be made from IoT. However, in order to fit the real-time data report, it will cost huge communication resources. Even worse, when false data are injected in IoT [9], it not only wastes the scarce communication bandwidth, but also causes inaccurate decisions made at the control center. Therefore, desirable mechanisms are expected to address the challenges, i.e., to reduce the communication costs and *early* filter the false injected data during the IoT data report.

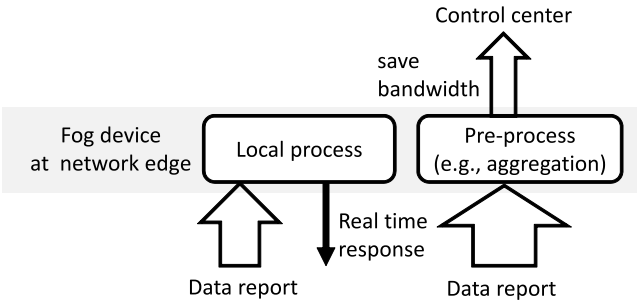


FIGURE 1. Fog computing paradigm extends the cloud computing capabilities to the network edge to provide i) real time response with local process, and ii) bandwidth saving with pre-process (e.g., aggregation).

The concept of fog computing was proposed by Cisco in 2012 [10], which aims to solve problems or pre-process parts of a problem by keeping data at the network edge with local fog devices (e.g., Cisco next generation routers), rather than routing everything through a central control center. Because of its network edge computing feature, fog computing can not only provide low latency, location awareness but also improve real-time and quality of services in network-based applications [11]–[13]. Since the objective of fog computing is not to replace the powerful cloud computing, but to extend the computing capabilities to the network edge, as shown in Fig. 1, the IoT challenges (bandwidth and security) mentioned-above can be resolved by fog computing paradigm. For example, we can deploy a fog device (as a gateway) at the network edge, which is in charge of collecting all IoT devices' data, aggregating and forwarding them to the central control center, so as to save the bandwidth. In addition, in case there exist injected false data by external attackers, the fog device can also execute source authentication to *early* filter injected false data at the network edge. Although the above fog computing-enhanced IoT is promising, the trust issue of fog device cannot be ignored. Since it is deployed at the network edge, the fog device cannot be fully trusted. Therefore, in some IoT applications, e.g., smart grid, in order to protect each individual IoT device's data, the fog device cannot see each individual data when running the aggregation. As a result, the privacy-preserving data aggregation schemes are desirable in IoT applications.

Soothly, when the buzz word "fog computing" becomes popular recently, several privacy-preserving data aggregation schemes have been proposed [1], [6], [14]–[27], and they can be well fit in fog computing-enhanced IoT. However, most of them only support one type of aggregation (e.g, sum or mean) for homogenous IoT devices. In case there are hybrid IoT devices in some real IoT applications, those schemes cannot aggregate all data into one. Therefore, there is a high desire to design a privacy-preserving data aggregation for hybrid IoT devices.

In this paper, in order to address the above challenge, we propose a new lightweight privacy-preserving data aggregation scheme, called Lightweight Privacy-preserving Data Aggregation (LPDA), for fog computing-enhanced IoT. The proposed LPDA is characterized by employing the Chinese Remainder Theorem to aggregate hybrid IoT devices' data into one, and use the one-way hash chain to run the source authentication at the network edge to *early* filter the injected false data. Specifically, the contributions of this paper are threefold as follows.

- First, we propose our LPDA aggregation scheme for fog computing-enhanced IoT, which combines the homomorphic Paillier encryption [28], Chinese Remainder Theorem, and one-way hash chain techniques to enable a fog device at the network edge to not only aggregate hybrid IoT devices' data into one for saving scarce bandwidth, but also *early* filter injected false data for security enhancement.
- Secondly, we give the detailed analysis to show that our proposed LPDA is really secure under our defined security model. Particularly, we also use the differential privacy techniques [29] to enhance the privacy-preservation of LPDA, i.e., LPDA can also resist differential attacks.
- Thirdly, we not only theoretically analyze the communication overheads of LPDA but also run extensive experiments to evaluate the computational costs of LPDA, and the results indicate LPDA is really lightweight and suits in fog computing-enhanced IoT.

The remainder of this paper is organized as follows. In Section II, we introduce our system model, security model and design goal. Then, we describe some preliminaries in Section III. In Section IV, we present our LPDA scheme, followed by security analysis and performance evaluation in Section V and Section VI, respectively. Related work is discussed in Section VII. Finally, we draw our conclusion in Section VIII.

II. MODELS AND DESIGN GOAL

In this section, we formalize our system model, security model, and identify our design goal.

A. SYSTEM MODEL

In our system model, we consider a hybrid IoTs network, which includes a set of heterogeneous IoT devices $\mathcal{D} = \{D_1, D_2, D_2, \cdots, D_N\}$, a fog device deployed at the network edge, a control center, and a trusted authority, as shown in Fig. 2.

- IoT devices $\mathcal{D} = \{D_1, D_2, D_2, \cdots, D_N\}$: a set of IoT devices are deployed at an area of interest, each device D_i is equipped with sensing and communication

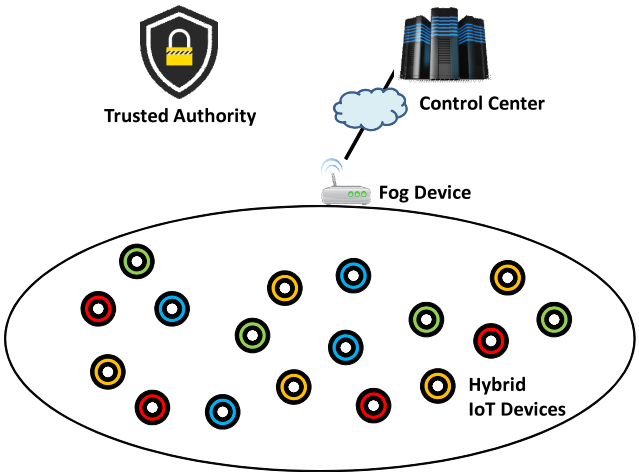


FIGURE 2. System model under consideration.

capabilities, which enables D_i to periodically report its sensing result x_i to the control center via the fog device. Due to the nature of heterogeneity, we can further divide IoT devices $\mathcal{D}$ into to several subsets according to their sensing functions, i.e., all IoT devices with the same sensing functions will be placed in the same subset. Without loss of generality, we assume $\mathcal{D}$ can be divided into k subsets: $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \cdots, \mathcal{D}_k$, where the size of $\mathcal{D}_i$ is $|\mathcal{D}_i| = N_i$, $\bigcup_{i=1}^{k} \mathcal{D}_i = \mathcal{D}$ and $\mathcal{D}_i \bigcap \mathcal{D}_j = \phi$ for any $i \neq j$. Note that, since IoT devices are usually not powerful, we cannot deploy time-consuming security algorithms, and thus lightweight security mechanisms are desirable to mount on these IoT devices.

- Fog device: In hybrid IoTs, the fog devices is a critical component for fog computing, which is deployed at network edge and serves as the relay between the IoT devices and the control center. In particular, the fog device will accomplish some fog computing functions, e.g., aggregate all IoT devices' data $(x_1, x_2, \cdots, x_N)$ and forward them to the control center, and also help the control center to locally filter some injected data from external attackers.
- Control center: The control center receives all IoT devices' data $(x_1, x_2, \cdots, x_N)$ via the fog device, and makes some data analytics according to some application requirements. Since all data come from heterogenous IoT devices, it is not accurate to directly operate on all data. Therefore, for each subset $\mathcal{D}_j \subset \mathcal{D}$, the control center will compute its mean $E(\mathcal{D}_j)$ and variance $Var(\mathcal{D}_j)$ in this work, where

$$E(\mathcal{D}_j) = \frac{\sum_{D_i \in \mathcal{D}_j} x_i}{N_j}, \; Var(\mathcal{D}_j) = \frac{\sum_{D_i \in \mathcal{D}_j} x_i^2}{N_j} - E(\mathcal{D}_j)^2$$

- Trusted authority: In our system model, the trusted authority is a trusted third party, whose duty is to bootstrap the system, manage key materials and assign keys to all IoT devices, the fog device, and the control center. Note that, after bootstrapping the system, the trusted authority will be offline, i.e., it will not participate in the subsequent actions.

B. SECURITY MODEL

In our security model, we consider the trusted authority is fully trusted, while the control center and fog device are *honest-but-curious*, because they may be affected by undetected malwares, and the malwares will eavesdrop IoT devices' data. Although the control center and fog device are *honest-but-curious*, i.e., they follow the protocols, but are also curious about IoT devices' data privacy, they will not collude with each other in this work.

Because IoT devices are usually not powerful, the security of IoT devices is always challenging. Since this work is focused on the privacy-preserving aggregation, we will not discuss the inside threats, that is, the IoT devices are not compromised in this work. Nevertheless, we still consider some IoT devices could be malfunctioning and stop reporting for some periods, and it is also possible for some external attackers to launch false data injection attack. Therefore, the fog device at the network edge is expected to filter these false data locally, and will not send them to the control center.

Note that, the external attackers may launch other active attacks, i.e., Denial of Service (DoS) attacks, to the hybrid IoT networks. Again, since the privacy-preserving aggregation is our focus, those active attacks are beyond the scope of this work, and will be discussed in future.

C. DESIGN GOAL

Under the aforementioned system model and security model, our design goal is to propose a lightweight privacy-preserving data aggregation scheme for hybrid IoT networks. In particular, the following four objectives should be achieved:

- Privacy: The proposed aggregation scheme should be privacy-preserving, that is, the control center can compute each subset $\mathcal{D}_j$'s mean and variance, and cannot obtain each individual IoT device's data.
- Security: The proposed aggregation scheme can resist against the false data injection attack from external attackers, that is, the fog device can filter false data locally at the network edge.
- Fault-Tolerance: The proposed aggregation scheme should be falut-tolerant, that is, even though some IoT devices are malfunctioning and stop reporting to the control center, the control center can still compute the mean and variance of the reported data in each subset.
- Efficiency: The proposed aggregation scheme should be efficient, that is, the computational costs at IoT devices, fog device, and control center should be as less as possible, and the communication overheads should also be minimal.

III. PRELIMINARIES

In this section, we briefly review the Chinese Remainder Theorem, one-way hash chain, and some properties under the

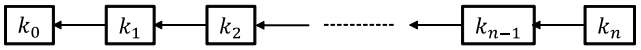


FIGURE 3. The structure of one-way hash chain.

modulo n^2, which will serve as the building blocks of the proposed LPDA scheme.

A. CHINESE REMAINDER THEOREM

The Chinese Remainder Theorem says that we can uniquely solve any pair of congruences that have relatively prime moduli, which enables us to devise an efficient data aggregation scheme for fog computing-enhanced IoT with hybrid IoT devices and is described as follows.

Theorem 1 (Chinese Remainder Theorem): Suppose that $q_1, q_2, \cdots, q_k$ are pairwise relatively prime positive integers, and let $a_1, a_2, \cdots, a_k$ be integers. Then, the system of congruences, $x \equiv a_i \bmod q_i$ for $1 \leq i \leq k$, has a unique solution modulo $Q = q_1 \times q_2 \times \cdots \times q_k$, which is given by

$$x \equiv a_1 Q_1 y_1 + a_2 Q_2 y_2 + \cdots + a_k Q_k y_k \bmod Q$$

where $Q_i = \frac{Q}{q_i}$ and $y_i \equiv \frac{1}{Q_i} \bmod q_i$ for $1 \leq i \leq k$.

B. ONE-WAY HASH CHAIN

One-Way Hash Chain is a very popular security technique, which has been widely discussed in data stream authentication, e.g., TESLA [30]. In this work, we will apply it to achieve lightweight authentication in hybrid IoTs.

Given a secure hash function $h : \{0, 1\}^* \rightarrow \{0, 1\}^l$, a one-way hash chain is defined as a set of values $\{k_0, k_1, \cdots, k_n\}$ for $n \in \mathbb{Z}$ such that $k_n \in \{0, 1\}^l$ is a randomly chosen value, and $k_i = h(k_{i+1})$ for $i = 0$ to $n - 1$.

Due to the one-wayness, given k_i in a hash chain, it is easy to compute k_j, where $j < i$; however, it is computationally infeasible to compute k_l, where $l > i$.

C. SOME PROPERTIES UNDER THE MODULO n^2

Let $p = 2p' + 1$ and $q = 2q' + 1$ be two safe primes, where p' and q' are also two primes. Compute $n = pq$, $\lambda = lcm(p - 1, q-1) = 2p'q'$, the least common multiple of $p-1$ and $q-1$. Then, we have the following properties under the modulo n^2.

1) For any $x \in \mathbb{Z}^*_{n^2}$, we have $x^{n\lambda} \equiv 1 \bmod n^2$.
2) For any $x_i \in \mathbb{Z}_n$, $i = 1, 2, \cdots, m$, we have

$$\prod_{i=1}^{m}(1 + n \cdot x_i) \equiv (1 + n \cdot \sum_{i=1}^{m} x_i) \bmod n^2 \qquad (1)$$

The first property has been applied in designing the popular Pailler Homomorphic Encryption, the detailed proof can be found in [28]. The second property can be easily proved with mathematical induction. When $m = 1$, the left side becomes $1 + n \cdot x_i$, which is equal to the right side. Assume, when $m = k$, we have $\prod_{i=1}^{k}(1 + n \cdot x_i) \equiv (1 + n \cdot \sum_{i=1}^{k} x_i) \bmod n^2$, we need to show it also holds for $m = k+1$. When $m = k+1$, the left side is

$$\prod_{i=1}^{k+1}(1 + n \cdot x_i) = \prod_{i=1}^{k}(1 + n \cdot x_i) \cdot (1 + n \cdot x_{k+1})$$

$$= (1 + n \cdot \sum_{i=1}^{k} x_i) \cdot (1 + n \cdot x_{k+1}) = (1 + n \cdot \sum_{i=1}^{k+1} x_i) \bmod n^2$$

which equals the right side and shows that the second property is also correct. In particular, when $m = \lambda$ and all $x_i = x$, we will have

$$(1 + n \cdot x)^{\lambda} \equiv (1 + n \cdot \lambda x) \bmod n^2 \qquad (2)$$

IV. PROPOSED LPDA SCHEME

In this section, we present our lightweight privacy-preserving data aggregation scheme (LDPA) for hybrid IoTs, which mainly consists of the following four parts: system initialization, IoT device report generation, fog device report aggregation, and control center report reading and analytics.

A. SYSTEM INITIALIZATION

As the trusted authority (TA) is a fully trusted entity in the system, it is reasonable to assume the TA bootstraps the whole system. Specifically, given the security parameters k_0, k_1, l, TA first randomly chooses two safe prime numbers p, q, where $p = 2p' + 1$, $q = 2q' + 1$, and $|p| = |q| = k_0$, compute $n = pq$, and $\lambda = lcm(p - 1, q - 1) = 2p'q'$, and defines a function $L(x) = \frac{x-1}{n}$. Then, consider there are total N IoT devices $\mathcal{D} = \{D_1, D_2, D_2, \cdots, D_N\}$ in the network, TA chooses $N + 2$ random numbers $s_0, s_1, s_2, \cdots, s_N, s_{N+1}$ such that

$$\sum_{i=0}^{N+1} s_i \equiv 0 \bmod \lambda \qquad (3)$$

Consider there are k subsets $\mathcal{D}_1, \mathcal{D}_2, \mathcal{D}_3, \cdots, \mathcal{D}_k$ in the network, and the IoT device's sensing data range of each subset $\mathcal{D}_j$ is $[0, X_j]$. Then, we can define $X = \max\{X_1, X_2, \cdots, X_k\}$. Note that, the range $[0, X]$ is still a small message space in comparison with $\mathbb{Z}_n$. With these knowledge, the TA chooses $k + 1$ prime numbers $\alpha_0, q_1, q_2, \cdots, q_k$, and computes

$$\begin{cases} Q = q_1 \times q_2 \times \cdots \times q_k \\ Q_i = \dfrac{Q}{q_i}, \quad y_i \equiv \dfrac{1}{Q_i} \bmod q_i \\ \alpha_i = Q_i \cdot y_i \end{cases} \qquad (4)$$

where each prime q_i is of the same length, i.e., $|q_i| = k_1$ for $1 \leq i \leq k$. The conditions of the above parameters follow

$$\begin{cases} N \cdot X^2 \leq \alpha_0, \quad N \cdot (X^2 + X \cdot \alpha_0) < q_i \\ k_1 \cdot (k + 1) + \lg k < |n| \end{cases} \qquad (5)$$

which enable us to aggregate all data into one ciphertext.

Next, the TA chooses two secure cryptographic hash functions h, H, where $h : \{0, 1\}^* \rightarrow \{0, 1\}^l$ and

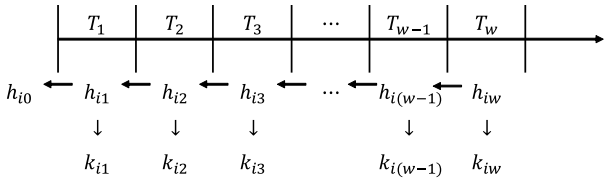


FIGURE 4. The time slot division and hash chain and key generation.

$H : \{0,1\}^* \rightarrow \mathbb{Z}_n^*$, and another random number $t_0 \in \{0,1\}^l$ as a secret key. As the IoT devices will periodically report their data to the control center. we divide the report time period into w time slots, as shown in Fig. 4, and at every time slot, each IoT device will report its sensing data. To cater for this setting, the TA chooses a random number t_0 and builds N one-way hash chains $\mathcal{HC}_1, \mathcal{HC}_2, \cdots, \mathcal{HC}_N$, each chain $\mathcal{HC}_i = \{h_{i0}, h_{i1}, h_{i2}, \cdots, h_{iw}\}$ is of length $w+1$, where $h_{iw} \in \{0,1\}^l$ is a randomly chosen number, and

$$h_{ij} = h(h_{i(j+1)}||T_j) \quad j = 0, 1, \cdots, w-1 \tag{6}$$

In addition, for each h_{ij}, $1 \leq j \leq w$, the TA also computes its corresponding key

$$k_{ij} = h(h_{ij}||t_0) \tag{7}$$

where h_{ij} will be used for one-time authentication in time slot T_j and k_{ij} will be used for encryption in time slot T_j. The TA also makes a signature σ on all hash chains' heads $(h_{10}, h_{20}, \cdots, h_{N0})$ so as to ensure all hash chains are indeed valid for authentication. Finally, the TA also chooses AES as the encryption algorithm in the system and sets $params = \{n, q_i : i = 1, 2, \cdots k, \alpha_j : j = 0, 1, \cdots k, h, H, L(x), \text{AES}\}$ as the system public parameters.

After the above parameter settings, the TA will assign the key materials to all entities. Specifically,

- For each IoT device $D_i \in \mathcal{D}$, the TA will assign the secret key s_i, the secret hash chain $\mathcal{HC}_i = \{h_{i0}, h_{i1}, h_{i2}, \cdots, h_{iw}\}$, the corresponding keys $\mathcal{K}_i = \{k_{i1}, k_{i2}, \cdots, k_{iw}\}$, and the public parameters $params$ to D_i via a secure channel.
- For the fog device, the TA will first choose a random number sk as the shared key between the fog device and the control center, and assign the shared key sk, the signed hash chain heads $(h_{10}, h_{20}, \cdots, h_{N0})$ and σ, together with the secret keys (s_{N+1}, t_0) and the public parameters $params$ to the fog device.
- For the control center, the TA will assign the same shared key sk and the secret keys (s_0, λ), together with the public parameters $params$ to the control center.

B. IoT DEVICE REPORT GENERATION

At every time slot T_s, each IoT device D_i will report its sensing data x_i by running the following steps:

- Step 1: If the IoT device D_i belongs to the subset $\mathcal{D}_j$, D_i uses its secret key s_i and (α_0, α_j) to compute

$$c_{is} = [1 + n \cdot \alpha_j \cdot (x_i \cdot \alpha_0 + x_i^2)] \cdot H(T_s)^{n \cdot s_i} \bmod n^2 \tag{8}$$

  and uses the key k_{is} to compute $C_{is} = \text{AES}_{k_{is}}(c_{is})$, which can avoid the control center directly obtain c_{is} to get D_i's individual data.
- Step 2: D_i uses the hash value h_{is} in its hash chain $\mathcal{HC}_i$ to compute

$$mac_{is} = h(C_{is}||h_{is}) \tag{9}$$

- Step 3: D_i forwards $(C_{is}, h_{is}, mac_{is})$ to the fog device.

Note that, the IoT device D_i can very efficiently run the above steps. In particular, when $H(T_s)^{n \cdot s_i}$ is pre-computed in advance, then only fast multiplication operations and AES encryption are required.

C. FOG DEVICE REPORT AGGREGATION

After receiving $(C_{is}, h_{is}, mac_{is})$ in time slot T_s, the fog device runs the following steps to check its validity.

- Step 1: As the fog device holds the authenticated h_{i0} from σ, it is easy to verify the validity of each h_{ij} on the chain $\mathcal{HC}_i$. For example, once $h_{i(s-1)}$ has been authenticated in the previous time slot T_{s-1}, the fog device can verify h_{is} by checking $h_{i(s-1)} \stackrel{?}{=} h(h_{is}||T_s)$. If it does hold and h_{is} has not been received previously, h_{is} is accepted, otherwise rejected.
- Step 2: Once h_{is} is valid, the fog device can verify C_{is} by computing $mac'_{is} = h(C_{is}||h_{is})$ and checking whether $mac'_{is} \stackrel{?}{=} mac_{is}$. If it does hold, C_{is} will be accepted, otherwise C_{is} will be filtered out.
- Step 3: Once C_{is} is accepted, the fog device computes $k_{is} = h(h_{is}||t_0)$ and uses k_{is} to recover c_{is} from $C_{is} = \text{AES}_{k_{is}}(c_{is})$.

After receiving all $(c_{1s}, c_{2s}, \cdots, c_{Ns})$ from all IoT devices $\mathcal{D}$ in time slot T_s, the fog device uses the secret key s_{N+1} to compute $H(T_s)^{n \cdot s_{N+1}}$ and runs the following data aggregation operations

$$\begin{cases} C_s = \left(\prod_{i=1}^{N} c_{is}\right) \cdot H(T_s)^{n \cdot s_{N+1}} \bmod n^2 \\ mac_s = h(C_s||T_s||sk) \end{cases} \tag{10}$$

and forwards (C_s, mac_s) to the control center.

D. CONTROL CENTER REPORT READING AND ANALYTICS

Upon receiving (C_s, mac_s) in time slot T_s, the control center first verifies C_s by checking $mac_s = h(C_s||T_s||sk)$. If C_s is valid, the control center runs the following steps for report reading and analytics.

- Step 1: The control center uses the secret key s_0 to compute $H(T_s)^{n \cdot s_0}$. Note that, $H(T_s)^{n \cdot s_0}$ can also be pre-computed, and then the report reading can be accelerated.

- Step 2: The control center computes

$$\begin{aligned}
C_s' &= C_s \cdot H(T_s)^{n\cdot s_0} \bmod n^2 \\
&= \prod_{i=1}^{N} c_{is} \cdot H(T_s)^{n\cdot(s_0+s_{N+1})} \bmod n^2 \\
&\xrightarrow{\text{where } \alpha_j^* \text{ is one element in set } \{\alpha_1, \alpha_2, \cdots \alpha_k\}} \\
&= \left(\prod_{i=1}^{N}[1+n\cdot\alpha_j|_{j\in\{1,2,\cdots,k\}}\cdot(x_i\cdot\alpha_0+x_i^2)]\cdot H(T_s)^{n\cdot s_i}\right) \\
&\quad \times H(T_s)^{n\cdot(s_0+s_{N+1})} \bmod n^2 \\
&= \prod_{i=1}^{N}[1+n\cdot\alpha_j^*\cdot(x_i\cdot\alpha_0+x_i^2)]\cdot\prod_{i=0}^{N+1}H(T_s)^{n\cdot s_i} \bmod n^2 \\
&= \prod_{i=1}^{N}[1+n\cdot\alpha_j^*\cdot(x_i\cdot\alpha_0+x_i^2)]\cdot H(T_s)^{n\cdot\sum_{i=0}^{N+1}s_i} \bmod n^2 \\
&\xrightarrow{\sum_{i=0}^{N+1}s_i\equiv 0 \bmod \lambda \Rightarrow \sum_{i=0}^{N+1}s_i=\kappa\cdot\lambda \text{ for some integer } \kappa} \\
&= \prod_{i=1}^{N}[1+n\cdot\alpha_j^*\cdot(x_i\cdot\alpha_0+x_i^2)]\cdot H(T_s)^{n\cdot\lambda\cdot\kappa} \bmod n^2 \\
&\xrightarrow{x^{n\lambda}\equiv 1 \bmod n^2 \Rightarrow H(T_s)^{n\cdot\lambda\cdot\kappa}\equiv 1 \bmod n^2} \\
&= \prod_{i=1}^{N}[1+n\cdot\alpha_j^*\cdot(x_i\cdot\alpha_0+x_i^2)] \bmod n^2 \\
&\xrightarrow{\text{from Eq. (1)}} \\
&= 1+n\cdot\sum_{i=1}^{N}\alpha_j^*\cdot(x_i\cdot\alpha_0+x_i^2) \bmod n^2 \\
&\xrightarrow{\text{because we consider there are } k \text{ subsets of IoT devices}} \\
&= 1+n\cdot\sum_{j=1}^{k}\alpha_j\left(\sum_{D_i\in\mathcal{D}_j}(x_i\cdot\alpha_0+x_i^2)\right) \bmod n^2 \qquad (11)
\end{aligned}$$

- Step 3: Because we set the conditions $x_i < X$ and $N \cdot (X^2 + X \cdot \alpha_0) < q_i$, we will have

$$\begin{aligned}
&\sum_{i=1}^{N_j}(x_i\cdot\alpha_0+x_i^2) \le \sum_{i=1}^{N_j}(X\cdot\alpha_0+X^2) \\
&< \sum_{i=1}^{N}(X\cdot\alpha_0+X^2) = N(X\cdot\alpha_0+X^2) = a_j < q_j \qquad (12)
\end{aligned}$$

and then from $Q = q_1 \times q_2 \times \cdots \times q_k$, $|q_i| = k_1$, $k_1 \cdot (k+1) + \lg k < |n|$, we have

$$\begin{aligned}
\sum_{j=1}^{k}\alpha_j\left(\sum_{i=1}^{N_j}(x_i\cdot\alpha_0+x_i^2)\right) &< \sum_{j=1}^{k}\alpha_j\cdot a_j \\
&< k\cdot 2^{k_1\cdot(k+1)} < n \qquad (13)
\end{aligned}$$

Therefore, in step 3, the control center computes

$$\sum_{j=1}^{k}\alpha_j\left(\sum_{i=1}^{N_j}(x_i\cdot\alpha_0+x_i^2)\right) = L(C_s') = \frac{C_s'-1}{n} \qquad (14)$$

and

$$M = \sum_{j=1}^{k}\alpha_j\left(\sum_{i=1}^{N_j}(x_i\cdot\alpha_0+x_i^2)\right) \bmod Q \qquad (15)$$

- Step 4: From the Chinese Remainder Theorem and the condition $N \cdot X^2 \le \alpha_0$, for each subset $\mathcal{D}_j \in \mathcal{D}$, the control center can compute its mean and variance as follows

$$M_j = M \bmod q_j = \sum_{i=1}^{N_j}(x_i\cdot\alpha_0+x_i^2) \qquad (16)$$

$$E(\mathcal{D}_j) = \frac{M_j - (M_j \bmod \alpha_0)}{\alpha_0\cdot N_j} \qquad (17)$$

$$Var(\mathcal{D}_j) = \frac{M_j \bmod \alpha_0}{N_j} - E(\mathcal{D}_j)^2 \qquad (18)$$

FAULT TOLERANCE

In case that one IoT device D_a in subset $\mathcal{D}_b$ is malfunctioning, D_a stops reporting to the control center. Then, after aggregating other devices' data into C_s', the fog device reports "D_a is malfunctioning" to the control center, together with C_s^*, which does not include D_a's data. Because the condition $\prod_{i=0}^{N+1} H(T_s)^{n\cdot s_i} \equiv 1 \bmod n^2$ does not hold, the control center uses the secret key λ to run the following steps:

- Step 1: Because C_s^* is in the form of

$$\begin{aligned}
C_s^* = &\left(1+n\cdot\sum_{i=1,i\neq a}^{N}\alpha_j^*\cdot(x_i\cdot\alpha_0+x_i^2)\right) \\
&\cdot\prod_{i=1,i\neq a}^{N+1}H(T_s)^{n\cdot s_i} \bmod n^2 \qquad (19)
\end{aligned}$$

The control center computes

$$\begin{aligned}
M_s^* &= C_s^{*\lambda} \bmod n^2 \\
&\xrightarrow{(1+n\cdot x)^\lambda\equiv(1+n\cdot\lambda x) \bmod n^2,\quad x^{\lambda n}\equiv 1 \bmod n^2} \\
&= 1+n\cdot\lambda\cdot\sum_{i=1,i\neq a}^{N}\alpha_j^*\cdot(x_i\cdot\alpha_0+x_i^2) \bmod n^2 \qquad (20)
\end{aligned}$$

and obtains M by computing

$$M = \left(\frac{M_s^*-1}{n\cdot\lambda} \bmod n\right) \bmod Q \qquad (21)$$

- Step 2: For each subset $\mathcal{D}_j \subset \mathcal{D}$, except the subset $\mathcal{D}_b$, the control center uses Eqs.(16)-(18) to compute its mean $E(\mathcal{D}_j)$ and variance $Var(\mathcal{D}_j)$.
- Step 3: For the subset $\mathcal{D}_b$, the control center first computes M_b by Eq. (16), and gains the mean $E(\mathcal{D}_b)$ and variance $Var(\mathcal{D}_b)$ by

$$E(\mathcal{D}_b) = \frac{M_b - (M_b \bmod \alpha_0)}{\alpha_0\cdot(N_b-1)} \qquad (22)$$

$$Var(\mathcal{D}_b) = \frac{M_b \bmod \alpha_0}{N_b-1} - E(\mathcal{D}_b)^2 \qquad (23)$$

Therefore, even though some IoT devices are malfunctioning, the proposed LPDA scheme is still workable, i.e., the control center can still make data analytics from the rest normal IoT devices. As a result, the proposed LPDA scheme achieves *fault-tolerant*.

V. SECURITY ANALYSIS

In this section, we discuss security properties of the proposed LPDA scheme. In particular, following our design goal, we will focus on analyzing how the proposed LPDA scheme is secure against the false data injection attack from external attacks, and achieve the privacy-preserving data aggregation.

• *The proposed LPDA scheme can resist against the false data injection from the external attacks.* In order to authenticate the source of data at each time slot, we apply the one-way hash chain technique. For each IoT device D_i, the hash value $h_{i(s-1)}$ was released in time slot T_{s-1}. From $h_{i(s-1)} = h(h_{i(s)}||T_s)$, we can authenticate $h_{i(s)}$ from time slot T_s, as it is hard to get $h_{i(s)}$ from $h_{i(s-1)}$ due to the one-wayness of the hash function. In addition, as each D_i can directly communicate with the fog device, only if D_i reports its data in time slot T_s, the fog device can always receive a fresh h_{is}. If h_{is} is not fresh in time slot T_s, it may be a false data injection attack launched by an external attacker by replaying h_{is}, and the fog device can identify it and will reject the false data. Therefore, the proposed LPDA scheme can resist against the false data injection from the external attacks.

• *The proposed LPDA scheme is privacy preserving.* In the proposed LDPA scheme, for each individual ciphertext $c_{is} = [1 + n \cdot \alpha_j \cdot (x_i \cdot \alpha_0 + x_i^2)] \cdot H(T_s)^{n \cdot s_i} \bmod n^2$ and the aggregated ciphertext $C_s = \left(1 + n \cdot \sum_{i=1}^{N} \alpha_j^* \cdot (x_i \cdot \alpha_0 + x_i^2)\right) \cdot \prod_{i=1}^{N+1} H(T_s)^{n \cdot s_i} \bmod n^2$, if we look the item $\alpha_j \cdot (x_i \cdot \alpha_0 + x_i^2)$ as a message $\overline{m}_i$, the item $H(T_s)^{\cdot s_i}$ as a random number r_i, the item $\sum_{i=1}^{N} \alpha_j^* \cdot (x_i \cdot \alpha_0 + x_i^2)$ as a message $\overline{M}$, and the item $\prod_{i=1}^{N+1} H(T_s)^{n \cdot s_i}$ as a random number R, then both $c_{is} = (1+n \cdot \overline{m}_i) \cdot r_i^n \bmod n^2$ and $C_s = (1+n \cdot \overline{M}) \cdot R^n \bmod n^2$ are valid Paillier ciphertexts. Because Pailler encryption is IND-CPA (indistinguishable under the chosen plaintext attack) secure, an external attacker cannot read $\overline{m}_i$ and $\overline{M}$. For the fog device, it may be curious about $\overline{m}_i$ and $\overline{M}$. However, similar as the external attacker, without knowing the secret keys s_0 and λ, it has no idea to recover them. For the control center, it has the ability to recover $\overline{M}$, and it may be curious about each individual $\overline{m}_i$. Because $c_{is} = (1 + n \cdot \overline{m}_i) \cdot r_i^n \bmod n^2$ is encrypted with $C_{is} = \mathrm{AES}_{k_{is}}(c_{is})$, only if there is a collusion between the fog device and the control center, the control center can recover $\overline{m}_i$. However, under our defined security model, the control center cannot collude with the fog device and subsequently cannot recover $\overline{m}_i$. From the above analysis, our proposed LPDA scheme can achieve privacy-preserving in our defined security model.

ENHANCED PRIVACY WITH DIFFERENTIAL PRIVACY TECHNIQUE

The above analysis shows that the proposed LPDA can achieve privacy-preserving when there is no malfunctioning IoT devices. However, once there is one IoT device D_a in subset $\mathcal{D}_b$ is malfunctioning, it is possible for the control center to use the differential attack to gain D_a's data. For example, in time slot T_{s-1}, the device D_a in $\mathcal{D}_b$ did not report its data but other devices in $\mathcal{D}_b$ normally reported their data, and thus the control center can get the aggregated data $A_{1(s-1)} = \sum_{D_i \in \mathcal{D}_b/\{D_a\}} x_i$; while in time slot T_s, D_a is recovered and also reports its data x_a, then the control center can get aggregated data $A_{1s} = A'_{1s} + x_a = \sum_{D_i \in \mathcal{D}_b/\{D_a\}} x_i + x_a = \sum_{D_i \in \mathcal{D}_b} x_i$. If other devices' reports are stable in time slots T_{s-1} and T_s, there may exist some correlation between $A_{1(s-1)} = \sum_{D_i \in \mathcal{D}_b/\{D_a\}} x_i$ and $A'_{1s} = \sum_{D_i \in \mathcal{D}_b/\{D_a\}} x_i$, e.g., $A_{1(s-1)} = A'_{1s}$. Then, in this case, the control center can gain the IoT device D_a's data x_a in time slot T_s by computing $x_a = A_{1s} - A_{1(s-1)}$. In order to avoid this kind of privacy disclosure, we can use differential privacy technique to enhance privacy.

DIFFERENTIAL PRIVACY TECHNIQUE

Since the seminal work was introduced in 2006 [29], differential privacy techniques has received considerable attention in privacy-preserving data statistics. The core idea of differential privacy technique is to add some reasonable noises, e.g., noises extracted from symmetrical geometric distribution, Laplace distribution, etc., to make the outputs from similar inputs indistinguishable. Formally, we call a randomized algorithm $A(.)$ can achieve ε-differential privacy, if for any two data sets DS_1 and DS_2 differing on a single element, for every subset $S \subseteq \mathit{OutputRange}(A)$, $\Pr[A(DS_1) \in S] \leq \exp(\varepsilon) \cdot \Pr[A(DS_2) \in S]$ holds. As the aggregated data are discrete in the proposed LPDA scheme, we consider noises extracted from geometric distribution. The noises generation from geometric distribution was introduced by Ghosh *et al.* [31], where the noise is chosen from a symmetric geometric distribution $\mathit{Geom}(\alpha)$ with $0 < \alpha < 1$. Then, the $\mathit{Geom}(\alpha)$ can be viewed as a discrete approximation of Laplace distribution $\mathit{Lap}(\lambda)$, where $\alpha \approx \exp(-\frac{1}{\lambda})$. The probability density function (PDF) of geometric distribution $\mathit{Geom}(\alpha)$ is

$$\Pr[X = x] = \frac{1-\alpha}{1+\alpha} \cdot \alpha^{|x|} \tag{24}$$

When the sensitivity of the algorithm $A(DS)$ is $\Delta A = \max_{DS_1, DS_2} ||A(DS_1) - A(DS_2)||_1$ for all the data sets DS_1 and DS_2 differing in at most one element, then by adding geometric noise r randomly chosen from $\mathit{Geom}(\exp(-\frac{\varepsilon}{\Delta A}))$ to the original aggregated data, the perturbed results can achieve ε-differential privacy, i.e., for any integer $k \in \mathit{OutputRange}(A)$, $\Pr[A(DS_1) + r = k] \leq \exp(\varepsilon) \cdot \Pr[A(DS_2) + r = k]$.

Assume that each subset $\mathcal{D}_j$ has at most one malfunctioning device at each time slot. In order to avoid the differential attack from the control center, no matter there is a malfunctioning IoT device in $\mathcal{D}_j$, the fog device will run the following steps:

- Step 1: Let $\mathcal{C}$ be the ciphertext set received from all normal IoT devices in time slot T_s. Due to the existence of some malfunctioning devices, the size of $\mathcal{C}$

is $|\mathcal{C}| \leq N$. For each subset $\mathcal{D}_j$, the aggregated data $\sum_{D_i \in \mathcal{D}_j} x_i$ and $\sum_{D_i \in \mathcal{D}_j} x_i^2$ both can be recovered by the control center later. Let $A_1(\mathcal{D}_j) = \sum_{D_i \in \mathcal{D}_j} x_i$ and $A_2(\mathcal{D}_j) = \sum_{D_i \in \mathcal{D}_j} x_i^2$, then $|A_1(\mathcal{D}_{j1}) - A_1(\mathcal{D}_{j2})| \leq X$, and $|A_2(\mathcal{D}_{j1}) - A_2(\mathcal{D}_{j2})| \leq X^2$. Therefore, we can set $\Delta A_1 = X$ and $\Delta A_2 = X^2$. In this step, the fog device chooses the random noises $\mathbf{x}_{j1}$ from $Geom(\exp(-\frac{\varepsilon}{\Delta A_1}))$ and $\mathbf{x}_{j2}$ from $Geom(\exp(-\frac{\varepsilon}{\Delta A_2}))$ to implicitly add them to $A_1(\mathcal{D}_j)$ and $A_2(\mathcal{D}_j)$ by computing

$$C_s^* = \left(\prod_{c_i \in \mathcal{C}} c_i\right) \cdot \prod_{j=1}^{k} \left(1 + n \cdot \alpha_j (\mathbf{x}_{j1} \cdot \alpha_0 + \mathbf{x}_{j2})\right) \bmod n^2$$

- Step 2: The fog device forwards $(C_s^*, mac_s = h(C_s||T_s||sk))$ to the control center.

Because C_s^* does not include $H(T_s)^{n \cdot s_{N+1}}$, the condition $\prod_{i=0}^{N+1} H(T_s)^{n \cdot s_i} \equiv 1 \bmod n^2$ does not hold, the control center can only use the similar methods in Eqs. (20) and (16)-(18) to obtain $E(\mathcal{D}_j)$ and $Var(\mathcal{D}_j)$, where

$$E(\mathcal{D}_j) = \frac{\sum_{D_i \in \mathcal{D}_j} x_i + \mathbf{x}_{j1}}{N_j} \tag{25}$$

$$Var(\mathcal{D}_j) = \frac{\sum_{D_i \in \mathcal{D}_j} x_i^2 + \mathbf{x}_{j2}}{N_j} - E(\mathcal{D}_j)^2 \tag{26}$$

With the above differential privacy technique, we can show both $\sum_{D_i \in \mathcal{D}_j} x_i + \mathbf{x}_{j1}$ and $\sum_{D_i \in \mathcal{D}_j} x_i^2 + \mathbf{x}_{j2}$ can achieve ε-differential privacy. For example, for $\sum_{D_i \in \mathcal{D}_j} x_i + \mathbf{x}_{j1}$ aggregation, assume the control center obtains two perturbed aggregation $u + \mathbf{x}_{j1}^{(u)}$ and $v + \mathbf{x}_{j1}^{(v)}$, where u and v are two adjacent aggregation while $\mathbf{x}_{j1}^{(u)}$ and $\mathbf{x}_{j1}^{(v)}$ are the corresponding geometric noises from $Geom(\exp(-\frac{\varepsilon}{\Delta A_1}))$. Since $|u - v| \leq X$, for any integer k, we will have

$$\begin{aligned} \eta &= \frac{\Pr[u + \mathbf{x}_{j1}^{(u)} = k]}{\Pr[v + \mathbf{x}_{j1}^{(v)} = k]} = \frac{\Pr[\mathbf{x}_{j1}^{(u)} = k - u]}{\Pr[\mathbf{x}_{j1}^{(v)} = k - v]} \\ &= \frac{\frac{1-\alpha}{1+\alpha} \cdot \alpha^{|k-u|}}{\frac{1-\alpha}{1+\alpha} \cdot \alpha^{|k-v|}} = \alpha^{|k-u|-|k-v|} \end{aligned} \tag{27}$$

Because $-|u-v| \leq |k-u| - |k-v| \leq |u-v|$ and $0 < \alpha < 1$,

$$\begin{aligned} \alpha^X &\leq \alpha^{|u-v|} \leq \eta \leq \alpha^{-|u-v|} \leq \alpha^{-X} \\ &\xrightarrow{\text{since } \alpha \approx \exp(-\frac{\varepsilon}{X})} \\ &\Rightarrow (e^{-\frac{\varepsilon}{X}})^X \leq \eta \leq (e^{-\frac{\varepsilon}{X}})^{-X} \\ &\Rightarrow e^{-\varepsilon} \leq \eta \leq e^{\varepsilon} \end{aligned} \tag{28}$$

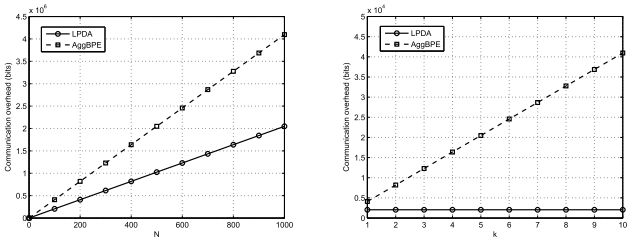
Therefore, $\sum_{D_i \in \mathcal{D}_j} x_i + \mathbf{x}_{j1}$ achieves ε-differential privacy. Similarly, we can also prove $\sum_{D_i \in \mathcal{D}_j} x_i^2 + \mathbf{x}_{j2}$ achieves ε-differential privacy. From the above analysis, we can conclude that the differential attack from the control center can be avoided, and the privacy is enhanced in the proposed LPDA scheme.

VI. PERFORMANCE EVALUATION

In this section, we evaluate our proposed LPDA scheme in terms of the communication overhead and computational costs at the IoT devices, the fog device, and the control center.

A. COMMUNICATION OVERHEAD

The proposed LPDA scheme achieves the privacy-preserving aggregation for hybrid IoT, i.e., it can aggregate different subsets of IoT devices' data into one, and the control center can recover each subset's *mean* and *variance*. In order to show the efficiency of LPDA, we compare LPDA with the aggregation with the basic Paillier encryption (AggBPE), i..e, the ciphertext is in the form of $c = g^m r^n \bmod n^2$, under the same parameter settings, i.e., the bit length of n^2 is $|n^2| = 2048$, i.e., $|n| = 1024$. For the fairness of comparison, we do not consider the authentication cost in both schemes. Then, for the communication from N IoT devices to the fog device, the communication overhead is $2048 \times N$ bits, because each device D_i encrypts x_i and x_i^2 into one ciphertext $c_{is} = [1 + n \cdot \alpha_j \cdot (x_i \cdot \alpha_0 + x_i^2)] \cdot H(T_s)^{n \cdot s_i} \bmod n^2$. However, if we use the AggBPE to aggregate IoT devices' data, D_i will encrypt x_i and x_i^2 into two ciphertexts. Then the communication costs become double, and the total communication overhead is $4096 \times N$ bits. For the communication overhead from the fog device to the control center, both of them are independent from the the number of IoT devices, because data have already been aggregated at the fog device. However, the communication cost of AggBPE is dependent on the number of subsets in $\mathcal{D}$. Assume there are k subsets in $\mathcal{D}$, then the communication overhead is $4096 \times k$ bits. While in our LPDA scheme, as all data in all subsets have been aggregated into one ciphertext, the communication overhead from the fog device to the control center is only 2048 bits. Fig. 5 plots the communication overheads from the IoT devices to the fog device and the overhead from the fog device to the control center. From the figures, it clearly indicates our proposed LPDA is much more efficient than AggBPE.



(a) (b)

FIGURE 5. Communication comparisons between LPDA and AggBPE. (a) Communication overhead from IoT devices to fog device with N varying from 1 to 1000. (b) Communication overhead from Fog device to control center with k varying from 1 to 10.

Note that, different from AggBPE, the proposed LPDA uses the Chinese Remainder Theorem to aggregate all data into one, which requires us to carefully choose the proper parameters. Luckily, because the message space is small in many real application scenarios, it is not difficult for us

to choose the proper parameters meeting the conditions in Eqs. (4) and (5). For example, if the message space is $[0, X = 2^8]$, the number of IoT devices is $N = 2^{10}$, we can choose α_0 with length $|\alpha_0| = 30$ to satisfy the condition $N \cdot X^2 \leq \alpha_0$. Further, we can choose each q_i with length $|q_i| = k_1 = 50$, then the condition $N \cdot (X^2 + X \cdot \alpha_0) < q_i$ is also satisfied. From the condition $k_1 \cdot (k + 1) + \lg k < |n|$, we can choose the maximal $k = 19$, which means the proposed LPDA can achieve at most 19 subsets' aggregation with the above parameter settings.

B. COMPUTATIONAL COSTS

In terms of computational costs, the proposed LPDA is lightweight, because if we only take the time-consuming module exponent operations into consideration in LPDA, there are at most 1 module exponent at each IoT device, the fog device, and the control center. When we further consider the module exponent operations can be pre-computed, LPDA will become more efficient. In this subsection, we check the computational costs of LPDA by implementing it with Java (JDK 1.8) and run our experiments on a Laptop with Intel i5-6300H 2.3 GHz processor, 12GB RAM, and Window 10 platform. The detailed parameter settings are shown in Table 1.

TABLE 1. The parameter settings.

Parameter	Value
k_0, k_1, l	$k_0 = 512$, $k_1 = 50$, $l = 160$
p, q	$\|p\| = \|q\| = k_0 = 512$
$n = pq$	$\|n\| = 2k_0 = 1024$, $\|n^2\| = 4k_0 = 2048$
q_i	$\|q_i\| = k_1 = 50$
N, k	$N = 1000$, $k = 10$: 1000 IoT devices in 10 subsets
N_j	$N_j = 100$: the size of each subset $\mathcal{D}_j$ is 100
α_0	$\|\alpha_0\| = 30$: the size of the parameter α_0
X	$X = 2^8$: the message space is $[0, 2^8]$
ε	$\varepsilon = 1$: the privacy parameter set in differential privacy
$\mathbf{x}_{j1}$	$\mathbf{x}_{j1} \in Geom(\exp(-\frac{\varepsilon}{X}))$: the 1st noise added in $\mathcal{D}_j$'s aggregation, i.e., $\sum_{D_i \in \mathcal{D}_j} x_i + \mathbf{x}_{j1}$
$\mathbf{x}_{j2}$	$\mathbf{x}_{j2} \in Geom(\exp(-\frac{\varepsilon}{X^2}))$: the 2nd noise added in $\mathcal{D}_j$'s aggregation, i.e., $\sum_{D_i \in \mathcal{D}_j} x_i^2 + \mathbf{x}_{j2}$

TABLE 2. The parameter settings.

Entities	Computational costs	
	without malfunctioning device and no differential privacy enhancement	enhanced with differential privacy
Each IoT device	0.328 ms	
Fog device	0.470 ms	0.578 ms
Control center	0.062 ms	0.156 ms

We run our experiments 1000 times, and the average running time values are recorded in Table 2. From the table, we can see our proposed LPDA is really efficient in terms of computational costs. Even when LPDA is privacy-enhanced with differential privacy techniques, it is still much efficient.

VII. RELATED WORK

In this section, we review some privacy-preserving data aggregation schemes [1], [6], [14]–[27], which are also fit for fog computing-enhanced IoT and closely related to our proposed LPDA.

In 2011, Shi *et al.* [14] proposed an efficient privacy-preserving data aggregation scheme for time-series data. In their scheme, the secret keys $(s_1, s_2, \cdots, s_n)$ of all participants and the secret key s_0 of the aggregator satisfy the condition $\sum_{i=0}^{n} s_i = 0 \bmod p$, and each ciphertext c_i is in the form of $c_i = g^{x_i} H(t)^{s_i}$. Then, only when all ciphertexts are aggregated, the aggregator can use $H(t)^{s_0}$ to make $H(t)^{\sum_{i=0}^{n} s_i} = 1$, and the aggregated result $\sum_{i=1}^{n} x_i$ can be recovered. However, in case there is a malfunctioning participant, the aggregated result cannot be correctly recovered, i.e., their scheme is not fault-tolerant. In the same year, Ruj *et al.* [15] used the original homomorphic Paillier encryption [28], i.e., the ciphertext is in the form of $c = g^m r^n \bmod n^2$, to design a privacy-preserving aggregation scheme for smart grid. Without the restriction $\sum_{i=0}^{n} s_i = 0 \bmod p$, their scheme is fault-tolerant, but each individual ciphertext could be directly recovered by the control center. In 2012, Alharbi and Lin [16] used the additive homomorphic encryption scheme proposed by Castelluccia *et al.* [32] to present a privacy-preserving data aggregation scheme for smart grid, their scheme is ligh-weight, but the fault tolerance is not considered. Aiming at solving the fault tolerance issue, Chan *et al.* [17] presented novel mechanisms to enhance Shi *et al.*'s scheme [14]. With the new mechanisms, the enhanced scheme is resilient to user failure and compromise, and can efficiently support dynamic joins and leaves. In order to support privacy-preserving multidimensional data aggregation, Lu *et al.* [6] presented EPPA by combining the Paillier encryption and the superincreasing sequence techniques, which can further reduce the communication overhead. Based on Boneh-Goh-Nissim (BGN) homomorphic encryption [33], Chen *et al.* [18] presented multifunctional data aggregation in privacy-preserving smart grid communications, which can support *average*, *variance*, and *one-way ANOVA* aggregation. Li *et al.* [19] also used the lattice cryptographic technique [34] to present a privacy-preserving dual-functional aggregation scheme for smart grid, which can support *mean* and *variance* aggregation at the same time. Later, by taking the requirements of fault tolerance, data integrity, and differential privacy into consideration, other novel privacy-preserving data aggregation schemes have been proposed [1], [20]–[27]. Although the above schemes are promising and can be fit for fog computing-enhanced IoT. However, for hybrid IoT, they cannot aggregate all IoT devices' data into one ciphertext.

Different from the above schemes, our proposed LPDA scheme uses the modified Paillier encryption, i.e., the ciphertext is in the form of $c = (1 + n \cdot m) r^n \bmod n^2$, the Chinese Remainder Theorem, the one-way hash chain, and the differential privacy techniques, to enable the fog device to aggregate hybrid IoT devices' data into one ciphertext

and *early* filter the injected false data at the network edge. In addition, fault tolerance and efficiency can also be well achieved. Note that, EPPA [6] may be applied for hybrid IoT data aggregation, but the multiple time-consuming exponential computations make it not as efficient as the proposed LPDA.

VIII. CONCLUSION

In this paper, we have proposed a lightweight privacy-preserving data aggregation scheme, called LPDA, for fog computing-enhanced IoT. With the fog device deployed at the network edge, LPDA can not only *early* filter false data injected by external attackers, but also support fault-tolerance and efficiently aggregate hybrid IoT devices' data into one. Detailed security analyses, especially the enhanced differential privacy analyses, show the proposed LPDA scheme is secure under our defined security model. In addition, extensive performance analyses and experiments are conducted, and the results indicate it is really lightweight in both communication overheads and computational costs. In future work, we will evaluate our proposed scheme in some realistic IoT scenarios, consider stronger adversarial model, and design new solutions under new model.

REFERENCES

[1] R. Lu, *Privacy-Enhancing Aggregation Techniques for Smart Grid Communications* (Wireless Networks). Switzerland: Springer, 2016. [Online]. Available: http://dx.doi.org/10.1007/978-3-319-32899-7

[2] X. Lin, R. Lu, X. Shen, Y. Nemoto, and N. Kato, "Sage: A strong privacy-preserving scheme against global eavesdropping for ehealth systems," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 4, pp. 365–378, May 2009.

[3] M. Khan, B. N. Silva, and K. Han, "Internet of Things based energy aware smart home control system," *IEEE Access*, vol. 4, pp. 7556–7566, 2016.

[4] M. Pouryazdan, B. Kantarci, T. Soyata, and H. Song, "Anchor-assisted and vote-based trustworthiness assurance in smart city crowdsensing," *IEEE Access*, vol. 4, pp. 529–541, 2016.

[5] S. L. Hoe, "Defining a smart nation: The case of singapore," *J. Inf., Comm, Ethics Soc.*, vol. 14, no. 4, pp. 323–333, 2016.

[6] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, "EPPA: An efficient and privacy-preserving aggregation scheme for secure smart grid communications," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 9, pp. 1621–1631, Sep. 2012.

[7] K.-Y. Lam and C.-H. Chi, "Identity in the Internet-of-Things (IoT): New challenges and opportunities," in *Proc. 18th Int. Conf. Inf. Commun. Secur. (ICICS)*, Singapore, Nov./Dec. 2016, pp. 18–26.

[8] M. M. Rathore, A. Paul, A. Ahmad, and G. Jeon, "IoT-based big data: From smart city towards next generation super city planning," *Int. J. Semantic Web Inf. Syst.*, vol. 13, no. 1, pp. 28–47, 2017.

[9] R. Lu, X. Lin, H. Zhu, X. Liang, and X. S. Shen, "BECAN: A bandwidth-efficient cooperative authentication scheme for filtering injected false data in wireless sensor networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 23, no. 1, pp. 32–43, Jan. 2012.

[10] F. Bonomi, R. A. Milito, J. Zhu, and S. Addepalli, "Fog computing and its role in the Internet of Things," in *Proc. 1st Ed. MCC Workshop Mobile Cloud Comput.*, Helsinki, Finland, Aug. 2012, pp. 13–16.

[11] A. V. Dastjerdi and R. Buyya, "Fog computing: Helping the Internet of Things realize its potential," *IEEE Comput.*, vol. 49, no. 8, pp. 112–116, Aug. 2016.

[12] R. Deng, R. Lu, C. Lai, and T. H. Luan, "Towards power consumption-delay tradeoff by workload allocation in cloud-fog computing," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 3909–3914.

[13] H. Freeman and T. Zhang, "The emerging era of fog computing and networking [the president's page]," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 4–5, Jun. 2016.

[14] E. Shi, T. H. Chan, E. G. Rieffel, R. Chow, and D. Song, "Privacy-preserving aggregation of time-series data," in *Proc. Netw. Distrib. Syst. Secur. Symp. (NDSS)*, San Diego, CA, USA, Feb. 2011.

[15] S. Ruj, A. Nayak, and I. Stojmenovic, "A security architecture for data aggregation and access control in smart grids," *CoRR*, 2011. [Online]. Available: http://arxiv.org/abs/1111.2619

[16] K. Alharbi and X. Lin, "LPDA: A lightweight privacy-preserving data aggregation scheme for smart grid," in *Proc. Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Huangshan, China, Oct. 2012, pp. 1–6.

[17] T.-H. H. Chan, E. Shi, and D. Song, "Privacy-preserving stream aggregation with fault tolerance," in *Proc. 16th Int. Conf. Financial Cryptogr. Data Secur.*, Kralendijk, Bonaire, Feb./Mar. 2012, pp. 200–214.

[18] L. Chen, R. Lu, Z. Cao, K. AlHarbi, and X. Lin, "MuDA: Multifunctional data aggregation in privacy-preserving smart grid communications," *Peer-Peer Netw. Appl.*, vol. 8, no. 5, pp. 777–792, Sep. 2015.

[19] C. Li, R. Lu, H. Li, L. Chen, and J. Chen, "PDA: A privacy-preserving dual-functional aggregation scheme for smart grid communications," *Secur. Commun. Netw.*, vol. 8, no. 15, pp. 2494–2506, Oct. 2015.

[20] H. Bao and R. Lu, "A lightweight data aggregation scheme achieving privacy preservation and data integrity with differential privacy and fault tolerance," *Peer-Peer Netw. Appl.*, vol. 10, no. 1, pp. 106–121, Jan. 2017.

[21] H. Bao and R. Lu, "Comment on 'privacy-enhanced data aggregation scheme against internal attackers in smart grid,'" *IEEE Trans. Ind. Informat.*, vol. 12, no. 1, pp. 2–5, Feb. 2016.

[22] H. Bao and R. Lu, "A new differentially private data aggregation with fault tolerance for smart grid communications," *IEEE Internet Things J.*, vol. 2, no. 3, pp. 248–258, Jun. 2015.

[23] L. Chen, R. Lu, and Z. Cao, "PDAFT: A privacy-preserving data aggregation scheme with fault tolerance for smart grid communications," *Peer-Peer Netw. Appl.*, vol. 8, no. 6, pp. 1122–1132, Nov. 2015.

[24] H. Bao and L. Chen, "A lightweight privacy-preserving scheme with data integrity for smart grid communications," *Concurrency Comput., Pract. Exper.*, vol. 28, no. 4, pp. 1094–1110, Mar. 2016.

[25] D. He, N. Kumar, and J.-H. Lee, "Privacy-preserving data aggregation scheme against internal attackers in smart grids," *Wireless Netw.*, vol. 22, no. 2, pp. 491–502, Feb. 2016.

[26] G. Zhuo, Q. Jia, L. Guo, M. Li, and P. Li, "Privacy-preserving verifiable data aggregation and analysis for cloud-assisted mobile crowdsourcing," in *Proc. 35th Annu. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, San Francisco, CA, USA, Apr. 2016, pp. 1–9.

[27] R. Lu, K. Alharbi, X. Lin, and C. Huang, "A novel privacy-preserving set aggregation scheme for smart grid communications," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, San Diego, CA, USA, Dec. 2015, pp. 1–6.

[28] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proc. Int. Conf. Theory Appl. Cryptogr. Tech. Adv. Cryptol. (EUROCRYPT)*, Prague, Czech Republic, May 1999, pp. 223–238.

[29] C. Dwork, "Differential privacy," in *Proc. 33rd Int. Colloq. Autom., Lang. Programm. (ICALP)*, Venice, Italy, Jul. 2006, pp. 1–12.

[30] A. Perrig, R. Canetti, J. D. Tygar, and D. Song, "Efficient authentication and signing of multicast streams over lossy channels," in *Proc. IEEE Symp. Secur. Privacy*, Berkeley, CA, USA, May 2000, pp. 56–73.

[31] A. Ghosh, T. Roughgarden, and M. Sundararajan, "Universally utility-maximizing privacy mechanisms," *SIAM J. Comput.*, vol. 41, no. 6, pp. 1673–1693, 2012.

[32] C. Castelluccia, A. C. Chan, E. Mykletun, and G. Tsudik, "Efficient and provably secure aggregation of encrypted data in wireless sensor networks," *ACM Trans. Sensor Netw.*, vol. 5, no. 3, pp. 20:1–20:36, 2009.

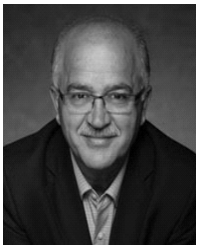
[33] D. Boneh, E.-J. Goh, and K. Nissim, "Evaluating 2-DNF formulas on ciphertexts," in *Proc. 2nd Theory Cryptogr. Conf. Theory Cryptogr. (TCC)*, Cambridge, MA, USA, Feb. 2005, pp. 325–341.

[34] Z. Brakerski and V. Vaikuntanathan, "Fully homomorphic encryption from ring-LWE and security for key dependent messages," in *Proc. 31st Annu. Cryptol. Conf. Adv. Cryptol. (CRYPTO)*, Santa Barbara, CA, USA, Aug. 2011, pp. 505–524.

RONGXING LU (S'09–M'10–SM'15) received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Waterloo, Canada, in 2012. He was a Post-Doctoral Fellow with the University of Waterloo from 2012 to 2013. He was an Assistant Professor with the School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, from 2013 to 2016. He has been an Assistant Professor with the Faculty of Computer Science, University of New Brunswick, Canada, since 2016. His research interests include applied cryptography, privacy enhancing technologies, and IoT-Big Data security and privacy. He currently serves as the Secretary of the IEEE ComSoc CIS-TC. He is currently a Senior Member of the IEEE Communications Society. He received the most prestigious Governor General's Gold Medal and the 8th IEEE Communications Society (ComSoc) Asia Pacific Outstanding Young Researcher Award, in 2013.

KEVIN HEUNG received the master's degree from Peking University in 2002 and the master's degree (Hons.) from the City University of Hong Kong in 2013. He has been with sizeable enterprises as a Developer and a System Analyst. He is currently the Graduate Student with the Faculty of Computer Science, University of New Brunswick, Canada. He has good background in computer programming and statistical analysis, linear and dynamic programming, number theory, and mathematical analytical methods.

ARASH HABIBI LASHKARI (M'10) is currently a Research Associate with the Canadian Institute for Cybersecurity on the Faculty of Computer Science, University of New Brunswick. He has over 21 years of academic and industry experience. In addition, he is the Author of nine books in English and Persian on topics, including cryptography, network security, and mobile communication and over 70 journals and conference papers concerning various aspects of computer security. His current research focuses on cyber security, big security data analysis, Internet traffic analysis, and the detection of malware and attacks. He received three gold medals and 12 silver and bronze medals in international competitions around the world.



ALI A. GHORBANI (SM'–) has held a variety of positions in academia for the past 35 years. He has been the Dean of the Faculty of Computer Science since 2008. He is currently the Canada Research Chair (Tier 1) in Cybersecurity. He is also the Director of the Canadian Institute for Cybersecurity. He has developed a number of technologies that have been adopted by high-tech companies. He co-founded two startups, Sentrant and EyesOver in 2013 and 2015, respectively. He is the Co-Inventor on three awarded patents in the area of network security and web intelligence and has published over 200 peer-reviewed articles during his career. He has supervised over 160 research associates, post-doctoral fellows, and graduate and undergraduate students during his career. His book, *Intrusion Detection and Prevention Systems: Concepts and Techniques*, (Springer, 2010). Since 2010, he has obtained over 10M to fund six large multi-project research initiatives. He was twice one of the three finalists for the Special Recognition Award at the 2013 and 2016 New Brunswick KIRA Award for the knowledge industry. In 2007, he received the University of New Brunswick's Research Scholar Award. He is the Co-Editor-In-Chief of *Computational Intelligence Journal*.