This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2673843, IEEE Transactions on Smart Grid

Differentially Private Smart Metering with Fault Tolerance and Range-Based Filtering

Jianbing Ni, *Student Member, IEEE,* Kuan Zhang, *Member, IEEE,* Khalid Alharbi, Xiaodong Lin, *Fellow, IEEE,* Ning Zhang, *Member, IEEE,* Xuemin (Sherman) Shen, *Fellow, IEEE*

***Abstract*—Smart grid enables two-way communications between operation centers and smart meters to collect power consumption and achieve demand response to improve flexibility, reliability and efficiency of electricity system. However, power consumption data may contain users' privacy, e.g., activities, references and habits. Many smart metering schemes have been proposed utilizing homomorphic encryption for users' privacy preservation. Unfortunately, some abnormality of smart meter reading, e.g., caused by electricity theft, cannot be discovered since data is encrypted. Meanwhile, operation centers could become curious in reality. To address the above issues, we propose a new privacy-preserving smart metering scheme for smart grid, which supports data aggregation, differential privacy, fault tolerance and range-based filtering simultaneously. Specifically, we extend Lifted ElGamal encryption to aggregate users' consumption reports at the gateway to reduce communication overhead, while supporting fault tolerance of malfunctioning smart meters effectively. We also leverage zero-knowledge range proof to filter abnormal measurements caused by electricity theft or false data injection attacks without exposing individual measurements. In addition, our scheme can resist differential attacks, by which the curious operation center can violate users' privacy through comparing two aggregations of the similar data set. Finally, we discuss the properties of the proposed scheme and evaluate its performance in terms of security and efficiency.**

Keywords: Smart grid, privacy preservation, data security, smart meters.

I. Introduction

Smart grid is the evolution of the aging power grid, which integrates the traditional power grid with information and communication technologies to achieve two-way electricity and information exchange between operation centers and smart meters, making it more reliable, efficient, secure and green [1]. The primary breakthrough of smart grid is to make sure that electricity generation matches the demand of users for avoiding system instability due to voltage changes. To achieve this, users' real-time electricity consumption is measured, collected and analyzed by the operation center through advanced metering infrastructure for learning the electricity demand in a residential area, thereby adjusting the electricity generation to guarantee the balance of demand and supply [2].

Although the advantages of smart grid are attractive, the infrastructure is confronted with various cyber security threats [3], [4], [5], [6]. The consequences of security incidents may range from benign disruptions to deliberate acts of sabotage, threatening lives of citizens and even national security. For example, in December 2015, more than 225,000 people in Ukraine suffered from a blackout due to a devastating cyber attack on a power station [7]. Moreover, from the perspective of users, privacy is a primary concern as it is possible to infer their daily activities, habits and other privacy witnessable references from the electricity consumption data. It is reported that the determination of personal behavior patterns is the second most serious privacy consequences of smart grid systems in 14 threat types defined by Electronic Privacy Information Center [8]. A relatively low and static daily consumption of a household may indicate that no one is at home; power variation every several hours throughout every night might indicate that this family has a new baby [9]. End-to-end encryption is a straightforward way to hide the communication content and preserve users' privacy, but at the same time increases heavy overhead on communication and computation.

To address these issues, many secure data aggregation schemes [10], [11] have been proposed using homomorphic encryption to allow a semi-trust gateway to aggregate consumption reports in a specific residential area. These schemes usually assume that the smart meters are honest to report the users' power consumption without any corruption. Unfortunately, in reality, smart meters cannot provide solid security guarantee, although they are physically protected and tamper-proofing. Electricity theft [12] is not unusual, as reported that B.C hydro lost up to 3% of electricity to theft and the legitimate users had to bear at least 850 GWh in 2013 [13]. An attacker can simply place a strong magnet on a smart meter to stop counting, while still providing electricity to the user [14]. Utilities invest a lot to detect electricity theft by observing any abnormality of electricity consumption of users. However, it will not work after data is encrypted. Moreover, attackers may attempt to introduce malicious measurements by directly compromising smart meters to inject false data to affect power grid status assessment, thereby reducing the operation center's level of situational awareness. For example, an IOActive security consultant proved the weaknesses of smart metering architecture and designed a self-propagating worm to get remote control of about 15,000 smart meters in a 24-hour time at US Black Hat 2009 [15]. Therefore, it is of importance to design an efficient scheme to eliminate the false measurements in smart grid.

Meanwhile, operation centers are usually considered fully trusted, but in reality, may share the users' consumption data to

Jianbing Ni, Kuan Zhang, Ning Zhang and Xuemin (Sherman) Shen are with Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario, Canada N2L 3G1. Email: {j25ni,k52zhang, n35zhang, sshen}@uwaterloo.ca.

Khalid Alharbi and Xiaodong Lin is with Faculty of Business and Information Technology, University of Ontario Institute of Technology, Oshawa, Ontario, Canada L1H 7K4, Email: {Khalid.Alharbi, xiaodong.lin}@uoit.ca.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2673843, IEEE Transactions on Smart Grid

their cooperators. As a result, the private information of users is directly disclosed to unauthorized and untrusted entities. To prevent a curious operation center and dishonest cooperators from invading users' privacy, some schemes [16], [17] utilize a subtle key management technique, i.e., the sum of all keys are zero, to decentralize the operation center's capacity so as to enhance the security of smart metering schemes. However, one major limitation of these schemes is that they cannot support fault tolerance [18]. In other words, the smart metering schemes may be disrupted if one of the individual consumption reports is invalid or one of the malfunctioning smart meters fails to submit the measurement. Another challenging issue is that these schemes are vulnerable to differential attacks [18], launched by a curious operation center. If the operation center obtains two aggregations of individual reports, one is for n users, the other is for $n-1$ users in n users, the difference of two aggregations may expose the privacy of the user absent from the second aggregation. The differential attack has been discussed in several works [18], [19], [20], [21], while a few of them [18], [20], [21] are able to support fault tolerance, but seldom can prevent electricity theft and false data injection attacks, such that they are impractical in real scenarios. Hence, how to design an efficient privacy-preserving, differentially private and fault tolerant smart metering scheme that can resist electricity theft and detect false data in smart grid still deserves further investigations and efforts.

Motivated by the aforementioned challenges, in this paper, we propose a novel Differentially Private smart metering scheme (DiPrism) with fault tolerance and range-based filtering. Initiatively, to achieve range-based filtering, we consider that the user's habit of electricity usage keeps unchanged overtime, and utilize the historical power usage data to determine the range of current measurements and thereby discover the abnormal readings. The goal of fault tolerance is to handle general consumption report failures. These failures include the malfunction of smart meters and the occasions that the measurements do not fall in the required range. Specifically, the main contributions of this paper are as follows.

- We propose the DiPrism with the desirable features of data aggregation and differential privacy, under a new security model, in which both the gateway and the operation center are honest-but-curious and the users may be malicious, e.g., stealing electricity. By employing Lifted ElGamal encryption [22] and Laplace noise [23], we realize efficient data aggregation without exposing individual measurements to other parties, including other users, the gateway and the operation center.
- We utilize the zero-knowledge range proof [24] to eliminate abnormal measurements in smart metering. Specifically, the operation center predicts the range of electricity consumption according to historical power consumption data, load of appliances and types of users, i.e., community residents, factories and commercial residents. Based on this range, a user generates a zero-knowledge range proof to prove that his/her power consumption falls in the range with privacy preservation, and the center can determine that the measurement is polluted with high probability if it is not in this range. Therefore, all the measurements can be filtered based on the predicted range and the abnormal readings can be eliminated and also recorded for further investigation of possible electricity theft.
- We extend Lifted ElGamal encryption to support fault tolerance. Each user utilizes the public key of the operation center and a specific secret key, which is assigned by a trusted authority, to encrypt the individual measurement. The gateway transforms the ciphertexts generated by users to pairing-based ciphertexts and aggregates them to generate a compressed ciphertext, along with an auxiliary ciphertext, which is computed from users' public keys and can be used to decrypt the aggregated ciphertext even if some smart meters fail to report the measurements or some measurements are corrupted and eliminated.

The remainder of this paper is organized as follows. In section II, we define the system model, security threats and security goals. Then, we revisit the preliminaries in III and describe our DiPrism in section IV, followed by the security discussion in section V. We evaluate the performance in section VI and review the related work in section VII. Finally, we conclude our paper in section VIII.

II. Problem Statement

In this section, we state the problem by formalizing system model and security threats, and identifying security goals.

A. System Model

Advanced metering infrastructure of smart grid consists of four entities, operation center, gateways, users and smart meters, as depicted in Fig. 1. The operation center has the supply of electricity from plants and controls electricity transmission system to distribute electricity to users. To achieve the balance of demand and supply, the operation center measures, collects the users' electricity consumption via advanced metering infrastructure. It also determines the range of electricity consumption for each user according to the historical power usage data and the load of appliances, which is used to filter the abnormal readings. Each user, equipped with a smart meter, consumes the electricity provided by the operation center and accesses his/her daily consumption and electricity prices through the Internet. The smart meter provides smart grid interface between the user and the operation center. It counts the electricity consumed by the appliances in the household through a home energy management system and submits the measurement to the operation center at every reporting point. The gateway, which is a wireless access point or base station, is deployed to connect the operation center and smart meters in home area network. They mainly perform two functions: aggregation and relaying. The system time is divided into several time slots, e.g., every 15 minutes, and all smart meters are synchronous in time slots. At the beginning of each slot, i.e., the reporting point, a smart meter submits real-time electricity consumption to the gateway via a local area network, along with a zero-knowledge range proof, proving that the measurement is in the pre-defined range. Upon receiving the consumption reports from smart meters in the residential area,

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2673843, IEEE Transactions on Smart Grid

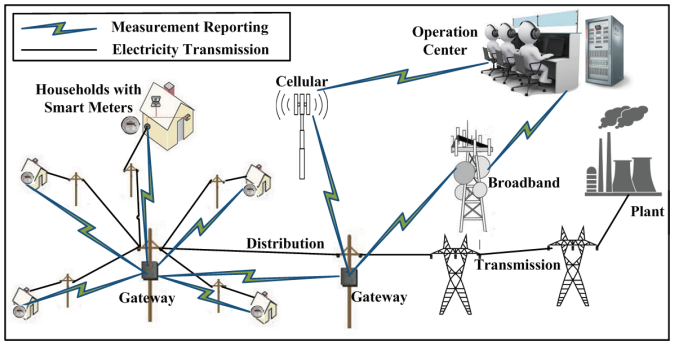


Fig. 1. System Model for Smart Metering

the gateway firstly verifies the range proofs to ensure that all the individual reports are valid, and then aggregates the reports to be a compressed one and forwards it to the center. The center recovers the sum of power consumption from the aggregated ciphertext and monitors electricity distribution and dynamic pricing to achieve the balance of demand and support.

B. Security Threats

Security threats in smart grid may come from both external and internal attackers. An external attacker may compromise a smart meter to inject false data or eavesdrop on the communication channel to violate user's privacy. The internal attackers include the users, the gateway and the operation center. The honest-but-curious gateway and operation center cannot tamper with the smart metering protocol, for example, they do not maliciously drop or distort any received message, but try all sorts of methods to seek and infer knowledge about users from their measurements and intermediate results. Meanwhile, we assume that the operation center and the gateway would not collude to pollute electricity measurements. A malicious user may not only be interested in the privacy of their neighbors, but also try to compromise smart meter for stealing electricity. The smart meters have protection against physical damage by using sophisticated microcontrollers from manufacturers, such as Analog Devices, Atmel and NXP Semiconductors. However, they are still vulnerable to electricity theft and false data injection attacks. In addition, a trusted authority (TA) is a fully trusted party, whose responsibility to issue random secret keys to users and the operation center through secure channels.

C. Security Goals

To enable smart metering under the aforementioned system model and resist the security threats, our scheme should achieve the following security goals.

- *Privacy Preservation*: Neither the individual power consumption reports nor the aggregated report would disclose any personal information of users. Therefore, an attacker cannot violate any user's privacy from the interactions between smart meters and the operation center.
- *Differential Privacy*: There should be a solid mechanism to resist the differential attacks. External attackers, the operation center and the local gateway are unable to infer any knowledge about users from the aggregations of two differential data sets.
- *Fault Tolerance*: The operation center should efficiently decrypt the aggregation of functioning smart meters, even in the presence of malfunctioning ones. It should also enable to recover the sum of the power consumption from the aggregation even if some individual readings are eliminated or some smart meters fail to report the measurements.
- *Range-Based Filtering*: The operation center should identify and filter the measurements those are illegally corrupted according to the range of electricity consumption.

III. Preliminaries

In this section, we briefly revisit the preliminaries used to construct the DiPrism, including Lifted ElGamal encryption scheme and differential privacy.

A. Lifted ElGamal Encryption

The Lifted ElGamal encryption scheme [22] is a variant of ElGamal encryption scheme, which is widely utilized in privacy-preserving applications due to its appealing additive homomorphism. Specifically, the cipherext of Lifted ElGamal encryption has two components. Given two ciphertexts $C = (C_1, C_2)$ and $C' = (C'_1, C'_2)$ of plaintexts m and m', respectively, anyone can compute the ciphertext C'' of $m'' = m + m'$ by computing the product $C'' = (C''_1, C''_2) = (C_1 C'_1 g^r, C_2 C'_2 h^r)$ for a random $r \in \mathbb{Z}_p^*$. Note that the calculation of C'' only requires two ciphertexts C and C'.

B. Differential Privacy

We use differential privacy to model the adversary model and discuss the noise generation technique. Generally, differential privacy guarantees that the presence or absence of any individual element in the database has only a limited impact on any output of a query.

ϵ-differential privacy [23]: A randomized algorithm $\mathcal{A}$ is *ϵ-differential privacy* if and only if for any datasets D and D' that differ on one element, and for all $T \subseteq \text{Range}(\mathcal{A})$,

$$\Pr[\mathcal{A}(D) \in T] \leq e^{\epsilon} \Pr[\mathcal{A}(D') \in T]. \quad (1)$$

More generally, differential privacy can be defined by requiring the equation above to hold on D and D' that are neighboring. An algorithm $\mathcal{A}$ provides *ϵ-differential privacy* if two outputs of a query on two datasets that differ on a single element only are not distinguishable. Therefore, an adversary cannot infer the value of any single element in the dataset. A smaller value of ϵ implies better privacy insurance, but lower accuracy of the query output.

The Laplace mechanism [23] is widely used to achieve *ϵ-differential privacy* by adding i.i.d Laplace noise $Lap(\lambda)$ to the accurate output of a query. The noise $Lap(\lambda)$ is sampled from a Laplace distribution, whose pdf is $\Pr[Lap(\lambda) = x] = \frac{1}{2\lambda} e^{|x|/\lambda}$. Assume a differentially private algorithm $\mathcal{A}$ is to answer a query on two datesets D and D' that differ on a single element, we have

$$\frac{\Pr(\mathcal{A}(D) = s)}{\Pr(\mathcal{A}(D') = s)} = \frac{\frac{1}{2\lambda} e^{\frac{|s - f(D)|}{\lambda}}}{\frac{1}{2\lambda} e^{\frac{|s - f(D')|}{\lambda}}} \leq e^{\frac{|f(D) - f(D')|}{\lambda}} \leq e^{\frac{\Delta f}{\lambda}}. \quad (2)$$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2673843, IEEE Transactions on Smart Grid

Here f is the numerical function and one wants to publish $f(D)$ in a way that satisfies ϵ-*differential privacy*. Δf denotes the global sensitivity of the function f, which is the maximum change of f between two neighboring datasets D and D', that is, $\Delta f = max_{D \simeq D'} |f(D) - f(D')|$, where $D \simeq D'$ denotes that D and D' are neighboring. Let $\epsilon = \Delta f / \lambda$, we have that $Pr[\mathcal{A}(D) \in T] \leq e^{\epsilon} Pr[\mathcal{A}(D') \in T]$, i.e., adding i.i.d Laplace noise $Lap(\lambda)$ to a query result for achieving ϵ-*differential privacy*, where λ denotes noise scale. The Laplace distribution $Lap(\lambda)$ is infinitely divisible. For every $n > 1$, $Lap(\lambda) = \sum_{i=1}^{n} (G(n, \lambda) - G'(n, \lambda))$, in which $G(n, \lambda)$ and $G'(n, \lambda)$ are i.i.d with Gamma distribution $g(x, n, \lambda)$. The Gamma distribution is defined as

$$g(x, n, \lambda) = \frac{(\frac{1}{\lambda})^{\frac{1}{n}}}{\Gamma(\frac{1}{n})} x^{\frac{1}{n}-1} e^{\frac{x}{\lambda}}, \tag{3}$$

where $\Gamma(\frac{1}{n})$ is the Gamma function evaluated at $1/n$.

In smart metering, f is the electricity consumption of a residential area and Δf is the maximum amount that any user can consume in a constant period. If the number of smart meters in a residential area is N, for each smart meter SM_i, we can add $G_i(N, \lambda) - G'_i(N, \lambda)$ to its measurement m_i before reporting. Thus, the sum of the power consumption in the residential area is

$$\sum_{i=1}^{N} m_i + \sum_{i=1}^{N} (G_i(N, \lambda) - G'_i(N, \lambda)) = \sum_{i=1}^{N} m_i + Lap(\lambda). \tag{4}$$

In this way, ϵ-*differential privacy* is satisfied.

IV. The Proposed Schemes

In this section, we propose the DiPrism, which includes four phases: system initialization, report generation, report aggregation and report reading. Then, we apply differential privacy to the DiPrism to against differential attacks.

A. The Basic DiPrism

The basic DiPrism mainly focuses on providing measurement aggregation, fault tolerance and range-based filtering for smart metering. Specifically, by means of Lifted ElGamal encryption scheme, the measurements of smart meters can be aggregated at the gateway in the residential area to improve the efficiency of smart metering. We also expand the Lifted ElGamal encryption to achieve ciphertext transformation, indicating that the ciphertexts obtained from Lifted ElGamal encryption can be transformed to pairing-based ciphertexts. The curious operation center can only decrypt the aggregated pairing-based ciphertext and handle general failures of measurement reports, but it is unable to decrypt the original ciphertexts generated by users. In addition, the electricity consumption of a household in a constant time period can be predicted since the residents generally follow their energy usage habits. Therefore, the measurement should lie in a specific range, which is pre-committed by the users or learnt from their historical data. The information flow of smart metering is shown in Fig.2. The construction of the basic DiPrism is described as follows:

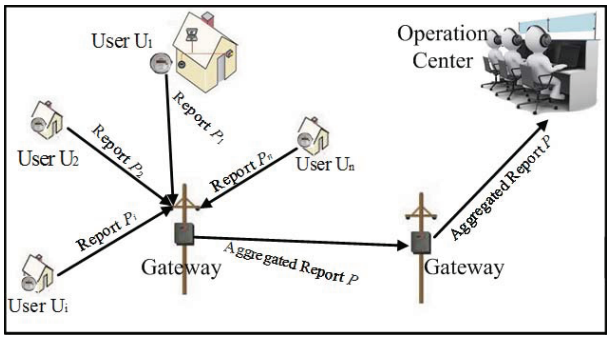


Fig. 2. Information Flow

1) System Initialization: The TA bootstraps the whole system at the beginning. Concretely, for the system initialization, given the security parameter κ, the TA generates a tuple $(p, \mathbb{G}, \mathbb{G}_1, \hat{e}, g, H, f)$, where p is a big prime. $\mathbb{G}$ and $\mathbb{G}_1$ are groups of the order p, $\hat{e} : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}_1$ is a bilinear map, $g \in_R \mathbb{G}$ is a random generator of $\mathbb{G}$, $H : \{0,1\}^* \rightarrow \mathbb{Z}_p^*$ is a cryptographic hash function and $f : \mathbb{Z}_p^* \times \mathbb{Z}_p^* \rightarrow \mathbb{Z}_p^*$ is a pseudo-random function with a secret key. The TA releases the system parameters $(p, \mathbb{G}, \mathbb{G}_1, \hat{e}, g, H, f)$.

Besides, the TA is responsible for assigning secret key materials to the residential users $U = \{U_1, \cdots, U_N\}$ and the operation center OC, where N is the number of users in the residential area. For each user $U_i \in U$, the TA firstly picks a random $s_i \in \mathbb{Z}_p$ and assigns s_i to U_i as its secret key through secure channels. The TA also chooses a random $k \in \mathbb{Z}_p$, which is assigned to all the residential users in U. To achieve the secure distribution of each (s_i, k), one recommended approach is to insert (s_i, k) in a trusted platform module and install this module on the smart meter during production. The TA also computes $s_0 \in \mathbb{Z}_p^*$, such that $s_0 + s_1 + \cdots + s_N = 0 \mod p$, and assigns s_0 to OC as its secret key.

The OC initializes the system by selecting a random $x \in_R \mathbb{Z}_p^*$ to compute $h = g^x \in \mathbb{G}$. OC releases h as its public key and keeps the corresponding secret key x secretly.

The gateway with an identifier GW also has its secret-public key pair (y, Y), where y is randomly picked from $\mathbb{Z}_p^*$ and $Y = \hat{e}(g, g)^y \in \mathbb{G}_1$.

A user U_i in the residential area is equipped with a smart meter with an identifier SM_i. U_i uses the assigned secret key s_i to compute the corresponding public key as $S_i = g^{s_i}$. In addition, U_i determines the range of the electricity consumption in a reporting period, $[\mathcal{W}_i, \mathcal{W}'_i]$, according to his/her usage habits and loads of appliances. This range $[\mathcal{W}_i, \mathcal{W}'_i]$ also can be resolved by OC based on the historical records of U_i's electricity usage. By means of $[\mathcal{W}_i, \mathcal{W}'_i]$, OC can determine the possible discrete readings of smart meters that fall in the range of $[\mathcal{W}_i, \mathcal{W}'_i]$, i.e., $w_{ij} \in [\mathcal{W}_i, \mathcal{W}'_i]$. Then, OC chooses a random value $v_i \in \mathbb{Z}_p^*$ and computes $V_i = g^{v_i}$, $W_{i,w_{ij}} = g^{\frac{1}{v_i + w_{ij}}}$ for each $w_{ij} \in [\mathcal{W}_i, \mathcal{W}'_i]$. Finally, OC publishes $(V_i, W_{i,w_{ij}})$ for U_i.

In addition, a digital signature scheme, e.g., DSA, is employed by users and the gateway to achieve the authentication and integrity of consumption reports during transmission from smart meters to the operation center.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2673843, IEEE Transactions on Smart Grid

2) Report Generation: Denote the reporting time points, e.g., every 15 min, as $\mathcal{T} = \{t_1, t_2, \cdots, t_{max}\}$ for a long time period. To report the real-time electricity consumption m_i at a specific time point $t_l \in \mathcal{T}$, U_i generates a consumption report P_i as follows:

- Compute $r_l = f(k, t_l)$, $C_l = g^{r_l}$, $D_i = g^{m_i} h^{r_l s_i}$;
- Randomly pick $u_i \in \mathbb{Z}_p^*$ to compute $Z_i = W_{i,m_i}^{u_i}$;
- Pick random values $\rho_{i1}, \rho_{i2}, \rho_{i3} \in \mathbb{Z}_p^*$, and compute $T_i = g^{\rho_{i1}} h^{r_l \rho_{i2}}$, $X_i = \hat{e}(Z_i, g)^{-\rho_{i1}} \hat{e}(g, g)^{\rho_{i3}}$;
- Compute $c_i = H(C_l, D_i, Z_i, t_l)$, $z_{m_i} = \rho_{i1} - c_i m_i$, $z_{s_i} = \rho_{i2} - c_i s_i$ and $z_{u_i} = \rho_{i3} - c_i u_i$;
- Set $P_i = (SM_i, C_l, D_i, Z_i, T_i, X_i, c_i, z_{m_i}, z_{s_i}, z_{u_i})$.

Finally, U_i sends P_i to the gateway GW.

3) Report Aggregation: Upon receiving the individual consumption reports from users in U, $P_i \in \{P_1, P_2, \cdots, P_n\}$, GW firstly checks whether all the measurements in consumption reports are valid by verifying whether the following equations are satisfied for each report P_i:

$$\hat{e}((T_i D_i^{-c_i} g^{-z_{m_i}})^{z_{s_i}^{-1}}, g) \stackrel{?}{=} \hat{e}(C_l, h), \tag{5}$$

$$X_i \stackrel{?}{=} \hat{e}(Z_i, V_i)^{c_i} \hat{e}(Z_i, g)^{-z_{m_i}} \hat{e}(g, g)^{z_{u_i}}. \tag{6}$$

If either of them does not hold for $P_{\hat{i}}$, the corresponding smart meter $SM_{\hat{i}}$ is compromised or the measurement is corrupted. GW removes the report $P_{\hat{i}}$ and aggregates the remaining individual reports into a compact report P. Otherwise, GW aggregates all the received reports to generate P. The procedures to generate P are described as follows:

- If all N smart meters work correctly, i.e., $n = N$, for each report P_i, GW uses its secret key y to compute
  $$C = \hat{e}(g^y, C_l) = \hat{e}(g^y, g^{r_l}),$$
  $$D = \prod_{i=1}^N \hat{e}(g^y, D_i) = \prod_{i=1}^N \hat{e}(g^y, g^{m_i} h^{r_l s_i}).$$
  For the range proofs, GW aggregates the individual reports as
  $$\widetilde{D} = \prod_{i=1}^N D_i^{c_i} = \prod_{i=1}^N g^{m_i c_i} h^{r_l s_i c_i},$$
  $$Z = \prod_{i=1}^N \hat{e}(Z_i, V_i)^{c_i} = \prod_{i=1}^N \hat{e}(W_{i,m_i}, g)^{v_i u_i c_i},$$
  $$F = \prod_{i=1}^N \hat{e}(Z_i, g) = \prod_{i=1}^N \hat{e}(W_{i,m_i}^{u_i}, g),$$
  $$T = \prod_{i=1}^N T_i = \prod_{i=1}^N g^{\rho_{i1}} h^{r_l \rho_{i2}},$$
  $$X = \prod_{i=1}^N X_i = \prod_{i=1}^N \hat{e}(Z_i, g)^{-\rho_{i1}} \hat{e}(g, g)^{\rho_{i3}},$$
  $$z_m = \sum_{i=1}^N z_{m_i} = \sum_{i=1}^N \rho_{i1} - c_i m_i,$$
  $$z_s = \sum_{i=1}^N z_{s_i} = \sum_{i=1}^N \rho_{i2} - c_i s_i,$$
  $$z_u = \sum_{i=1}^N z_{u_i} = \sum_{i=1}^N \rho_{i3} - c_i u_i.$$
  GW sets $P = (GW, C_l, C, D, \widetilde{D}, Z, F, T, X, z_m, z_s, z_u)$.
- If the smart meters of some users $\widehat{U} \subset U$ do not work, that is, $\widehat{U}$ do not report their data at time point t_l or report polluted measurements, GW uses y to compute
  $$C = \hat{e}(g^y, C_l) = \hat{e}(g^y, g^{r_l}),$$
  $$D = \prod_{i \in U \setminus \widehat{U}} \hat{e}(g^y, D_i) = \prod_{i \in U \setminus \widehat{U}} \hat{e}(g^y, g^{m_i} h^{r_l s_i}),$$
  $$\widehat{D} = \prod_{\hat{i} \in \widehat{U}} \hat{e}(C_l^y, S_{\hat{i}}) = \prod_{\hat{i} \in \widehat{U}} \hat{e}(g^{r_l y}, g^{s_{\hat{i}}}).$$
  In addition, GW calculates
  $$\widetilde{D} = \prod_{i \in U \setminus \widehat{U}} D_i^{c_i} = \prod_{i \in U \setminus \widehat{U}} g^{m_i c_i} h^{r_l s_i c_i},$$
  $$Z = \prod_{i \in U \setminus \widehat{U}} \hat{e}(Z_i, V_i)^{c_i} = \prod_{i \in U \setminus \widehat{U}} \hat{e}(W_{i,m_i}, g)^{v_i u_i c_i},$$
  $$F = \prod_{i \in U \setminus \widehat{U}} \hat{e}(Z_i, g) = \prod_{i \in U \setminus \widehat{U}} \hat{e}(W_{i,m_i}^{u_i}, g),$$
  $$T = \prod_{i \in U \setminus \widehat{U}} T_i = \prod_{i \in U \setminus \widehat{U}} g^{\rho_{i1}} h^{r_l \rho_{i2}},$$
  $$X = \prod_{i \in U \setminus \widehat{U}} X_i = \prod_{i \in U \setminus \widehat{U}} \hat{e}(Z_i, g)^{-\rho_{i1}} \hat{e}(g, g)^{\rho_{i3}},$$
  $$z_m = \sum_{i \in U \setminus \widehat{U}} z_{m_i} = \sum_{i \in U \setminus \widehat{U}} \rho_{i1} - c_i m_i,$$
  $$z_s = \sum_{i \in U \setminus \widehat{U}} z_{s_i} = \sum_{i \in U \setminus \widehat{U}} \rho_{i2} - c_i s_i,$$
  $$z_u = \sum_{i \in U \setminus \widehat{U}} z_{u_i} = \sum_{i \in U \setminus \widehat{U}} \rho_{i3} - c_i u_i.$$
  GW sets $P = (GW, C_l, C, D, \widehat{D}, \widetilde{D}, Z, F, T, X, z_m, z_s, z_u)$.

Finally, GW forwards P to OC.

4) Report Reading: Upon receiving the aggregated report P from the gateway, OC performs the following operations to recover the sum of power consumption in the residential area:

- If all N smart meters work correctly, OC verifies whether
  $$T \stackrel{?}{=} \widetilde{D} g^{z_m} C_l^{x z_s}, \qquad X \stackrel{?}{=} Z F^{-z_m} \hat{e}(g, g)^{z_u}.$$
  If both hold, OC computes $M = DC^{x s_0} = \hat{e}(g, g)^{y \sum_{i=1}^N m_i}$ and recovers the discrete log of M base Y using Pollard's lambda method [25] to obtain $m = \sum_{i=1}^N m_i$, which is the sum of the power consumption of residential users in U. Otherwise, OC has to use the divide-and-conquer approach to find the corrupted individual report, in which m_i does not fall in the predefined range $[\mathcal{W}_i, \mathcal{W}_i']$.
- If the smart meters of some users $\widehat{U} \subset U$ do not work, OC verifies whether
  $$T \stackrel{?}{=} \widetilde{D} g^{z_m} C_l^{x z_s}, \qquad X \stackrel{?}{=} Z F^{-z_m} \hat{e}(g, g)^{z_u}.$$
  If both hold, OC computes $M = DC^{x s_0} \widehat{D}^x = \hat{e}(g, g)^{y \sum_{i \in U \setminus \widehat{U}} m_i}$, and recovers the discrete log of M base Y using Pollard's lambda method [25] to obtain $m = \sum_{i \in U \setminus \widehat{U}} m_i$, which is the sum of the correct power consumption of residential users in U. Otherwise, OC also uses the divide-and-conquer approach to find the corrupted individual report.

B. The Enhanced DiPrism

To resist differential attacks, we propose the enhanced DiPrism with additional privacy preservation. In the enhanced DiPrism, the noise is added to the individual consumption and encrypted by the user, such that SO can only learn the sum of electricity measurements with noise. In addition, the noise is generated from a distributed approach by means of the infinite divisibility of Laplace distribution in Section III. In case that some smart meters fail to report the correct measurements, GW can complement the noise for malfunctioning smart meters. As a result, through the collaboration of smart meters and the gateway, ϵ-*differential privacy* can be achieved.

1) System Initialization: The TA bootstraps the whole system, the operation center, the gateway and the users U setup following the same procedures as those in the basic DiPrism.

2) Report Generation: To report the real-time electricity consumption m_i at $t_l \in \mathcal{T}$, U_i generates the consumption report P_i as follows:

- Use the secret values (s_i, k) to compute $r_l = f(k, t_l)$, $C_l = g^{r_l}$, $\alpha_i = G_{i,t_l}(N, \lambda) - G'_{i,t_l}(N, \lambda)$, $D_i = g^{m_i + \alpha_i} h^{r_l s_i}$, where $G_{i,t_l}(N, \lambda)$ and $G_{i,t_l}(N, \lambda)$ are two random values independently sampled from the Gamma distribution. Thus, $G_{i,t_l}(N, \lambda)$ and $G_{i,t_l}(N, \lambda)$ are i.i.d random variables having Gamma distribution with pdf

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2673843, IEEE Transactions on Smart Grid

$g(x, N, \lambda)$, where $x \geq 0$, and N is the number of users in the residential area;

- Randomly pick $u_i \in \mathbb{Z}_p^*$ to compute $Z_i = W_{i,m_i}^{u_i}$;
- Pick random values $\rho_{i1}, \rho_{i2}, \rho_{i3}, \rho_{i4} \in \mathbb{Z}_p^*$ and compute $T_i = g^{\rho_{i1}+\rho_{i4}} h^{r_l \rho_{i2}}$, $X_i = \hat{e}(Z_i, g)^{-\rho_{i1}} \hat{e}(g, g)^{\rho_{i3}}$;
- Compute $c_i = H(C_l, D_i, Z_i, t_l)$, $z_{m_i} = \rho_{i1} - c_i m_i$, $z_{s_i} = \rho_{i2} - c_i s_i$, $z_{u_i} = \rho_{i3} - c_i u_i$, $z_{\alpha_i} = \rho_{i3} - c_i \alpha_i$;
- Set $P_i = (SM_i, C_l, D_i, Z_i, T_i, X_i, c_i, z_{m_i}, z_{s_i}, z_{u_i}, z_{\alpha_i})$.

Finally, U_i sends P_i to GW.

3) Report Aggregation: Upon receiving $P_i \in \{P_1, P_2, \cdots, P_n\}$ from users in U, GW firstly checks whether all the measurements in consumption reports are valid by verifying whether the following equations are satisfied for each report P_i:

$$\hat{e}((T_i D_i^{-c_i} g^{-z_{m_i} - z_{\alpha_i}})^{z_{s_i}^{-1}}, g) \stackrel{?}{=} \hat{e}(C_l, h), \tag{7}$$

$$X_i \stackrel{?}{=} \hat{e}(Z_i, V_i)^{c_i} \hat{e}(Z_i, g)^{-z_{m_i}} \hat{e}(g, g)^{z_{u_i}}. \tag{8}$$

If either of them does not hold for $P_{\hat{i}}$, GW removes the report $P_{\hat{i}}$ and aggregates the remaining individual reports into P. Otherwise, GW aggregates all the received reports to generate P. The procedures to generate P are described as follows:

- If all N smart meters work correctly, i.e., $n = N$, for each report P_i, GW generates $(C_l, C, D, \widetilde{D}, Z, F, T, X, z_m, z_s, z_u)$ following the same procedures as those in the basic DiPrism and calculates $z_\alpha = \sum_{i=1}^{N} z_{\alpha_i}$. GW sets $P = (GW, C_l, C, D, \widetilde{D}, Z, F, T, X, z_m, z_s, z_u, z_\alpha)$.
- If the smart meters of some users $\widehat{U} \subset U$ do not work, GW uses its secret key y to compute

$$\begin{aligned} &\beta = \textstyle\sum_{\hat{i} \in \widehat{U}} G_{i,t_l}(N, \lambda) - G'_{i,t_l}(N, \lambda), \\ &C = \hat{e}(g^y, C_l), \\ &D = \textstyle\prod_{i \in U \setminus \widehat{U}} \hat{e}(g^y, D_i) \hat{e}(g^y, g^\beta), \\ &\widehat{D} = \textstyle\prod_{\hat{i} \in \widehat{U}} \hat{e}(C_l^y, S_{\hat{i}}). \end{aligned}$$

  In addition, GW generates $(\widetilde{D}, Z, F, T, X, z_m, z_s, z_u)$ following the same procedures as those in basic DiPrism and calculates $z_\alpha = \sum_{i \in U \setminus \widehat{U}} z_{\alpha_i}$. GW sets $P = (GW, C_l, C, D, \widehat{D}, \widetilde{D}, Z, F, T, X, z_m, z_s, z_u, z_\alpha)$.

Finally, GW forwards P to OC.

4) Report Reading: Upon receiving P from GW, OC performs the following operations to recover $\sum_{i=1}^{n} m_i + Lap(\lambda)$ as follows:

- If all N smart meters work correctly, OC verifies the equations $T \stackrel{?}{=} \widetilde{D} g^{z_m + z_\alpha} C_l^{x z_s}$ and $X \stackrel{?}{=} Z F^{-z_m} \hat{e}(g, g)^{z_u}$. If both hold, OC computes $M = D C^{x s_0}$ and uses Pollard's lambda method [25] to recover the discrete log of M base Y, that is, $m = \sum_{i=1}^{N} m_i + Lap(\lambda)$.
- If the smart meters of some users $\widehat{U} \subset U$ do not work, OC verifies the equations $T \stackrel{?}{=} \widetilde{D} g^{z_m + z_\alpha} C_l^{x z_s}$ and $X \stackrel{?}{=} Z F^{-z_m} \hat{e}(g, g)^{z_u}$. If both hold, OC computes $M = D C^{x s_0} \widehat{D}^x$ and uses Pollard's lambda method [25] to recover the discrete log of M base Y, that is, $m = \sum_{i \in U \setminus \widehat{U}} m_i + Lap(\lambda)$.

V. Security Discussion

In this section, we demonstrate that our schemes can achieve all the security goals defined in II-C, namely, privacy preservation, differential privacy, fault tolerance and range-based filtering.

Privacy Preservation. In basic DiPrism, we utilize the Lifted ElGamal encryption scheme with distributed random secret keys to protect the individual consumption of users. Since Lifted ElGamal encryption is key-less reproducible [26], the re-use of r_l among users in U does not affect the confidentiality of m_i. Because the Lifted ElGamal encryption scheme [22] is semantic secure against chosen plaintext attacks, an adversary $\mathcal{A}$ is not able to recover m_i or learn any knowledge about m_i without the private key of OC. Similarly, the ciphertexts of individual power measurements are aggregated to be (C, D), which has the same format as the individual report (C_l, D_i). Since (C, D) is still a ciphertext of Lifted ElGamal encryption in $\mathbb{G}_1$, $\mathcal{A}$ cannot learn any information about the individual measurements and the sum of all users' power consumption.

Differential Privacy. In DiPrism, although the user's power usage data is encrypted to prevent privacy disclosure, $\mathcal{A}$ still can launch differential attacks to threaten user's privacy, if it obtains the aggregated data of two adjacent data sets. To prevent differential attacks, in the enhanced DiPrism, we allow the normal smart meters and gateways to add appropriate Laplace noise collaboratively into the electricity consumption for achieving *ϵ-differential privacy* [23]. Specifically, through sampling from Gamma distribution, each smart meter integrates $G(N, \lambda) - G'(N, \lambda)$ into the individual measurement m_i before reporting and the sum of N random variables with Gamma distribution is Laplace distribution. $\sum_{i=1}^{N} m_i + Lap(\lambda)$ or $\sum_{i \in U \setminus \widehat{U}} m_i + Lap(\lambda)$ is calculated by OC in *Report Reading* phase. Therefore, the difference of the aggregated adjacent data sets does not expose the individual consumption to the operation center.

Fault Tolerance. We design a new approach to realize fault tolerance of smart meters' malfunction. Even if some smart meters $\widehat{U} \subset U$ do not report their electricity measurements correctly, OC still can use the secret key (k, s_0) to decrypt $(C, D, \widehat{D})$ and recover $m = \sum_{i \in U \setminus \widehat{U}} m_i$. Specifically, if only the aggregated ciphertext of electricity consumption from normal smart meters, i.e., $D = \prod_{i \in U \setminus \widehat{U}} \hat{e}(g^y, D_i) = \hat{e}(g^y, g^{\sum_{i \in U \setminus \widehat{U}} m_i} h^{r_l \sum_{i \in U \setminus \widehat{U}} s_i})$ is provided, OC cannot recover $\sum_{i \in U \setminus \widehat{U}} m_i$ since it does not know the value of $\sum_{i \in U \setminus \widehat{U}} s_i$. In order to give the decryption capability to OC, GW generates $\widehat{D} = \prod_{\hat{i} \in \widehat{U}} \hat{e}(C_l^y, S_{\hat{i}})$ for OC, which can be employed to aggregate with (C, D) for making $s_0 + s_1 + \cdots + s_N = 0 \mod p$ hold. Therefore, OC can utilize its secret key (k, s_0) to obtain $\sum_{i \in U \setminus \widehat{U}} m_i$, even some smart meters do not work correctly or fault measurements are eliminated.

Range-based Filtering. We utilize the range proof [24] to determine whether the individual consumption falls in the predicted or preset range. If a measurement m_i is out of the range $[\mathcal{W}_i, \mathcal{W}'_i]$, we deem that the measurement is corrupted. Specifically, each user U_i generates a proof $(Z_i, T_i, X_i, c_i, z_{m_i}, z_{s_i}, z_{u_i})$ to prove that $m_i \in [\mathcal{W}_i, \mathcal{W}'_i]$

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2673843, IEEE Transactions on Smart Grid

without disclosing the detailed measurement m_i to the adversary. GW checks the proof to determine whether there is a measurement m_i that does not fall in the preset range $[\mathcal{W}_i, \mathcal{W}_i']$ and discards the corrupted measurement. Then, GW aggregates the valid proofs from users in U to obtain $(Z, F, T, X, z_m, z_s, z_u)$, which can prove that the sum of measurements $(m_1, \cdots, m_n)$ are in the preset range for all the individual reports. Now we show that the soundness of range proof follows the extraction property of zero-knowledge proof and BBS signature [27]. The extraction property implies that if there is a user U_i^* can convince OC with a non-negligible probability ϵ, there exists an extractor, which interacts with U_i^* and outputs a witness (m_i, s_i, u_i) with a probability $ploy(\epsilon)$. Furthermore, if we assume the extractor's input consists of two transcripts, i.e., $(Z_i, T_i, X_i, c_i, z_{m_i}, z_{s_i}, z_{u_i})$ and $(Z_i, T_i, X_i, c_i', z_{m_i}', z_{s_i}', z_{u_i}')$, we can obtain the witness as $m_i = \frac{z_{m_i} - z_{m_i}'}{c_i' - c_i}$, $s_i = \frac{z_{s_i} - z_{s_i}'}{c_i' - c_i}$, $u_i = \frac{z_{u_i} - z_{u_i}'}{c_i' - c_i}$. It is obvious that the extractor succeeds if $c_i' - c_i$ is invertible in $\mathbb{Z}_p^*$. If $m_i \notin [\mathcal{W}_i, \mathcal{W}_i']$, U_i^* cannot generates a range proof to convince OC that $m_i \in [\mathcal{W}_i, \mathcal{W}_i']$ with a non-negligible probability $poly(\epsilon)$, which is the probability that BBS signature is broken under chosen-message attacks.

VI. Performance Evaluation

In this section, we evaluate the performance of DiPrism in terms of computational and communication overhead.

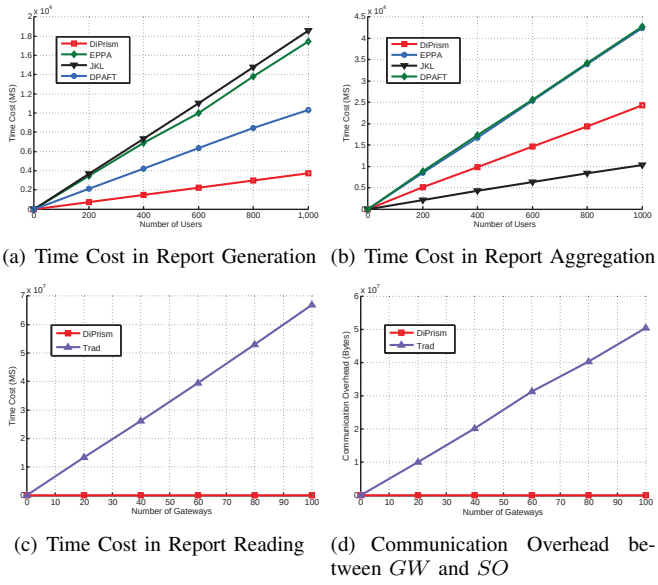
A. Computational Overhead

We count the number of the time-consuming cryptographic operations to demonstrate the computational overhead of the enhanced DiPrism. Exp, Exp_1, BP, Mul and Mul_1 denote the operations of exponentiation in $\mathbb{G}$, exponentiation in $\mathbb{G}_1$, bilinear pairing, multiplication in $\mathbb{G}$ and multiplication in $\mathbb{G}_1$. In *System Initialization* phase, OC initializes the system by executing $(N + |\mathcal{W}_1| + \cdots + |\mathcal{W}_N| + 1)Exp$ operations for N users in the residential area, where $|\mathcal{W}_i|$ denotes the number of possible discrete readings in range $[\mathcal{W}_i, \mathcal{W}_i']$ of U_i. GW needs to perform $BP + Exp$ operations to compute $Y = \hat{e}(g, g)^y$ and each user U_i is required to run Exp operation to generate S_i. In *Report Generation* phase, each user U_i generates P_i by running $6Exp + 2BP + 2Exp_1 + 2Mul + Mul_1$ operations. To further reduce the computational burden for U_i, OC can pre-compute $\hat{g} = \hat{e}(g, g)$ and $\widehat{W}_{i,w_{ij}} = \hat{e}(W_{i,w_{ij}}, g)$ for each $w_{ij} \in [\mathcal{W}_i, \mathcal{W}_i']$, such that X_i can be calculated as $X_i = \widehat{W}_{i,m_i}^{-u_i \rho_{i1}} \hat{g}^{\rho_{i3}}$. Therefore, the computation overhead of U_i can be reduced to $6Exp + 2Exp_1 + 2Mul + Mul_1$ and no bilinear pairing is required for U_i, which is the most time-consuming operation in these five operations. In *Report Aggregation* phase, GW receives n' individual reports and aggregates n reports to generate P, where n is the number of uncorrupted measurements and $N - n$ is the number of malfunctioning or compromised smart meters. To aggregate the reports, GW should perform $(3n' + n + 5)BP + (5n' + 2n + 2)Exp + n'Exp_1 + (N + 3n - 5)Mul + (2n - 2)Mul_1$ operations. Nevertheless, the verification of range proofs can be only performed by OC since electricity theft does not frequently happen. If $(\widetilde{D}, Z, F, T, X, z_m, z_s, z_u, z_\alpha)$ does not succeed, OC can further check the individual range proof. In this way, GW is only required to perform $(n + 4)BP + (2n + 2)Exp + (N + 3n - 5)Mul + (2n - 2)Mul_1$ operations to aggregate the individual reports. In *Report Reading* phase, OC verifies whether the power consumption is in the desired range and recovers the sum of measurements by executing $2Exp + 4Exp_1 + 2Mul + 4Mul_1$ operations.

We conduct an experiment on a notebook with Intel Core i5-4200U CPU @ 2.29GHz and 4.00GB memory. We use the MIRACL library to implement number-theoretic based methods of cryptography. The parameter p is approximately 160 bits and the elliptic curve is defined as $y = x^3 + 1$ over $\mathbb{F}_q$, where q is 512 bits. We simulate a smart metering scenario where an operation center manages the power consumption of 100,000 smart meters with 100 gateways. The operation center receives the aggregated reports from 100 gateways and each gateway controls 1000 smart meters. It costs 7.422ms for each user U_i to generate P_i and sends P_i to GW. After receiving n (i.e. n=1000) reports from users, GW executes 24124ms to aggregate the reports and sends the result to OC. OC receives 100 aggregated reports and spends 64324ms to recover the sum of the power consumption from all aggregated reports.

We also compare the computational overhead with some existing schemes, e.g., EPPA [10], DPAFT [18] and JKL [28]. Since EPPA [10], DPAFT [18] and JKL [28] do not support range-based filtering, we remove the computations of range-based filtering in DiPrism and compare the computational overhead on data aggregation. Firstly, due to the limitation of computational capability of smart meters, the computational overhead in *Report Generation* phase should be relatively low. We compare the enhance DiPrism with EPPA [10], DPAFT [18] and JKL [28] in terms of the execution time of encryption algorithm in *Report Generation* phase. As shown in Fig 3(a), our scheme costs less time than EPPA, JKL and DPARF for n smart meters to generate the ciphertexts of measurements, since the Lifted ElGamal scheme is more efficient than the Paillier encryption and BGN encryption used in EPPA, JKL and DPARF, respectively. Since each smart meter performs the encryption algorithm in *Report Generation* phase individually, the time cost for each smart meter is relatively low. Each gateway would receive 1000 individual reports from 1000 smart meters. Fig 3(b) shows the comparison results of computational overhead on GW when the number of smart meters increases from 1 to 1000. Our DiPrism is more efficient than EPPA and DPAFT because GW in DiPrism is required to perform less bilinear pairing than that in EPPA and DPAFT, in which two bilinear pairings are performed to check the availability of each user's signature. In *Report Reading* phase, OC prefers to execute the decryption algorithm for each aggregated report to obtain the sum of power consumption, rather than decrypting each report and then adding the recovered individual consumption together. Therefore, with data aggregation, computational time of OC can be significantly reduced. In Fig. 3(c), OC maximally receives 100 aggregated reports from gateways and decrypts each aggregated report to obtain the sum of power consumption of 1000 smart meters in

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TSG.2017.2673843, IEEE Transactions on Smart Grid



(a) Time Cost in Report Generation (b) Time Cost in Report Aggregation

(c) Time Cost in Report Reading (d) Communication Overhead between GW and SO

Fig. 3. Performance Evaluation

its coverage area. DiPrism is more efficient than the traditional method (Trad), in which OC reads 100,000 individual reports one by one.

B. Communication Overhead

To report the real-time power consumption to the operation center, each user U_i has to send $P_i = (SM_i, C_l, D_i, Z_i, T_i, X_i, c_i, z_{m_i}, z_{s_i}, z_{u_i}, z_{\alpha_i})$ to GW at a reporting time point $t_l \in \mathcal{T}$. The binary length of P_i is 504 bytes, if the identifier of SM_i is assumed to be 20 bytes. Thus, GW can receive $504n$ bytes if there are n smart meters reporting the measurements correctly. To reduce the communication overhead on the channel between GW and SO, GW aggregates n individual reports into a compact report $P = (GW, C_l, C, D, \widehat{D}, \widetilde{D}, Z, F, T, X, z_m, z_s, z_u, z_\alpha)$, which is a bit-string of 996 bytes, if the identifier of GW is 20 bytes. Therefore, with data aggregation, the length of transmission message from GW to OC can decrease from $504n$ bytes to 996 bytes, a constant length that does not depend on the number of workable smart meters. In Fig. 3(d), we demonstrate the efficiency of reports transmission between GW and OC. Each gateway would send the aggregated report generated from 1000 individual reports to OC and OC could maximally receive 100 aggregated reports. In DiPrism, each GW is only required to send 996 bytes to OC, which is shorter than that in the traditional method (Trad), where GW forwards all 1000 individual reports to OC without aggregation.

C. Implementation

We give a detailed description to show how to implement our DiPrism scheme. The implementation is quite important because a secure cryptographic scheme cannot provide sufficient security to systems if it is not implemented correctly. To implement DiPrism, public-key infrastructure (PKI) is needed to facilitate secure information transmission, which has been widely used in our daily life. In PKI, a Certificate Authority (CA) is responsible to issue and sign public certificates for both gateways and the operation center. The TA in DiPrism can be the CA or a trusted government institution, who bootstraps the smart metering system and assigns secret key materials to the residential users. All the secret keys should be kept securely and prevented from being lost, stolen or forged. They can be encrypted using passwords and stored on secure disks. The trusted platform module (TMP) can be used to generate secure keys and ensure their security. These keys are required to be at least 160 bits. The elliptic curve can be defined as $y = x^3 + 1$ over $\mathbb{F}_q$, where q is 512 bits. Moreover, there are several programming languages (e.g., C, C++, Java, Python) and cryptographic libraries (e.g., Miracl, PBC, NTL) to be used to implement DiPrism. Based on the PKI, proper parameters and the description in section IV, each algorithm in DiPrism can be implemented on the corresponding entities. For example, the report generation algorithm is deployed on smart meters, the report aggregation algorithm should be implemented on gateways and the report reading algorithm is executed on the operation center. In short, to implement DiPrism, just a short and simple program is required to be inserted into a proper position of the original code in each entity for data processing.

VII. Related Work

Privacy-preserving data aggregation is a critical technique in smart grid, which achieves crucial data collection with user's privacy preservation in an efficient manner. In past decade, many privacy-preserving data aggregation schemes [10], [11], [16], [17] have been proposed. Lu et al. [10] integrated super-increasing sequence with Paillier homomorphic encryption to design a privacy-preserving multi-dimensional data aggregation scheme, which guarantees data confidentiality, authentication and multi-dimensional data compression in a semi-honest model. Li et al. [11] adopted homomorphic encryption and aggregation tree to achieve distributed incremental data aggregation without exposing users' privacy. To prevent the failure of a single gateway, Garcia and Jacobs [17] combined homomorphic encryption and data sharing to propose an interactive energy metering scheme, in which the individual measurement is sliced into several shares and the gateways collaboratively aggregate all the received shares for the operation center. Consequently, to protect user's privacy against internal attackers, Fan et al. [16] designed a privacy-enhanced data aggregation scheme to against internal attackers, such as the operation center, by injecting blinding factors into the consumption data. Unfortunately, Bao and Lu [29] found that Fan et al.'s scheme cannot achieve their design goals since the user's private key can be easily derived from public parameters. As a result, the integrity of consumption data are violated. Ni et al. [30] utilized homomorphic signature and homomorphic encryption to propose a security-enhanced smart metering scheme that prevents malicious gateways from injecting false data into consumption reports. Jo et al. [28] utilized Paillier homomorphic encryption to design a lightweight privacy-preserving metering protocol and proposed a new distributed authentication method to improve the efficiency

of message authentication. Abdallah and Shen [31] proposed a privacy-preserving data aggregation scheme by leveraging lightweight NTRU cryptosystem to protect user's privacy and integrity of power consumption with low computational and communication overhead.

In addition, some works [18], [32] extended privacy-preserving data aggregation schemes to support various appealing properties, including fault tolerance, error detection and differential privacy. Jia et al. [33] formally defined a human-factor-aware differential aggregation attack and proposed a privacy-preserving smart metering scheme to achieve efficient data aggregation without leaking the individual power consumption and thwart differential aggregation attacks. Bao and Lu [18] proposed a privacy-preserving data aggregation scheme with differential privacy and fault tolerance. A novel key agreement protocol is designed to support fault tolerance of malfunctioning smart meters flexibly, and the BGN cryptosystem is employed to achieve data aggregation with user's privacy protection. Furthermore, differential privacy technique is integrated to resist differential attacks. However, this scheme requires two-way interactions to distribute random values and collect power consumption in each time period of reading reporting. Shi et al. [32] introduced a diverse grouping-based aggregation protocol by considering the lifetime of smart meters as exponential distribution and utilizing differential privacy to achieve grouping-based private stream aggregation and efficient error detection. Different from the existing works, we propose a privacy-preserving smart metering scheme in smart grid using Lifted ElGamal encryption and differential privacy, which supports data aggregation, user's privacy preservation, fault tolerance, range-based filtering and resistant to differential attacks simultaneously.

VIII. Conclusions

In this paper, we have proposed a differentially private data aggregation scheme with fault tolerance and range-based filtering in smart grid. This scheme is secure under a more challenging threat model, in which the operation center and gateways are semi-honest and the users may launch electricity theft and false data injecting attacks. DiPrism can prevent the curious operation center from acquiring the individual electricity consumption, and ensure the measurements are in the acceptable range without disclosing the exact readings. In addition, we have extended Lifted ElGamal encryption to support power consumption collection even some smart meters fail to report measurements and utilized differential privacy technique to resist differential attacks. In the further work, we will design a privacy-preserving demand response scheme with error detection for two-way communications in smart grid.

References

[1] S. Maharjan, Q. Zhu, Y. Zhang, S. Gjessing, and T. Basar, "Demand response management in the smart grid in a large population regime," *IEEE Trans. Smart Grid*, vol. 7, no. 1, pp. 189–199, 2016.

[2] X. Deng, L. He, C. Zhu, M. Dong, K. Ota, and Lin Cai, "QoS-aware and load-balance routing for IEEE 802.11s based neighborhood area network in smart grid," *Wirel. Pers. Commun.*, vol. 89, no. 4, pp. 1065–1088, 2014.

[3] Y. Zhang, S. Gjessing, H. Liu, H. Ning, L. Yang, and M. Guizani, "Securing vehicle-to-grid communications in the smart grid," *IEEE Wirel. Commun.*, vol.20, no.6, pp. 66–73, 2013.

[4] H. Liu, H. Ning, Y. Zhang, and L. Yang, "Aggregated-proofs based privacy-preserving authentication for V2G networks in smart grid," *IEEE Trans. Smart Grid*, vol.3, no.4, pp. 1722–1733, 2012.

[5] X. Li, X. Liang, R. Lu, X. Shen, X. Lin, and H. Zhu, "Securing smart grid: cyber attacks, countermeasures, and challenges," *IEEE Commun. Mag.*, vol. 58, no. 8, pp. 38–45, 2012.

[6] R. Lu, "Privacy-enhancing aggregation techniques for smart grid communications," *Springer– Wireless Networks series*, 2016.

[7] BBC News, "Hackers caused power cut in western Ukraine," 2016.

[8] Electronic Privacy Information Center, "The smart grid and privacy," 2015.

[9] Z. Erkin, J.R. Troncoso-Pastoriza, R.L. Lagendijk, and F. Perez-Gonzalez, "Privacy-preserving data aggregation in smart metering systems: an overview," *IEEE Signal Process. Mag.*, vol. 30, no. 2, pp. 75–86, 2013.

[10] 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, 2012.

[11] F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption," in *Proc. of SmartGridComm*, 2010, pp. 327–332.

[12] S. McLaughlin, B. Holbert, A. Fawaz, R. Berthier, and S. Zonouz, "A multi-sensor energy theft detection framework for advanced metering infrastructures," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 7, pp. 1319–1330, 2013.

[13] P. Kelly-Detwiler, "Electricity theft: a bigger issue than you think," http://www.forbes.com/sites/peterdetwiler/2013/04/23/electricity-theft-a-bigger-issue-than-you-think/#242147fc72ef, 2013.

[14] Krebsonsecurity, "FBI: smart meter hacks likely to spread," http:// krebson security.com/2012/04/fbi-smart-meter-hacks-likely-to-spread/, 2012.

[15] ENISA, "Smart grid security: recommendations for Europe and member states," 2012.

[16] C.-I Fan, S.-Y. Huang, and Y.-L. Lai, "Privacy-enhanced data aggregation scheme against internal attackers in smart grid," *IEEE Trans. Ind. Inform.*, vol. 10, no. 1, pp. 666–675, 2014.

[17] F.D. Garcia and B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," in *Proc. of STM*, 2011, pp. 226–238.

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

[19] J. Zhao, T. Jung, Y. Wang, and X. Li, "Achieving differential privacy of data disclosure in the smart grid," in *Proc. of IEEE INFOCOM*, 2014, pp. 504–512.

[20] T.-H. Chan, E. Shi, and D. Song, "Privacy-preserving stream aggregation with fault tolerance," in *Proc. of FC*, 2012, pp. 200–214.

[21] J. Wu, M. Dong, K. Ota, Z. Zhou, and B. Duan, "Towards fault-tolerant fine-grained data access control for smart grid," *Wirel. Pers. Commun.*, vol. 75, no. 3, pp. 1787–1808, 2014.

[22] J. Liu, N. Asokan, and B. Pinkas, "Secure deduplication of encrypted data without additional independent servers," in *Proc. of ACM CCS*, 2015, pp. 874–885.

[23] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in privacy data analysis," in *Proc. of TCC*, 2006, pp. 265–284.

[24] J. Camenisch, R. Chaabouni, and A. Shelat, "Efficient protocols for set menbership and range proofs," in *Proc. of ASIACRYPT*, 2008, pp 234–252.

[25] J.M. Pollard, "Kangaroos, monopoly and discrete logarithms," *J. Cryptology*, vol. 13, no. 4, pp 437–447, 2000.

[26] B. Libert, K.G. Paterson, and E.A. Quaglia, "Anonymous broadcast encryption: adaptive security and efficient constructions in the standard model," in *Proc. of PKC*, 2012, pp. 206–224.

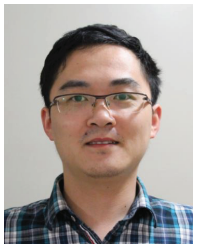
[27] D. Boneh and X. Boyen, "Efficient selective-ID identity-based encryption without random oracles," in *Proc. of EUROCRYPT*, 2004, pp. 223–238.

[28] H.J. Jo, I.S. Kim, and D.H. Lee, "Efficient and privacy-preserving metering protocols for smart grid systems," *IEEE Trans. Smart Grid*, vol. 7, no. 3. pp. 1732–1742, 2015.

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

[30] J. Ni, K. Alharbi, X. Lin, and X. Shen, "Security-enhanced data aggregation against malicious gateways in smart Grid," in *Proc. of IEEE GLOBECOM*, 2015, pp. 1–6.

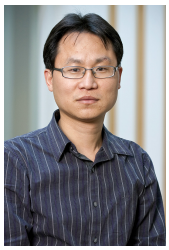
[31] A. Abdallah and X. Shen, "Lightweight security and privacy preserving scheme for smart grid customer-side networks," *IEEE Trans. on Smart Grid*, to appear.
[32] Z. Shi, R. Sun, R. Lu, L Chen, J. Chen, and X. Shen, "Diverse grouping based aggregation protocol with error detection for smart grid communications," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2856–2867, 2015.
[33] W. Jia, H. Zhu, Z. Cao, X. Dong, and C. Xiao, "Human-factor-aware privacy-preserving aggregation in smart grid," *IEEE Syst. J.*, vol. 8, no. 2, pp. 598–607, 2014.
[34] 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, 2015.



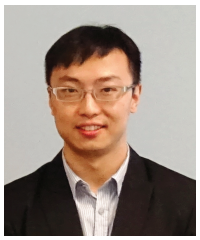
Jianbing Ni (S'16) received the B.E. degree and the M.S. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2011 and 2014, respectively. He is currently working toward the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. His research interests are applied cryptography and information security, with current focus on cloud computing, e-healthcare, smart grid and big data.

Kuan Zhang (S'13-M'16) received his B.Sc. degree in communications engineering and M.Sc. degree in computer science from Northeastern University, China, in 2009 and 2011, respectively. He received his Ph.D. degree in electrical and computer engineering from the University of Waterloo, Canada, in 2016. Currently, he is a postdoctoral fellow with the Department of Electrical and Computer Engineering, University of Waterloo. His research interests include security and privacy for mobile social networks, e-healthcare systems, and cloud computing.

Khalid Alharbi received the B.Sc. degree in Mathematics, Saudi Arabia, in 1999, the Master of Information Technology Security (MITS) in 2012 and the Ph.D. degree in Computer Science in 2017 at the University of Ontario Institute of Technology (UOIT), Canada. He is an instructor at Northern Border University, Saudi Arabia. His research interests include applied cryptography, and security and privacy issues in web applications, cloud computing, mobile social networks, and smart grid.



Xiaodong Lin (M'09-SM'12-F'17) received the Ph.D. degree in information engineering from Beijing University of Posts and Telecommunications, Beijing, China, and the Ph.D. degree in electrical and computer engineering (with Outstanding Achievement in Graduate Studies Award) from the University of Waterloo, ON, Canada. He is currently an Associate Professor of information security with the Faculty of Business and Information Technology, University of Ontario Institute of Technology (UOIT), Canada. His research interests include wireless network security, applied cryptography, computer forensics, software security, and wireless networking and mobile computing.



Ning Zhang (S'12-M'16) received the Ph.D degree from the University of Waterloo in 2015. He received his B.Sc. degree from Beijing Jiaotong University and the M.Sc. degree from Beijing University of Posts and Telecommunications, Beijing, China, in 2007 and 2010, respectively. From May 2015 to Apr. 2016, he was a postdoc research fellow at BBCR lab in the University of Waterloo. Since May 2016, he has been a postdoc research fellow at Wireless Computing lab at the University of Toronto. He is now an associate editor of International Journal of Vehicle Information and Communication Systems and a guest editor of Mobile Information System. He is the recipient of the Best Paper Award at IEEE Globecom 2014 and IEEE WCSP 2015. His current research interests include sensor networks, next generation wireless networks, software defined networking, green communication, and physical layer security.

Xuemin (Sherman) Shen (M'97-SM'02-F'09) received the B.Sc.(1982) degree from Dalian Maritime University (China) and the M.Sc. (1987) and Ph.D. degrees (1990) from Rutgers University, New Jersey (USA), all in electrical engineering. Dr. Shen is a Professor and University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo, Canada. He is also the Associate Chair for Graduate Studies. His research focuses on resource management in interconnected wireless/wired networks, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks. Dr. Shen served as the Technical Program Committee Chair/Co-Chair for IEEE Globecom'16, Infocom'14, IEEE VTC'10 Fall, and Globecom'07, the Symposia Chair for IEEE ICC'10, the Tutorial Chair for IEEE VTC'11 Spring and IEEE ICC'08, the General Co-Chair for ACM Mobihoc'15, Chinacom'07 and QShine'06, the Chair for IEEE Communications Society Technical Committee on Wireless Communications, and P2P Communications and Networking. He also serves/served as the Editor-in-Chief for IEEE Internet of Things Journal, IEEE Network, Peer-to-Peer Networking and Application, and IET Communications; a Founding Area Editor for IEEE Transactions on Wireless Communications; an Associate Editor for IEEE Transactions on Vehicular Technology, Computer Networks, and ACM/Wireless Networks, etc. Dr. Shen received the Excellent Graduate Supervision Award in 2006, and the Outstanding Performance Award in 2004, 2007, 2010, and 2014 from the University of Waterloo, the Premier's Research Excellence Award (PREA) in 2003 from the Province of Ontario, Canada, and the Distinguished Performance Award in 2002 and 2007 from the Faculty of Engineering, University of Waterloo. Dr. Shen is a registered Professional Engineer of Ontario, Canada, an IEEE Fellow, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, a Royal Society of Canada Fellow, and a Distinguished Lecturer of IEEE Vehicular Technology Society and Communications Society.