Peer-to-Peer Netw. Appl. (2018) 11:966–978
https://doi.org/10.1007/s12083-017-0618-9

FGDA: Fine-grained data analysis in privacy-preserving smart grid communications

Shanshan Ge[1] · Peng Zeng[1] · Rongxing Lu[2] · Kim-Kwang Raymond Choo[3]

Received: 27 February 2017 / Accepted: 29 October 2017 / Published online: 9 November 2017

Abstract In a smart grid environment, smart meters periodically collect and report information such as electricity consumption of users to a control center for timely monitoring, billing and other analytical purposes. There is, however, a need to ensure the privacy of user data, particularly when the data is combined with data from other sources. In this paper, we propose a new fine-grained data analysis (hereafter referred to as FGDA) scheme for privacy preserving smart grid communications. FGDA is designed to compute multifunctional data analysis (such as average, variance, and skewness) based on users' ciphertexts, as well as supporting fault tolerance feature. We remark that FGDA can still function when some smart meters fail. Compared to existing schemes providing both the properties of multifunction and fault tolerance, FGDA is more efficient in terms of computation overheads. This is because FDGA does not use bilinear map or Pollard's lambda method during decryption. We also demonstrate that FGDA achieves a higher communication efficiency, as the gateway only needs to send the ciphertext to the control center once even for different statistical functions.

Keywords Smart grid security · Smart grid privacy · Privacy-preserving smart grid communications · Finer-grained data analysis

✉ Peng Zeng
pzeng@sei.ecnu.edu.cn

Shanshan Ge
shanshangeecnu@outlook.com

Rongxing Lu
rlu1@unb.ca

Kim-Kwang Raymond Choo
raymond.choo@fulbrightmail.org

[1] Shanghai Key Laboratory of Trustworthy Computing, East China Normal University, Shanghai, China

[2] Faculty of Computer Science, University of New Brunswick, Fredericton, NB, Canada

[3] Department of Information Systems and Cyber Security and Department of Electrical and Computer Engineering, University of Texas at San Antonio, San Antonio, TX 78249, USA

1 Introduction

Due to worldwide interest in energy saving and emission reduction, green power and sustainable development, smart grids have also been the subject of recent research focus [1–3]. This is not surprising due to the potential for smart grids to improve the quality of service to users, as well as offering smart grid operators the opportunity to collect and analyze data to provide better insights into the market (user preference, usage patterns, etc) that can be used to inform other decision making. It is, however, important to ensure the security of the grids to reduce the potential and/or impact of a successful cyber attack or a natural disaster [4, 5].

With rapid advances in intelligent electronic devices (IEDs), such as smart meters (SMs), electric vehicles (EVs) and outdoor smart/Internet-connected "things", it is essential to ensure that these components within a smart grid infrastructure are compliant with international standards and are designed according to best practices [6, 7]. However, how to process the data collected by IEDs in a privacy-preserving way remains one of several research challenges in smart grid research.

Generally, a smart grid consists of a control center (CC), several SMs and several local gateways (GWs). A simple example is shown in Fig. 1, where each house has an SM installed. The installed SM periodically collects information, such as electricity usage of the IEDs in the house, and sends it to a local GW. The latter then aggregates data from all SMs deployed in the same residential area and reports these aggregated data to CC for further analysis and processing (such as balance electricity load and optimize energy consumption). In such a communication flow, if the collected data are leaked to an adversary $\mathcal{A}$ (e.g. by exploiting vulnerabilities in the deployed SMs), then $\mathcal{A}$ can use these data to analyze individual user habits, behaviors, activities and even preferences. For example, a low, or lack of, daily electricity consumption indicates that the house owner may be away, while an extremely high electricity consumption during certain times of the day may help an attacker plan their malicious activities (e.g. in kidnapping for ransom, or to steal). Also, a significant higher than average electricity usage from a particular user may also suggest that the particular address is growing cannabis or marijuana. Undeniably, privacy-preservation of user data is a topic of interest for smart grid operators as well as users.

In order to protect user's privacy, a number of public-key cryptosystems have been employed for smart grids (see [8–17]). One of the most popular systems is the Paillier scheme [18], whose additive homomorphic property enables a GW to compute addition operations on the encrypted data. However, Paillier scheme uses the same decryption key for the original data and the aggregated data; thus, anyone who obtains the decryption key can decrypt not only the final aggregated result but also the individual ciphertext. This contradicts the principal of user privacy-preserving. In addition, overheads of the decryption are significant and will increase with the size of the plaintext space. Another popular system is Boneh-Goh-Nissim cryptosystem [19]. However, the computation overheads due to the Bilinear map and the limitations of the plaintext space make it difficult to be applied in practice. Most existing schemes such as those reported in [14, 16] can be used to compute only the summation (or average) of user data, which limits the ability of CC to perform other essential and complex statistical and data analysis. In other words, how to perform complex statistical and data analysis of user data without compromising user privacy remains a research challenge and opportunity.

Chen et al. proposed a multifunctional data aggregation (MuDA) scheme [10], designed to achieve privacy-preserving aggregation of multiple functions such as average, variance, and one-way ANOVA. MuDA, however, uses Pollard's lambda method in decryption to obtain the aggregation value, which is computationally expensive. In order to improve the efficiency of the MuDA scheme and enable CC to compute more complex statistical operations in a privacy-preserving way, we propose a fine-grained data analysis (FGDA) scheme for the smart grid communications.

The main contributions of the proposed FGDA scheme are three-fold:

- Different from most existing schemes, FGDA provides fault tolerance for users. That is, FGDA can still function even some SMs fail to report the electricity usage data.
- FGDA allows CC to compute more statistical functions. In our FGDA scheme, each SM needs to report power usage data in privacy preserving form only once and CC obtains not only the summation, but also any other values such as average, variance, and skewness of the user usage data. With this fine-grained data analysis, CC can make better decisions.
- Compared to existing similar schemes that have both multifunction and fault tolerance, FGDA does not use either bilinear map or Pollard's lambda method in the decryption. Thus, FGDA has a lower computation overhead. Furthermore, FGDA enables CC to compute

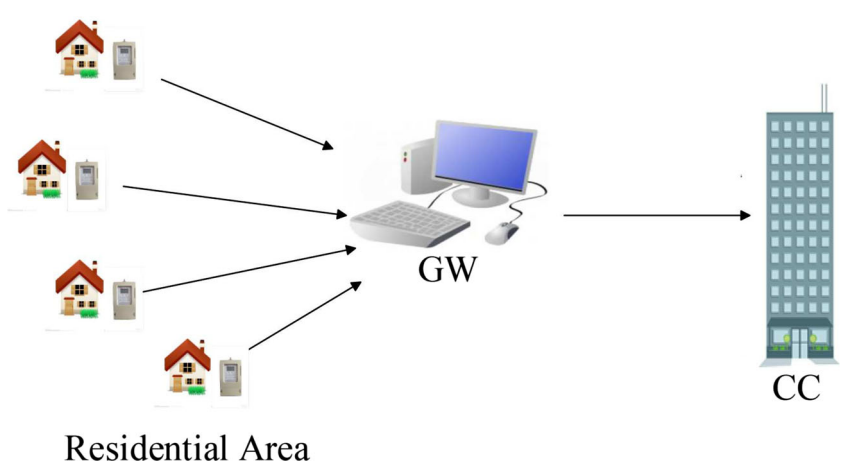


Fig. 1 A simply smart grid scenario

different statistical operations on the same ciphertext data from GW, while other related schemes require different data for different statistical operations. This allows FGDA to achieve a higher communication efficiency because it reduces the number of transmission to only one.

The rest of this paper is organized as follows. In Section 2, we review related work. We introduce the system model, security model and design goal in Section 3. The basic FGDA scheme and the advanced version are presented in Section 4 and Section 5, respectively. We give the security analysis in Section 6 and the performance evaluation in Section 7. Conclusion is drawn in the last section.

2 Related work

There are a number of data aggregation schemes designed for smart grid communications in the literature. In earlier literature, such as those of [20–22], aggregation security is achieved in a hop-by-hop manner. Specifically, electricity usage data are encrypted before reporting to a local GW, which then decrypts all received data for aggregation. After that, GW encrypts the aggregated result and forwards the encrypted data to CC. Consequently, such schemes have low efficiency. Furthermore, to protect users' privacy, there is a need to establish a secret key between the sender and the neighbors, which increases the communication costs.

Due to the proposal of homomorphic encryption techniques, a number of efficient privacy-preserving aggregation schemes have been proposed for smart grid communications [8–17]. The homomorphic property allows an aggregator (GW or CC in general) to perform operations directly on ciphertexts under the same key, without the need to first decrypt the data. For example, Lu et al. [14] proposed an efficient privacy-preserving data aggregation scheme in which CC can obtain a multidimensional electricity usage data of users by employing Paillier cryptosystem and a superincreasing sequence. The scheme can significantly improve communication efficiency and satisfy the real-time data collection requirements in smart grid communications. Sui et al. [23] proposed a robust secure data aggregation scheme using the Chinese Remainder Theorem and hash-based message authentication codes. However, the hop-by-hop communication mode in this scheme increases the communication costs.

In any real-world smart grid deployment, it is likely to have a few malfunction SMs. Therefore, Shi et al. [16] proposed a diverse grouping-based aggregation protocol. In their scheme, a key management center (KMC) classifies all SMs into different groups according to the distribution of their lifetimes. For each group, KMC distributes a random key k_i for each SM in this group such that the summation of all keys is equal to zero under module operation. In the event that some SMs fail to report the data, the aggregated data can still be decrypted using brute-force. However, solving the discrete logarithm problem significantly increases the computation overheads. In [9], the authors proposed a privacy-preserving data aggregation scheme with the ability of fault tolerance of both SMs and servers. In their scheme, when some SMs malfunction, a trusted authority (TA) will provide dummy ciphertexts (instead of the malfunctioned SMs) to the CC. This scheme uses the threshold secret sharing technology to handle the failure of servers.

A common limitation of the schemes mentioned above is that they can only be used to compute the summation of the usage data, while CC may need to perform more statistical analysis to manage the entire smart grid. In order to solve this problem, Lu et al. [10] proposed a multifunctional data aggregation scheme, which supports several aggregations including average, variance, and one-way ANOVA. The scheme needs to make several bilinear pair operations during aggregation and uses the Pollard's lambda method to decrypt the aggregated results by brute-forcing. However, these are computationally expensive operations.

Seeking to address existing limitations, we propose an efficient privacy-preserving aggregation scheme, FGDA, for smart grid communications in this paper.

3 Models and design goal

In this section, we introduce the system model, security model, and design goal for privacy-preserving smart grid communications.

3.1 System model

In this work, we mainly focus on the challenge of allowing CC to compute more data analysis (such as average, variance, and skewness) in a privacy-preserving and fault tolerant way. We consider a typical model (see Fig. 2), which includes a user set $U = \{u_1, u_2, \cdots, u_n\}$ living in a residential area (RA), a trusted authority TA, a control center CC, and a local gateway GW for RA.

- **TA**: TA is a trusted authority in charge of the entire system (e.g. CPS Energy in Texas), and whose main duties include initializing the system and distributing keys for CC and the users. In general, TA will be offline after initializing the system. In other words, it would not directly participate in the communication unless some exceptions occur in the reporting.
- **CC**: CC acts as an indispensable "brain" of a smart grid, whose duty is to collect users' near real-time electricity

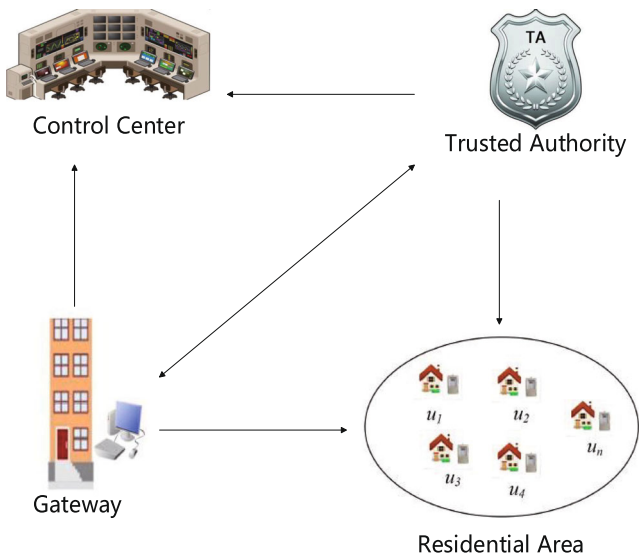


Fig. 2 A typical system model

usage data, compute various statistics of these data, and make decisions for the smart grid. In this paper, CC can compute not only the summation of users' data, but also average, variance, and skewness. For example, by computing the summation of users' data, CC can make real-time power pricing decisions [24, 25], detect power leakage [26]. Also, by computing the variance of users' data, CC understands the uniformity of electricity usage distribution; and by computing the skewness of users' data, CC knows the difference between sample distribution and normal distribution, and the distribution of the users' electricity usage data. With such fine-grained data analysis, CC can make informed decisions for smart grid operation.

– **GW**: GW is a powerful entity in this system, which collects user usage data and forwards the aggregated result to CC. We assume that GW will report user identities to TA in the event that any user's SM malfunction; otherwise, it will not be possible to replace or repair the malfunctioned SM.
– **SMs**: For each user $u_i \in U$, there is a unique SM_i that periodically collects and reports the electricity usage data of the IEDs of u_i to GW.

3.2 Security model

We assume that there is a probabilistic polynomial time (PPT) adversary $\mathcal{A}$ whose goal is to compromise the privacy of as many users as possible. $\mathcal{A}$ is not only able to eavesdrop on the communications among the users, GW, and CC, but is also capable of compromising the database of CC to exfiltrate the stored data or any other messages. Furthermore, GW and CC are assumed to be honest-but-curious. In other words, both parties will not deviate from the defined protocol, but will attempt to learn private information of the users.

We will now consider the security requirements for a smart grid.

– **Data Confidentiality**: SMs transmit the electricity usage data to GW. Because these data are related to user behavior and habits, it might be used to analyze a targeted user's sensitive information. Thus, GW should be able to aggregate user data in a privacy-preserving way (i.e. in a ciphertext form). Then, the users' data confidentiality can be assured.
– **Authentication and Data Integrity**: Determining whether an encrypted report is sent from a legitimate user is important in smart grid communications. The final result may be inaccurate due to the interferences made by $\mathcal{A}$, who might impersonate a "honest" user and send a report to GW. The malicious operations should be detected and GW accepts only the reports from legitimate users. If $\mathcal{A}$ compromises CC's database, then any user's private data cannot be disclosed since the values in the database are not about an individual user's electricity usage data. This allows user data to be protected and the integrity of users' data assured.

In addition, the goal of $\mathcal{A}$ is to violate the users' privacy. In the security analysis, we demonstrate that $\mathcal{A}$ is unable to infer the users' private keys. Thus, the security of communication flows in our proposed scheme can be guaranteed.

3.3 Design goal

Under the aforementioned system model and security model, our design goal is to develop an efficient fine-grained data analysis scheme while preserving user data privacy in smart grid communications. The following targets should be achieved.

– *Computations of fine-grained data analysis should be allowed.* In order to ensure a smooth operation of the smart grid and detect abnormal/suspicious conditions, CC should be able to perform complex statistical operations of the user data, such as average, variance, and skewness. It is important to also ensure that the proposed scheme allows GW to compute aggregation without decryption, and only CC can decrypt the final result.
– *Security requirement should be satisfied.* If the communications in smart grids are insecure, then the users will not use the system. In addition, any security leakage or compromise of user data may result in civil litigation against the smart grid operator, resulting in financial and legal implications or consequences.

– *Efficiency of communication should be achieved.* In order to reduce the computation overheads and communication costs, the users should be able to encrypt the electricity usage data and report the encrypted data to GW only once in a billing/collection period. CC can then perform several statistical operations on the user data; thus, resulting in a significant reduction of computation overheads and communication costs.

4 Basic FGDA scheme

The basic FGDA scheme presented in this section consists of five components, namely: system initialization, key generation, user report generation, privacy-preserving report aggregation, and secure report reading.

4.1 System initialization

In the initialization procedure, TA bootstraps the entire system. Given a security parameter κ,TA chooses a large prime number p satisfying $|p| = \kappa$ and a secure hash function $H : \{0, 1\}^* \to \mathbb{Z}_p^*$. The system parameters are $params := (p, H)$.

4.2 Key generation

In this phase, TA distributes the private keys to each user in $U = \{u_1, u_2, \ldots, u_n\}$ and CC as follows.

– For each user $u_i \in U$, $i = 1, 2, \ldots, n$, TA chooses a random number $\alpha_i \in \mathbb{Z}_p^*$ and sends α_i to u_i as his/her private key via a secure channel.
– TA computes $\alpha_0 \in \mathbb{Z}_p^*$ such that

$$1 = \alpha_0 \cdot \prod_{i=1}^{n} \alpha_i \bmod p^4 = \prod_{i=0}^{n} \alpha_i \bmod p^4 \tag{1}$$

and sends α_0 to CC as its private key via a secure channel.

4.3 User report generation

Assume that the reporting time points are fixed as $T = \{t_1, t_2, \ldots, t_\ell\}$ for a sufficient long runtime period and the power usage data of user u_i at time point t_γ is $m_{i,\gamma}$, $1 \le i \le n$, $1 \le \gamma \le \ell$. In practice, the electricity usage data for each residential user should not be extremely high (and this is a sensible expectation). Thus, we have the following relational assumption:

$$M = \max\{m_{i,\gamma} | 1 \le i \le n, 1 \le \gamma \le \ell\} < \sqrt[3]{p}/n. \tag{2}$$

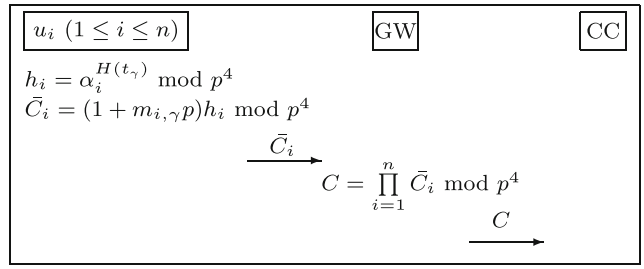


Fig. 3 Communication flows of the basic FGDA scheme

At time point t_γ, each user $u_i \in U$ collects its usage data $m_{i,\gamma}$ and performs the following steps (refer to Fig. 3):

– u_i computes a hash value

$$h_i = \alpha_i^{H(t_\gamma)} \bmod p^4$$

with his/her private key α_i for the reporting time point t_γ.
– u_i computes

$$\bar{C}_i = (1 + m_{i,\gamma} \cdot p) \cdot h_i \bmod p^4$$

as the encrypted version of data $m_{i,\gamma}$.
– u_i reports $\bar{C}_i$ to GW.

4.4 Privacy-preserving report aggregation

As shown in Fig. 3, after receiving the reporting messages of all users in U, GW computes the encrypted aggregation under different modules according to the maximum degree of statistical functions (as polynomial functions of n variables) required for various applications. Specially, if the maximum degree of considering statistic functions is λ, then the modulus should be set to $p^{\lambda+1}$. In this paper, we consider three statistical functions, namely: average, variance, and skewness, of which the skewness has the maximum degree 3. Thus, we set the modulus to be p^4. That is, GW computes

$$\begin{aligned} C &= \prod_{i=1}^{n} \bar{C}_i \bmod p^4 \\ &= \prod_{i=1}^{n} (1 + m_{i,\gamma} \cdot p) \cdot h_i \bmod p^4 \\ &= \prod_{i=1}^{n} (1 + m_{i,\gamma} \cdot p) \cdot \alpha_i^{H(t_\gamma)} \bmod p^4 \\ &= \prod_{i=1}^{n} (1 + m_{i,\gamma} \cdot p) \left(\prod_{i=1}^{n} \alpha_i \right)^{H(t_\gamma)} \bmod p^4. \end{aligned}$$

Then, GW reports C to CC for the fine-grained data analysis.

4.5 Secure report reading

After receiving the encrypted aggregation result C from GW, CC can compute multiple statistical functions based on the different requirements.

4.5.1 Average

If CC wishes to know the average of the electricity usage data at time point t_γ, then it computes the following:

– With the private key α_0, CC computes

$$\begin{aligned} S' &= C \cdot \alpha_0^{H(t_\gamma)} \bmod p^2 \\ &= \prod_{i=1}^{n}(1 + m_{i,\gamma} \cdot p) \left(\prod_{i=1}^{n} \alpha_i\right)^{H(t_\gamma)} \alpha_0^{H(t_\gamma)} \bmod p^2 \\ &= \prod_{i=1}^{n}(1 + m_{i,\gamma} \cdot p) \left(\alpha_0 \cdot \prod_{i=1}^{n} \alpha_i\right)^{H(t_\gamma)} \bmod p^2 \\ &= \prod_{i=1}^{n}(1 + m_{i,\gamma} \cdot p) \bmod p^2 \\ &= 1 + p \cdot \sum_{i=1}^{n} m_{i,\gamma}. \end{aligned}$$

The last two equations hold because we have $\prod_{i=0}^{n} \alpha_i = 1 \bmod p^2$ by Eq. 1 and $\sum_{i=1}^{n} m_{i,\gamma} < p$ by Eq. 2, respectively.

– CC obtains the summation of all user usage data as

$$S = \frac{S' - 1}{p} = \sum_{i=1}^{n} m_{i,\gamma}.$$

– The average of all user usage data is straightforward to compute:

$$A = \frac{1}{n} \cdot S.$$

It is clear that the usage data $m_{i,\gamma}$ of each individual user u_i at time point t_γ is protected during the calculation processes.

4.5.2 Variance

The variance has a central role in statistics and it informally measures the spread of the set of user electricity usage data from their average for smart grids. With the encrypted aggregation data C received from GW and the summation S (or the average A) computed in Section 4.5.1, CC further makes the following computations for the variance.

– CC computes

$$\begin{aligned} B &= C \cdot \alpha_0^{H(t_\gamma)} \bmod p^3 \\ &= \prod_{i=1}^{n}(1 + m_{i,\gamma} \cdot p) \left(\prod_{i=1}^{n} \alpha_i\right)^{H(t_\gamma)} \alpha_0^{H(t_\gamma)} \bmod p^3 \\ &= \prod_{i=1}^{n}(1 + m_{i,\gamma} \cdot p) \left(\alpha_0 \cdot \prod_{i=1}^{n} \alpha_i\right)^{H(t_\gamma)} \bmod p^3 \\ &= \prod_{i=1}^{n}(1 + m_{i,\gamma} \cdot p) \bmod p^3 \\ &= 1 + p \cdot S + p^2 \cdot \sum_{1 \le i < j \le n} m_{i,\gamma} m_{j,\gamma}. \end{aligned}$$

The last two equations hold because we have $\prod_{i=0}^{n} \alpha_i = 1 \bmod p^3$ by Eq. 1 and $\sum_{1 \le i < j \le n} m_{i,\gamma} m_{j,\gamma} < p$ by Eq. 2, respectively.

– CC obtains

$$B_1 = \frac{B - 1 - p \cdot S}{p^2} = \sum_{1 \le i < j \le n} m_{i,\gamma} m_{j,\gamma}.$$

– With S and B_1, CC can get

$$\begin{aligned} B_2 &= S^2 - 2 \cdot B_1 \\ &= \left(\sum_{i=1}^{n} m_{i,\gamma}\right)^2 - 2 \sum_{1 \le i < j \le n} m_{i,\gamma} m_{j,\gamma} \\ &= \sum_{i=1}^{n} m_{i,\gamma}^2. \end{aligned}$$

– The variance of the users' electricity usage data can be computed as

$$V = \frac{1}{n} \cdot B_2 - \frac{1}{n^2} \cdot S^2.$$

4.5.3 Skewness

The skewness is a measure of the asymmetry of the probability distribution of a random variable about its mean. When CC needs to know the skewness of electricity data sample, it operates as follows.

– With the encrypted aggregation data C and its private key α_0, CC computes

$$\begin{aligned} C_1 &= C \cdot \alpha_0^{H(t_\gamma)} \bmod p^4 \\ &= \prod_{i=1}^{n}(1 + m_{i,\gamma} \cdot p) \left(\prod_{i=1}^{n} \alpha_i\right)^{H(t_\gamma)} \alpha_0^{H(t_\gamma)} \bmod p^4 \end{aligned}$$

$$= \prod_{i=1}^{n}(1 + m_{i,\gamma} \cdot p)\left(\alpha_0 \cdot \prod_{i=1}^{n}\alpha_i\right)^{H(t_\gamma)} \mod p^4$$
$$= \prod_{i=1}^{n}(1 + m_{i,\gamma} \cdot p) \mod p^4$$
$$= 1 + pS + p^2 B_1 + p^3 \sum_{1\le i<j<k\le n} m_{i,\gamma} m_{j,\gamma} m_{k,\gamma}.$$

The penultimate equation and the last equation hold because we have $\prod_{i=0}^{n}\alpha_i = 1 \mathrm{mod} p^4$ by Eq. 1 and $\sum_{1\le i<j<k\le n} m_{i,\gamma} m_{j,\gamma} m_{k,\gamma} < p$ by Eq. 2, respectively.

– CC computes

$$C_2 = \frac{C_1 - 1 - pS - p^2 B_1}{p^3} = \sum_{1\le i<j<k\le n} m_{i,\gamma} m_{j,\gamma} m_{k,\gamma}.$$

– CC gets C_3 as

$$C_3 = 3 \cdot C_2 + S \cdot (B_2 - B_1) = \sum_{i=1}^{n} m_{i,\gamma}^3.$$

The above equation holds because we have

$$\sum_{i=1}^{n} m_{i,\gamma}^3 - 3 \cdot C_2$$
$$= \sum_{i=1}^{n} m_{i,\gamma}^3 - 3 \sum_{1\le i<j<k\le n} m_{i,\gamma} m_{j,\gamma} m_{k,\gamma}$$
$$= \left(\sum_{i=1}^{n} m_{i,\gamma}\right)\left(\sum_{i=1}^{n} m_{i,\gamma}^2 - \sum_{1\le i<j\le n} m_{i,\gamma} m_{j,\gamma}\right)$$
$$= S \cdot (B_2 - B_1).$$

– CC obtains the skewness of user electricity usage data as

$$W = \frac{C_3 - \frac{3}{n} S B_2 + \frac{2}{n^2} S^3}{n V \sqrt{V}}.$$

5 Advanced FGDA scheme

In this section, we present an advanced version of the basic FGDA scheme described in Section 4. In the basic scheme, each user $u_i \in U$ introduces the "disturbed" item $\alpha_i^{H(t_\gamma)}$ into his/her ciphertext and any adversary is unable to disclose the true electricity usage data of u_i without the private key α_i. On the other hand, as shown in Eq. 1, the product of the n private keys α_i $(1 \le i \le n)$ of users and the private key α_0 of CC is designed to be 1 modulo p^4. This enables CC to cancel out the disturbance during the decryption processes and recover the summation of all users' data without learning any individual user's data. User privacy is, thus, protected.

A major limitation of the basic FGDA scheme is that it is not fault tolerant. That is, if any single SM fails to report the data or any data sent from legal SMs are tampered by an adversary, CC would not be able to learn anything as the disturbance in reported data cannot be canceled out. This is a realistic issue that needs to be resolved, as hardware failure or attacks are unavoidable in practice [10]. Thus, we propose the advanced FGDA scheme with message authentication to ensure the authenticity of reported data and locate the malfunction SMs in real-time. The new scheme has five components, described in the following subsections.

5.1 System initialization

Given a security parameter κ, TA chooses a large prime number p satisfying $|p| = \kappa$ and a secure hash function $H : \{0,1\}^* \to \mathbb{Z}_p^*$. The system public parameters are $params = (p, H)$.

5.2 Key generation

TA distributes a private key $\alpha_i \in \mathbb{Z}_p^*$ to each user u_i in $U = \{u_1, u_2, \ldots, u_n\}$ and a private key $\alpha_0 \in \mathbb{Z}_p^*$ to CC via secure channels, such that $1 = \alpha_0 \cdot \prod_{i=1}^{n}\alpha_i \mathrm{mod} p^4$ (i.e. Eq. 1 in Section 4.2). Furthermore, we assume that there exists a session key $K_{u_i-\mathrm{GW}}$ between each user u_i and GW, $i = 1, 2, \ldots, n$.

5.3 User report generation

Similar to Section 4.3, we assume that the reporting time points are $\{t_1, t_2, \ldots, t_\ell\}$ and each user u_i has an identity ID_i and its electricity usage data is $m_{i,\gamma}$ at time point t_γ, $1 \le i \le n$, $1 \le \gamma \le \ell$. Further, we denote by U_γ the set of users whose SMs are functioning at time point t_γ. To report $m_{j,\gamma}$ to GW, each user $u_j \in U_\gamma$ (refer to Fig. 4):

– computes $h_j = \alpha_j^{H(t_\gamma)} \mathrm{mod} p^4$ with his/her private key α_j;

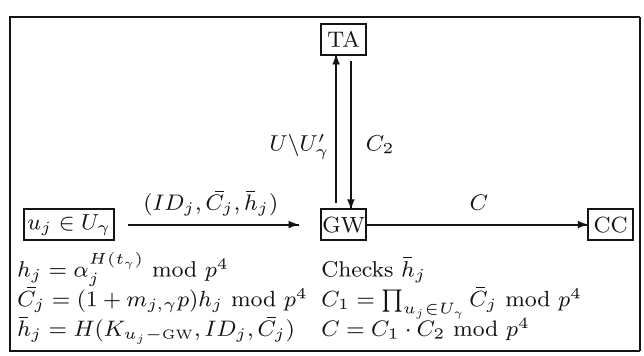


Fig. 4 Communication flows of the advanced FGDA scheme

– computes the encrypted data

$$\bar{C}_j = (1+m_{j,\gamma}\cdot p)\cdot h_j = (1+m_{j,\gamma}\cdot p)\cdot \alpha_j^{H(t_\gamma)} \bmod p^4$$

and hash value $\bar{h}_j = H(K_{u_j-\mathrm{GW}}, ID_j, \bar{C}_j))$;
– sends $(ID_j, \bar{C}_j, \bar{h}_j)$ to GW.

5.4 Privacy-preserving report aggregation

At each time point t_γ, we assume that GW maintains an initial empty table denoted by U'_γ. Upon receiving the report messages from users at time point t_γ, GW executes the following steps:

– For each message $(ID_j, \bar{C}_j, \bar{h}_j)$ of user u_j, GW checks whether $\bar{h}_j = H(K_{u_j-\mathrm{GW}}, ID_j, \bar{C}_j)$ with the received $ID_j, \bar{C}_j$ and the shared session key $K_{u_j-\mathrm{GW}}$ between the user u_j and GW. If yes, then GW accepts the message and records the identity ID_j in U'_γ; otherwise, GW discards the message as it implies a distortion or a forgery.

  In the following, we denote by Idx_1 and Idx_2 the index sets of the users in U'_γ and $U\backslash U'_\gamma$, respectively. That is,

$$\mathsf{Idx}_1 = \{j|u_j \in U'_\gamma\} \text{ and } \mathsf{Idx}_2 = \{k|u_k \in U\backslash U'_\gamma\}.$$

– GW aggregates all valid data by computing

$$\begin{aligned} C_1 &= \prod_{j\in\mathsf{Idx}_1} \bar{C}_j \bmod p^4 \\ &= \prod_{j\in\mathsf{Idx}_1} (1+m_{j,\gamma}\cdot p)\left(\prod_{j\in\mathsf{Idx}_1}\alpha_j\right)^{H(t_\gamma)} \bmod p^4. \end{aligned}$$

– For all users u_k, $k \in \mathsf{Idx}_2$, GW sends their identities ID_k to TA for the corresponding dummy ciphertexts.
– For all requirements from GW, TA finds the corresponding private key α_k, $k \in \mathsf{Idx}_2$, in TA's local database and computes

$$C_2 = \left(\prod_{k\in\mathsf{Idx}_2}\alpha_k\right)^{H(t_\gamma)} \bmod p^4.$$

TA returns C_2 to GW.

– Upon receiving C_2 from TA, GW computes the final aggregated result as

$$\begin{aligned} C &= C_1\cdot C_2 \bmod p^4 \\ &= \prod_{j\in\mathsf{Idx}_1}(1+m_{j,\gamma}\cdot p)\left(\prod_{j\in\mathsf{Idx}_1}\alpha_j\right)^{H(t_\gamma)} \\ &\quad\cdot\left(\prod_{k\in\mathsf{Idx}_2}\alpha_k\right)^{H(t_\gamma)} \bmod p^4 \\ &= \prod_{j\in\mathsf{Idx}_1}(1+m_{j,\gamma}\cdot p)\left(\prod_{i=1}^{n}\alpha_i\right)^{H(t_\gamma)} \bmod p^4 \end{aligned}$$

and reports the ciphertext C together with the number $r = |\mathsf{Idx}_1|$ of the functioning SMs to CC.

5.5 Secure report reading

After obtaining C and r from GW at time point t_γ, CC first cancels out the disturbed item $\prod_{i=1}^n \alpha_i^{H(t_\gamma)}$ in C with its private key α_0 by computing

$$\begin{aligned} \hat{C} &= C\cdot\alpha_0^{H(t_\gamma)} \bmod p^4 \\ &= \prod_{j\in\mathsf{Idx}_1}(1+m_{j,\gamma}\cdot p)\left(\prod_{i=1}^{n}\alpha_i\right)^{H(t_\gamma)}\alpha_0^{H(t_\gamma)} \bmod p^4 \\ &= \prod_{j\in\mathsf{Idx}_1}(1+m_{j,\gamma}\cdot p)\left(\alpha_0\cdot\prod_{i=1}^{n}\alpha_i\right)^{H(t_\gamma)} \bmod p^4 \\ &= \prod_{j\in\mathsf{Idx}_1}(1+m_{j,\gamma}\cdot p) \bmod p^4. \end{aligned}$$

Then, by using the index set Idx_1 and the integer r (instead of $\{1, 2, \ldots, n\}$ and n, respectively), CC can compute the statistic functions average, variance, and skewness of the r valid usage data $\{m_{j,\gamma}|j \in \mathsf{Idx}_1\}$ of the users in U_γ as in Section 4.5. As an example, the average of the data $\{m_{j,\gamma}|j \in \mathsf{Idx}_1\}$ can be computed in a privacy-preserving form as follows.

– With $\hat{C}$, CC computes

$$S' = \hat{C} \bmod p^2 = 1 + p\cdot\sum_{j\in\mathsf{Idx}_1} m_{j,\gamma}.$$

– CC obtains the summation of all valid usage data as

$$S = \frac{S'-1}{p} = \sum_{j\in\mathsf{Idx}_1} m_{j,\gamma}.$$

– The average of all valid data is straightforward to compute:

$$A = \frac{1}{r}\cdot S.$$

6 Security analysis

In this section, we discuss the security issues of the basic and advanced FGDA schemes. As discussed in Section 3.2, we consider an adversary $\mathcal{A}$ who can eavesdrop on communication flows in the system and compromise CC's database(s). We also assume that $\mathcal{A}$ can tamper with the report sent from legitimate SMs. We explain how our proposed schemes can resist attacks launched by $\mathcal{A}$ in our security model.

- *User electricity usage data are protected from eavesdropping.* In order to snoop on users' individual activities, $\mathcal{A}$ may hide in RA and eavesdrop on wireless communication flows between the users and GW. Assume that $\mathcal{A}$ has eavesdropped on a ciphertext $\bar{C}_i = (1 + m_{i,\gamma} \cdot p)\alpha_i^{H(t_\gamma)} \bmod p^4$ of user u_i at time point t_γ. Since $m_{i,\gamma}$ is usually a small value in one time interval, $\mathcal{A}$ may attempt to launch a brute-force attack by exhaustively testing each possible value of $m_{i,\gamma}$. Without u_i's secret key α_i, however, $\mathcal{A}$ is unable to cancel out the disturbance of $\alpha_i^{H(t_\gamma)}$. Thus, $\mathcal{A}$ is unable to obtain the real-time sensitive value $m_{i,\gamma}$ and the privacy of users' electricity usage data is guaranteed.
- *User electricity usage data and final result will not be disclosed at GW.* Recall that one main responsibility of GW is to collect all user ciphertexts and aggregate them into a single ciphertext $C = \prod_{i=1}^{n}(1 + m_{i,\gamma} \cdot p)\prod_{i=1}^{n} \alpha_i^{H(t_\gamma)} \bmod p^4$. Assume that $\mathcal{A}$ has obtained the ciphertext C by eavesdropping on the communication between GW and CC. Without CC's private key α_0, $\mathcal{A}$ will learn nothing from C because of the disturbed item $\prod_{i=1}^{n} \alpha_i^{H(t_\gamma)}$ of C. In addition, since u_i's private key α_i is distributed by TA via a secure channel and the disturbed item $\alpha_i^{H(t_\gamma)}$ of ciphertext $\bar{C}_i = (1 + m_{i,\gamma} \cdot p)\alpha_i^{H(t_\gamma)} \bmod p^4$ is changed at different reporting time points, $\mathcal{A}$ will learn nothing about α_i even if $\mathcal{A}$ has $\bar{C}_i$ in hand, without having the private key α_0 of CC. For a honest-but-curious GW, it is unable to obtain useful information from the encrypted data C or $\bar{C}_i$ without α_i and α_0. Thus, individual user's electricity usage data is protected.
- *Non-compromised user electricity usage data will not be revealed.* Assume that $\mathcal{A}$ has compromised several users and obtained the secret information stored in the SMs of these users. Since the private key of each user is randomly chosen by TA, and knowing one user's private key reveals nothing about other private keys. Thus, $\mathcal{A}$ is unable to violate the non-compromised users' privacy simply by learning the secret information of compromised users. Even in the extreme scenario that $\mathcal{A}$ compromised $n - 1$ users and obtained their private keys, $\mathcal{A}$ is still unable to obtain the remaining non-compromised user's private key and electricity usage data.
- *User electricity usage data will not be forged.* We consider a stronger adversary $\mathcal{A}$ for the advanced FGDA scheme who can forge or tamper a ciphertext of user u_i as the form $\bar{C} = (1 + \bar{m}_{i,\gamma} \cdot p) \cdot \bar{\alpha}_i^{H(t_\gamma)}$. However, $\mathcal{A}$ does not know the shared secret key $K_{u_i-\mathrm{GW}}$ between u_i and GW. So, the adversary has to choose a random number h (rather than the real hash value of $H(K_{u_i-\mathrm{GW}}, ID_j, \bar{C})$). When GW receives the forged report $(ID_i, \bar{C}, h)$, it will reject this report because the integrity check $h = H(K_{u_i-\mathrm{GW}}, ID_j, \bar{C})$ will fail. This ensures that GW aggregates only the ciphertexts of legitimate users' electricity usage data.

7 Performance evaluation

In this section, we compare the proposed scheme with Chen et al.'s MuDA scheme [10], in terms of computation overheads and communication costs.

7.1 Computation overheads

We focus on the computation overheads of common statistical functions (i.e. average and variance) in both MuDA and FGDA schemes. $T_{b,p}$ denotes the computation time of bilinear pairing, $T_{m,m}$ the time of modular exponent multiplication, $T_{m,p}$ the time of modular exponent power, and $T_{p,l}$ the time of Pollard's lambda method.

To compute the average of n users' power usage data in a privacy-preserving form, GW in the MuDA scheme needs to compute $A_{1,\gamma} = \prod_{i=1}^{n} C_{i,\gamma}$ after receiving users' reports $C_{i,\gamma} = H(t_\gamma)^{m_{i,\gamma}} \cdot h^{r_{i,\gamma}}$, where $1 \le i \le n$. The total computation time of aggregation is $(n-1)T_{m,m}$. While CC in the MuDA scheme needs to compute $A_{1,\gamma}^{p}$ and obtain the sum of users' data $\sum_{i=1}^{n} m_{i,\gamma}$ using the Pollard's lambda method, which results in a computation overhead $T_{m,p} + T_{p,l}$. In the FGDA scheme, the total computation time of GW (computing $C = \prod_{i=1}^{n} \bar{C}_i$) is the same as in the MuDA scheme, but CC (computing $S' = C \cdot \alpha_0^{H(t_\gamma)} \bmod p^2$) only has a

Table 1 Computation overhead comparison between MuDA and FGDA

		GW	CC
Average	MuDA	$(n-1)T_{m,m}$	$T_{m,p} + T_{p,l}$
	FGDA	$(n-1)T_{m,m}$	$T_{m,p} + T_{m,m}$
Variance	MuDA	$2(n-1)T_{m,m}$ $+(n+1)T_{b,p}$	$2T_{m,p} + 2T_{p,l}$
	FGDA	$(n-1)T_{m,m}$	$T_{m,p} + T_{m,m}$

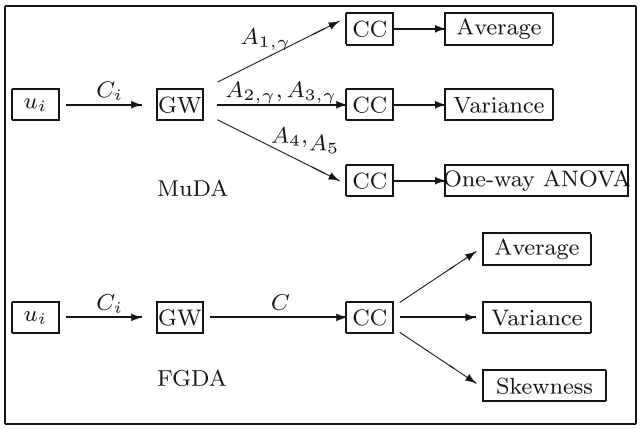


Fig. 5 Communication flow comparison between MuDA and FGDA

computation overhead of $T_{m,p}+T_{m,m}$, which is significantly lower than $T_{m,p}+T_{p,l}$ in the MuDA scheme.

As for the variance of n users' power usage data, GW in the MuDA scheme needs to compute

$$A_{2,\gamma}=e\left(\prod_{i=1}^{n} C_{i,\gamma}, \prod_{i=1}^{n} C_{i,\gamma}\right) \text{ and } A_{3,\gamma}=\prod_{i=1}^{n} e(C_{i,\gamma}, C_{i,\gamma}),$$

which results in a computation overhead of $2(n-1)T_{m,m}+(n+1)T_{b,p}$. On the other hand, CC in the MuDA scheme needs to compute $A_{2,\gamma}^{p}$ and $A_{3,\gamma}^{p}$ and then uses the Pollard's lambda method on $A_{2,\gamma}^{p}$ and $A_{3,\gamma}^{p}$ to obtain $\sum_{i=1}^{n} m_{i,\gamma}^2$ and $(\sum_{i=1}^{n} m_{i,\gamma})^2$. Thus, the computation overhead of CC is $2T_{m,p}+2T_{p,l}$. In the proposed FGDA scheme, GW makes the same computation for all three statistical functions, which has a computation overhead of $(n-1)T_{m,m}$. For CC, the main computation overhead comes from the calculation $B = C \cdot \alpha_0^{H(t_\gamma)}$, which results in a computation overhead $T_{m,p}+T_{m,m}$.

A comparative summary is presented in Table 1. Considering that both the Pollard's lambda method and the bilinear pairing operation are computationally expensive compared to modular exponent multiplication, it is clear that the proposed FGDA scheme has a lower computation overhead than the MuDA scheme.

7.2 Communication costs

We divide the communication flows into two parts: (1) from user u_i, $1 \le i \le n$, to GW and (2) from GW to CC. In part (1), each user u_i in the MuDA scheme reports $C_{i,\gamma} = H(t_\gamma)^{m_{i,\gamma}} h^{r_{i,\gamma}}$ to GW (in which h is a subgroup generator and $r_{i,\gamma}$ is a random number chosen by u_i [10]), while u_i in our FGDA scheme reports $C_i = (1+m_{i,\gamma})\alpha_i^{H(t_\gamma)}$ to GW (we omit the modulus for simplicity). This shows that the MuDA and FGDA schemes have a similar communication costs in part (1).

For part (2), GW in the MuDA scheme needs to send $A_{1,\gamma}$ to CC for the average function, $A_{2,\gamma}$ and $A_{3,\gamma}$ for the variance function, and $A_4 = e(g,g)^{\sum_{j=1}^{s}\sum_{i=1}^{n} m_{i,j}^2} \cdot e(g,h)^{R_4}$, $A_5 = e(g,g)^{\sum_{j=1}^{s}(\sum_{i=1}^{n} m_{i,j})^2} \cdot e(g,h)^{R_5}$ for the one-way ANOVA function [10]. While in our FGDA scheme, GW needs to send only one aggregated result $C = \prod_{i=1}^{n}(1+m_{i,\gamma} \cdot p) \cdot \prod_{i=1}^{n} \alpha_i^{H(t_\gamma)}$ to CC for all the three statistical functions average, variance, and skewness (refer to Fig. 5). Therefore, the FGDA scheme has a higher communication efficiency than the MuDA scheme.

Now we consider the simulation results of the communication costs using two metrics: individual user costs and overall costs. We set the security parameter $\kappa = 512$ and thus, the size of u_i's report at a time point is $|\bar{C}_i| = 512$ bits (as in [10], we do not consider other payloads such as user ID and time stamp, which are relatively short compared

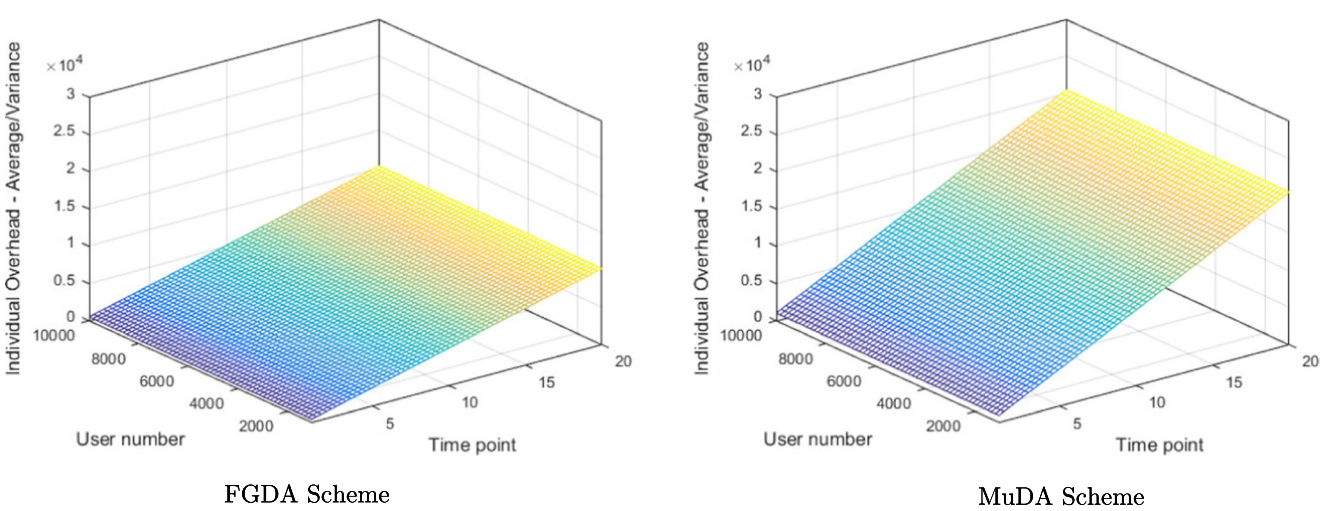


Fig. 6 Individual communication costs in the FGDA and MuDA schemes

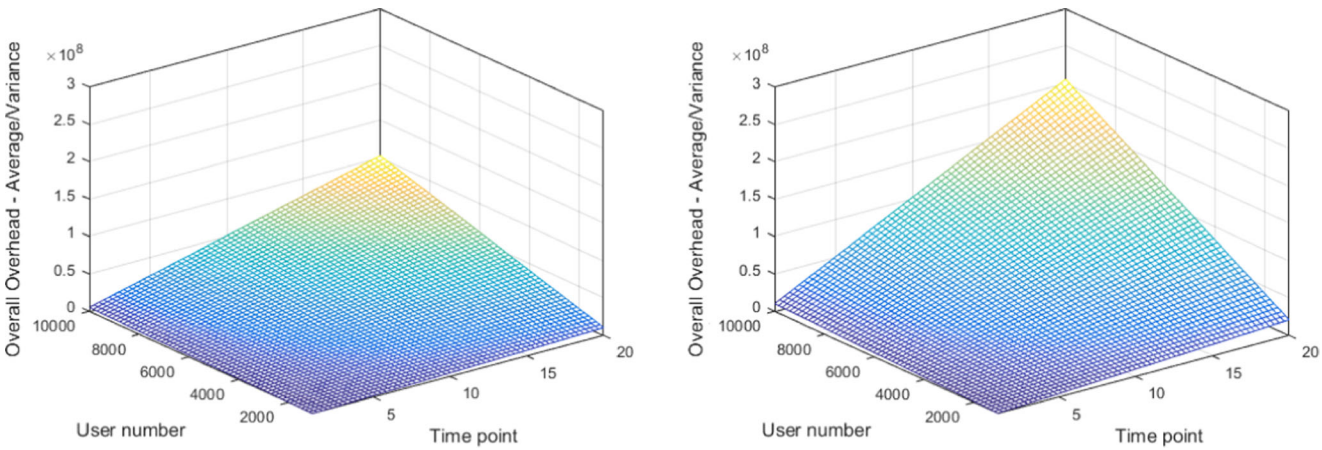


Fig. 7 Overall communication costs in the FGDA and MuDA schemes

with the report). We denote L_p as the length of p and thus, we have $L_p = 512$ bits. The communication costs of individual user in our scheme are always L_p at every time point, while the communication costs in the MuDA scheme are $2 \cdot L_p = 1024$ bits if the prime number p has the same size. Each user u_i in both schemes sends only one ciphertext to GW at each report time point. Figure 6 illustrates the comparative summary of the individual communication costs for the average and variance aggregations.

Finally, we consider the overall communication costs of both the FGDA and MuDA schemes. In the FGDA scheme, when CC wishes to compute the average of n users' electricity usage data, GW collects n users' reports and aggregates them into one report C for CC. Summing all n users' communication costs and the costs from GW to CC, we arrive at a total communication cost of $(n+1)L_p$. On the other hand, the total communication costs of average aggregation in the MuDA scheme are $2(n+1)L_p$. When CC wishes to calculate the variance of n users' data, the total communication costs of our FGDA scheme are still $(n+1)L_p$. The corresponding costs of the MuDA scheme are $2(n+2)L_p$ because GW needs to send $A_{2,\gamma}$ and $A_{3,\gamma}$ to CC for the computation of the variance. As shown in Fig. 7, it is easy to see that the proposed FGDA scheme significantly reduces the overall communication costs, especially when the number n of users is large.

8 Conclusion

With the rapidly urbanization of our society, there is tremendous pressure on traditional urban infrastructures. Hence, it is not surprising that smart city and related initiatives (e.g. smart grids) have been on the agendas of governments and the research communities worldwide. While smart city and related initiatives create new economic development opportunities and can potentially enhance the quality of living, there are underpinning security and privacy issues that need to be resolved, particularly in the fast evolving cyber threat landscape. One particular challenge is preserving the privacy of users in smart grids.

In this paper, we proposed a fine-grained data analysis (FGDA) scheme for smart grids. FGDA not only allows the control center to compute several statistical functions (average, variance, skewness, etc) of users' electricity usage data in a privacy-preserving form, but also supports fault tolerant feature. We then demonstrated the security of the proposed scheme, as well as evaluating its performance. Specifically, we demonstrated that the FGDA scheme achieves higher communication efficiency and lower computation overhead.

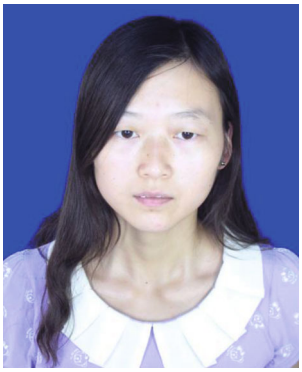
Future research includes implementing a prototype of the proposed scheme in collaboration with a smart grid operator in a monitored environment. This would allow us to evaluate and finetune the scheme, if necessary, to support other desirable properties.

Acknowledgments The work was supported in part by the NSFC-Zhejiang Joint Fund for the Integration of Industrialization and Informatization under Grant No. U1509219, the National Natural Science Foundation of China under Grant No. 61632012, the Shanghai Natural Science Foundation under Grant No. 17ZR1408400, and the Shanghai Sailing Program under Grant No. 17YF1404300. This work and the preparation of this publication were funded in part by monies provided by CPS Energy through an agreement with The University of Texas at San Antonio.

References

1. Heydt GT (2010) The next generation of power distribution systems. IEEE Trans Smart Grid 1(3):225–235

2. Li B, Lu R, Wang W et al (2017) Distributed host-based collaborative detection for false data injection attacks in smart grid cyber-physical system. J Parallel Distrib Comput 103:32–41
3. Liu J, Yu X, Xu Z et al (2017) A cloud-based taxi trace mining framework for smart city. Soft Practice and Exper 47(8):1081–1094
4. Choo KKR (2014) A conceptual interdisciplinary plug-and-play cyber security framework. In: ICTs and the millennium development goals. Springer US, pp 81–99
5. Yan Y, Qian Y, Sharif H et al (2012) A survey on cyber security for smart grid communications. IEEE Commun Surv Tutorials 14(4):998–1010
6. Li X, Liang X, Lu R et al (2012) Securing smart grid: cyber attacks, countermeasures, and challenges. IEEE Commun Mag 8(50):38–45
7. Pindoriya NM, Dasgupta D, Srinivasan D et al (2013) Infrastructure security for smart electric grids: a survey. In: Optimization and security challenges in smart power grids. Springer Berlin Heidelberg, pp 161–180
8. Bao H, Lu R (2015) A new differentially private data aggregation with fault tolerance for smart grid communications. IEEE Internet Things J 2(3):248–258
9. Chen L, Lu R, Cao Z (2015) PDAFT: A privacy-preserving data aggregation scheme with fault tolerance for smart grid communications. Peer-to-Peer Netw Appl 8(6):1122–1132
10. Chen L, Lu R, Cao Z et al (2015) MuDA: Multifunctional data aggregation in privacy-preserving smart grid communications. Peer-to-peer Netw Appl 8(5):777–792
11. Erkin Z (2015) Private data aggregation with groups for smart grids in a dynamic setting using CRT. In: 2015 IEEE international workshop on information forensics and security (WIFS), pp 1–6
12. Erkin Z, Tsudik G (2012) Private computation of spatial and temporal power consumption with smart meters. In: ACNS, vol 12, pp 561–577
13. Lu R, Alharbi K, Lin X et al (2015) A novel privacy-preserving set aggregation scheme for smart grid communications. In: 2015 IEEE global communications conference (GLOBECOM), pp 1–6
14. Lu R, Liang X, Li X et al (2012) Eppa: An efficient and privacy-preserving aggregation scheme for secure smart grid communications. IEEE Trans Parallel Distrib Syst 23(9):1621–1631
15. Ni J, Zhang K, Lin X et al (2016) EDAT: Efficient data aggregation without TTP for privacy-assured smart metering. In: 2016 IEEE international conference on communications (ICC), pp 1–6
16. Shi Z, Sun R, Lu R et al (2015) Diverse grouping-based aggregation protocol with error detection for smart grid communications. IEEE Trans Smart Grid 6(6):2856–2868
17. Xie CR, Zhang RY (2015) Privacy-preserving power consumption data measuring protocol for smart grid. In: Proceedings of international conference on computer information systems and industrial applications, CISIA
18. Paillier P (1999) Public-key cryptosystems based on composite degree residuosity classes. In: Eurocrypt, vol 99, pp 223–238
19. Boneh D, Goh EJ, Nissim K (2005) Evaluating 2-DNF formulas on ciphertexts, vol 3378, pp 325–341
20. Hu L, Evans D (2003) Secure aggregation for wireless networks. In: 2003 symposium on applications and the internet workshops, 2003. Proceedings, pp 384–391
21. Mahimkar A, Rappaport TS (2004) SecureDAV: a secure data aggregation and verification protocol for sensor networks. In: Global telecommunications conference, 2004. GLOBECOM'04, vol 4. IEEE, pp 2175–2179
22. Przydatek B, Song D, Perrig A (2003) SIA: Secure information aggregation in sensor networks. In: Proceedings of the 1st international conference on Embedded networked sensor systems, pp 255–265
23. Sui Z, Niedermeier M, de Meer H (2015) RESA: a robust and efficient secure aggregation scheme in smart grids. In: International conference on critical information infrastructures security. Springer, Cham, pp 171–182
24. Liang X, Li X, Lu R et al (2013) UDP: Usage-based dynamic pricing with privacy preservation for smart grid. IEEE Trans Smart Grid 4(1):141–150
25. Lin HY, Tzeng WG, Shen ST et al (2012) A practical smart metering system supporting privacy preserving billing and load monitoring. In: Applied cryptography and network security. Springer Berlin/Heidelberg, pp 544–560
26. Kursawe K, Danezis G, Kohlweiss M (2011) Privacy-friendly aggregation for the smart-grid. In: International symposium on privacy enhancing technologies symposium. Springer, Berlin, Heidelberg, pp 175–191



Shanshan Ge received the Bachelor degree in mathematics from Huaibei Normal University in 2015. She is currently a postgraduate at East China Normal University, Shanghai, China. Her research interests are cryptography, network security, and privacy-preservation in smart grids.

Peng Zeng received the PhD degree in computer science and technology from Shanghai Jiao Tong University, Shanghai, China, in 2009. He is currently an associate professor with the East China Normal University, Shanghai, China. His current research interests include applied cryptography, network information security, and coding theory.

Rongxing Lu has been an assistant professor at the Faculty of Computer Science, University of New Brunswick (UNB), Canada, since August 2016. Before that, he worked as an assistant professor at the School of Electrical and Electronic Engineering, Nanyang Technological University (NTU), Singapore from April 2013 to August 2016. His research interests include applied cryptography, privacy enhancing technologies, and IoT-Big Data security and privacy.

Kim-Kwang Raymond Choo received the Ph.D. degree in information security from the Queensland University of Technology, Australia, in 2006. He currently holds the Cloud Technology Endowed Professorship at The University of Texas at San Antonio. He serves on the editorial board of Computers & Electrical Engineering, Cluster Computing, Digital Investigation, IEEE Access, IEEE Cloud Computing, IEEE Communications Magazine, Future Generation Computer Systems, Journal of Network and Computer Applications, PLoS ONE, Soft Computing, etc. He also serves as the Special Issue Guest Editor of ACM Transactions on Embedded Computing Systems (2017), ACM Transactions on Internet Technology (2016), Computers & Electrical Engineering (2017), Digital Investigation (2016), Future Generation Computer Systems (2016, 2018), IEEE Cloud Computing (2015), IEEE Network (2016), IEEE Transactions on Cloud Computing (2017), IEEE Transactions on Dependable and Secure Computing (2017), Journal of Computer and System Sciences (2017), Multimedia Tools and Applications (2017), Personal and Ubiquitous Computing (2017), Pervasive and Mobile Computing (2016), Wireless Personal Communications (2017) etc. In 2016, he was named the Cybersecurity Educator of the Year - APAC (Cybersecurity Excellence Awards are produced in cooperation with the Information Security Community on LinkedIn), and in 2015 he and his team won the Digital Forensics Research Challenge organized by Germany's University of Erlangen-Nuremberg. He is the recipient of ESORICS 2015 Best Paper Award, 2014 Highly Commended Award by the Australia New Zealand Policing Advisory Agency, Fulbright Scholarship in 2009, 2008 Australia Day Achievement Medallion, and British Computer Society's Wilkes Award in 2008. He is also an IEEE Senior Member, and a Fellow of the Australian Computer Society.