PPP A: A Practical Privacy-Preserving Aggregation Scheme for Smart Grid Communications

Min Lu†, Zhiguo Shi†‡, Rongxing Lu‡, Ruixue Sun†, and Xuemin (Sherman) Shen‡
†Department of Information and Electronic Engineering, Zhejiang University, Hangzhou 310027, China
‡Dept. ECE, University of Waterloo, Canada; §School of EEE, Nanyang Technological University, Singapore

Email: {minlu, zhigz, sunrx}@zju.edu.cn; rxl@ntu.edu.sg; xshen@b crater. uwaterloo.ca

Abstract—Characterized by time-of-using pricing, better capacity, and usage planning, smart grid has attracted considerable attention in recent years. However, one of the research challenges in smart grid is the privacy issue due to too much sensitive and real-time user data involved. In this paper, we propose a practical privacy-preserving aggregation scheme, named PPPA, for secure smart grid communications. PPPA utilizes the lightweight cryptographic aggregation technique to achieve provable security guarantee, the differential privacy technique to achieve privacy preservation of each individual user, and the quad tree structure to achieve failure tolerance. For data communication from users to control center, data aggregation is conducted directly on ciphertexts at local gateway without decryption, and the aggregation results are reported by relays to the control center. Through extensive analysis, we demonstrate that PPPA can not only provide authentication and guarantee the data integrity, but also achieve strong privacy-preserving for each user, and thus it is feasible in practical smart grid scenarios.

Keywords—Smart Grid, Security, Privacy Preservation, Aggregation, Failure Tolerance

I. INTRODUCTION

Suffering from load imbalance and lack of effective real-time diagnosis, the current power grid is not always so stable, and may cause huge damages to our daily life. For example, the North America electrical blackout in August 2003 affected more than 100 power plants and paralyzed tens of millions of users [1]. Targeting such challenges, smart grid, a promising new generation of electric power system, has emerged [2], [3]. According to the U.S. Department of Energy (DOE)’s Modern Grid Initiative, a smart grid can integrate advanced sensing technologies, control methods and integrated communications into current electricity grid both at transmission and distribution levels [4], [5]. Since smart grid enables a distributed computing approach with potentials for self diagnosing/healing, reliable multi-user communication and fast real-time control, it has become a promising technology for our future power grid, as shown in Fig. 1.

Since electricity is difficult to be stored, the ideal case for a power grid is that the instantaneous power generation matches the power consumption by users, considering both the system efficiency and stability [6]. To achieve this, people try to integrate high-speed and two-way communication technologies in smart grid to establish a dynamic and interactive infrastructure with new energy management capabilities, such as advanced metering infrastructure (AMI) and demand response [7], [8]. Therefore, large number of smart meters, deployed at user sides, collect users’ electricity usage information and periodically report these rich and real-time data to the control center. However, such information could have unintended and negative consequences to users. For example, if the power consumption of a family is observed by a criminal, it is easy for him to know the habit of the family and burglarize their home when nobody is there [9]. Potential threats targeting user privacy would, without doubt, prevent smart grid from stepping into flourish. To combat this, many researches [9]–[11] on privacy-preserving data aggregation have been carried out for smart grid. However, these aggregation schemes mainly depend on some “ad-hoc” type techniques such as anonymizing or blind signing, which may induce terrible incidents like AOL data release in 2006 [12]. In addition, the computation cost and communication overhead have to be considered due to large scale of smart grid communication network. Therefore, how to achieve an efficient, privacy-preserving and failure-tolerant aggregation for smart grid still deserves further investigations.

To address the challenge, in this paper, we incorporate the differential privacy technique [13], [14] and propose a new practical privacy-preserving aggregation (PPPA) scheme for smart grid. Specifically, in the proposed PPPA scheme, the control center can monitor the users’ electricity usage situation in a real-time manner while achieving strong privacy of each individual user, by jointly using cryptographic technique and differential privacy technique. Compared with the binary protocol [15], the proposed PPPA aggregation scheme, due to its replacing binary tree with quad tree, is more efficient in terms of computation cost and communication overhead. In addition, the binary protocol does not address the practical challenges in smart grid scenarios, while our proposed...
PPP scheme is considered under the specific communication-oriented smart grid framework, which thus can achieve more practical applicability.

The remainder of this paper is organized as follows. In Section II, we introduce the system model, security model and design goal. In Section III, we propose our PPP scheme, followed by its security analysis and performance evaluation in Section IV and Section V, respectively. Related works are discussed in Section VI. Finally, we draw our conclusions in Section VII.

II. MODELS AND DESIGN GOALS

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

A. System Model

According to [7], the smart grid communication network can be simplified into the system model shown in Fig. 2. Specially, a typical residential area (RA) is comprised of several residential users $\mathcal{U} = \{U_1, U_2, \cdots, U_n\}$, and a local gateway (GW), a control center (CC) and several relays $\mathcal{R} = \{R_1, R_2, \cdots, R_m\}$.

![Fig. 2. System model under consideration](image)

- $CC$: CC is responsible for the monitor of electricity usage information in the residential area (RA) in a real-time manner, so as to make more efficient electricity use by various ways such as dynamic price.
- $GW$: GW mainly performs two functions: aggregation and relaying. The responsibility of aggregation component is to aggregate the electricity usage data of residential users into a compressed one, while the responsibility of relaying component is to help forward the aggregated reports to the nearest relay.
- $U_i$: $U_i \in \mathcal{U}$ is equipped with various smart appliances and smart meters ($SM_i$) to electronically record the real-time electricity usage data. For every time period, $U_i$ uploads the electricity usage report to the control center with the help of GW.
- $R_i$: $R_i \in \mathcal{R}$ is a relay node that performs traffic directing in smart grid communications, where the aggregated electricity usage report is typically forwarded from one relay to another until it reaches the control center.

B. Security Model

In our security model, we focus on how to guarantee the strong privacy of the electricity usage of each user in RA. We consider CC, GW and $R_i \in \mathcal{R}$ are trustable, and ensure that CC could monitor the electricity usage in RA in real-time manner. However, there exists an outside adversary $\mathcal{A}$, which will launch one or more of the following attacks simultaneously:

- **Communication attacks**: Eavesdrop, forge, modify or undue delay the electricity usage report in order to abuse smart grid;
- **Malware attacks**: Deploy undetectable malwares at control center and some users to steal the detailed electricity information;

Note that each user remains uncompromised independently with probability $\gamma$ in our model. In addition, the adversary $\mathcal{A}$ could also collect arbitrary information about a targeted user without compromising its key through ways such as publicly available datasets and personal knowledge.

C. Design Goal

With the aforementioned system model and security model, our goal is to develop a practical privacy-preserving aggregation scheme for secure smart grid communications. In particular, the following goals should be achieved.

- **The security and privacy requirements should be guaranteed in the proposed scheme.** As stated above, if the smart grid does not consider the security, the system may be abused and users’ privacy may be disclosed, and both will cause huge damage.
- **The proposed scheme should consider the practical applicability in smart grid communication architecture as much as possible.** A real-world smart grid is easy to suffer from user failures. In addition, smart meters usually have limited bandwidth and computational ability. Considering the huge number of users and long term development, the proposed scheme should also achieve computation-efficiency and communication-efficiency. Moreover, since data report is transmitted through GW and relays, the data flow path should be authenticated to prevent malicious attacks such as injecting falsified information during transmission.

III. THE PROPOSED PPPA SCHEME

In this section, we propose our PPPA scheme, which consists of five parts: System Initialization, User Report Generation, User Report Aggregation, Aggregation Report Relaying and Aggregation Report Reading. Before going into the details, we first review some building blocks, namely the notion of computational differential privacy and the Private Stream Aggregation (PSA) scheme proposed by Shi et. al [16], which serve as the basis of the proposed PPPA scheme.

A. Building Blocks

1) **Differential Privacy**: Differential privacy is a notion of privacy tailored to private data analysis, where the goal is to learn information about the population as a whole, while protecting the privacy of each individual [17]. In this work, we tailor computational differential privacy notion [15] to our design, as it is more applicable in smart grid scenario.
Definition 1 (Computational Differential Privacy): Let $\Delta$ be a small integer, and $x = \{x_1, x_2, \ldots, x_n\}$, $y = \{y_1, y_2, \ldots, y_n\}$ be two vectors of user electricity usage that differ only at position $i$, where each $x_i, y_i \in \{0, 1, \ldots, \Delta\}$ represents user $U_i$’s electricity usage data. Let $C_i$ denote the event that user $U_i$ remains uncompromised. A randomized protocol $\Pi$ preserves computational $(\epsilon, \delta)$-differential $(\epsilon, \delta > 0)$ privacy against a compromise process $C$ if there exists a negligible function $\eta : \mathbb{N} \rightarrow \mathbb{R}^+$ such that the following condition can be achieved:

$$
\Pr[c_i(\Pi(x, C)) = b] \leq e^\epsilon \times \Pr[c_i(\Pi(y, C)) = b] + \delta + \eta(\lambda)
$$

for all $\lambda \in \mathbb{N}$, for all $i \in [1 : n]$, for all probabilistic polynomial-time algorithms $A$, for all $x$ and $y$, and for any output $b \in \{0, 1\}$. Note that the probability is taken over the randomness of $A$, $\Pi$, and $C_i$.

Geometric distribution can serve as the randomized protocol $\Pi$ to achieve privacy-preserving data analysis. Usually, diluted geometric distribution is in lieu of geometric distribution for data utility consideration. Both of them are formally defined in [15] as follows.

Definition 2 (Geometric Distribution): Let $\alpha > 1$. We denote by $\text{Geom}(\alpha)$ the symmetric geometric distribution that takes integer values such that the probability function at $k$ is

$$
\frac{1}{\alpha} \cdot \alpha^{-\lfloor k \rfloor}.
$$

Definition 3 (Diluted Geometric Distribution): Let $0 < \beta \leq 1$, $\alpha > 1$. A random variable has $\beta$-diluted Geometric Distribution $\text{Geom}^{\beta}(\alpha)$ if with probability $\beta$ it is sampled from $\text{Geom}(\alpha)$, and with probability $(1 - \beta)$ it is set to 0.

2) Private Stream Aggregation Scheme: The Private Stream Aggregation (PSA) scheme [16] deals with the private-preserving aggregation issue in presence of an untrusted aggregator by adding appropriate noise in a distributed fashion. Concretely, the PSA scheme is compromised of three algorithms: key generation, encryption and decryption.

- Key Generation: Given the security parameter $\lambda$ and the number of users $n$, a trusted entity first outputs a cyclic group $G$ with the generator $g$ of the large prime order $p$ where $|p| = \lambda$. Then it chooses a secret key $sk_i \in \mathbb{Z}_p^*$ for each user $i$, as well as a secret key $sk_0 \in \mathbb{Z}_p^*$ for aggregator. The secret keys satisfy the condition $sk_0 + sk_1 + \cdots + sk_n = 0 \mod p$.

- Encryption: For time step $t$, each user $i$ computes $c_i = g^{vt_i + r_i} \cdot H(t)^{sk_i}$ as ciphertext, where $x_i$ denotes the user data (its space is small), $r_i$ denotes the random noise from some diluted geometric distribution, and $H$ denotes a cryptographic hash function $H() : \{0, 1\}^* \rightarrow G$.

- Decryption: Upon receiving ciphertexts $c_1, c_2, \ldots, c_n$, aggregator first computes $V = H(t)^{sk_0} \prod_{i=1}^{n} c_i$. It is not hard to have the following deductions: $V = H(t)^{sk_0 + sk_1 + \cdots + sk_n} \cdot g^{\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} r_i} = g^{\sum_{i=1}^{n} x_i + \sum_{i=1}^{n} r_i}$. Since the noisy sum of users’ data is a small value, it can be computationally easy to obtain by brute-force search.

B. Description of PPPA Scheme

We first give an informal description of PPPA under a special case where there are only 16 users in RA. Different sizes of user blocks are formed among users corresponding to all nodes in a quad tree with 16 leaf nodes, as shown in Fig. 3. Each block $B$ corresponds to a set of users. For example in Fig. 3(a), the second block in rank, $B_2^1$, corresponds to users $U_5, U_6, U_7, U_8$. Normally the sum estimate of user block $B_7^1$ can be used as estimate of the electricity usage of the residential area. However, when a user, for example, $U_1$ fails to respond, the estimate of $B_7^1$ can not be easily retrieved since PSA scheme [16] serves as the basic construction unit. In this situation, CC would find a set of disjoint blocks to uniquely cover the remaining functioning users. Thus, in case of $U_1$’s failure, CC could calculate the sum estimates of all the yellow blocks as RA’s overall electricity usage, as shown in Fig. 3(b).

![Fig. 3. A special case of the proposed PPPA scheme](image)

We then give a formal description of the proposed PPPA scheme. For ease of exposition, assume that $n$ is a power of 4. (If $n$ is not a power of 4, we can add some dummy leaves for a quad tree construction.) In this way, a quad tree with $n$ leaf nodes can be constructed. User blocks are formed in a similar way to the above case. In particular, given integers $k \geq 0$ and $j \geq 1$, the $j$th block of the rank $k, B^j_k$, corresponds to users $\{U_i | i = 4^k(j-1) + l, 1 \leq l \leq 4^k\}$. Then, it can be observed that the total number of user blocks is $\frac{4n - 1}{3}$.

1) System Initialization: Given security parameter $\lambda$ and privacy parameters $(\epsilon, \delta)$, a key management center (KMC) first outputs a cyclic group $G$ with generator $g$ of the large prime order $p$ where $|p| = \lambda$ and a secure cryptographic hash function $H() : \{0, 1\}^* \rightarrow G$ to CC, GW, relays and users. Then KMC generates and distributes secret keys among all parties for each block. In particular, suppose block $B$ corresponds to users $U_a, U_{a+1}, \ldots, U_b$ (Let $[b - a] + 1 = |B|$ be the block size), then KMC can choose $|B| + m + 2$ secret keys $sk_{a-1,B}, sk_{a,B}, \ldots, sk_{a+|B|-1,B}, \ldots, sk_{a+|B|+m,B} \in \mathbb{Z}_p^*$, where $sk_{a-1,B} + sk_{a,B} + \cdots + sk_{a+|B|+m,B} = 0 \mod p$. After that, the keys are distributed to all parties in a secure way.
As a result, CC obtains the key \( sk_{a-1,B} \) serving as the capacity to obtain RA’s overall electricity usage. \( U_i \) (i ∈ [a : b]) obtains the corresponding secret key \( sk_{i,B} \). GW obtains the secret key \( sk_{a+|B|,B} \). And \( R_i ∈ R \) obtains one secret key from \( sk_{a+|B|+1,B} \) to \( sk_{a+|B|+m,B} \) in a sequential way.

2) User Report Generation: For every time period \( t \), each user \( U_i ∈ U \) performs the following steps to report the electricity usage data \( x_i (x_i ∈ \{0, 1, \cdots, \Delta\} \) is a small integer) for each block \( B \) containing itself:

- \textbf{Step 1:} Choose a random noise \( r_i,B \) from the diluted geometric distribution \( Geom^\beta(\exp(\frac{\rho}{\beta})) \) ( \( \beta = \min\{\frac{1}{\ln \frac{1}{\rho}}, 1\} \) ) and compute

\[
\hat{x}_{i,B} = x_i + r_i,B
\]

- \textbf{Step 2:} Use the secret key \( sk_{i,B} \) to encrypt the perturbed data \( \hat{x}_{i,B} \)

\[
c_{i,B} = g^{\hat{x}_{i,B}} \cdot H(t)^{sk_{i,B}}
\]

where \( t \) can not only support time-series aggregation but also resist the potential replay attack.

- \textbf{Step 3:} Report the encrypted electricity usage data and block number \( c_{i,B}||B \) to the local gateway GW in the residential area.

3) User Report Aggregation: We consider it as a user block failure if the received user reports for this user block is not equal to its block size. Upon receiving a collection of user reports, GW would first check the failure situation of user blocks in rank 0 to detect malfunctioning users. Then all user blocks containing malfunctioning users will be treated as failure. Then GW searches from top to down to find a set of functioning blocks to uniquely cover the functioning users. For example, under the special case above, when \( U_1 \) fails to respond, GW is easy to detect the failures of user block \( B^0_1, B^1_1, B^2_1 \). After that, CC can easily find that the yellow blocks, shown in Fig. 3(b), can uniquely cover the remaining 15 users. For each users used as covering the functioning users, for example, considering a block \( B \) corresponding to users \( U_a, U_{a+1}, \cdots, U_b \), GW performs the following steps for privacy-preserving report aggregation:

- \textbf{Step 1:} Compute the aggregated and encrypted data \( C_{0,B} \)

\[
C_{0,B} = H(t)^{sk_{a+|B|,B}} \prod_{i=a}^{b} c_{i,B}
\]

- \textbf{Step 2:} Report the aggregated and encrypted data with block number \( C_{0,B}||B \) to relay \( R_1 \).

4) Aggregation Report Relaying: Upon receiving the aggregation report \( C_{i-1,B} \) for block \( B \) (Let block \( B \) corresponds to users \( U_a, U_{a+1}, \cdots, U_b \) ) from \( R_{i-1} \) (GW can be treated as \( R_0 \), \( R_i \) performs the following steps for relaying:

- \textbf{Step 1:} Add \( H(t)^{sk_{a+|B|+i,B}} \) to \( C_{i-1,B} \) as a path signature

\[
C_{i,B} = H(t)^{sk_{a+|B|+i,B}} \cdot C_{i-1,B}
\]

- \textbf{Step 2:} Report the result and block number \( C_{i,B}||B \) to \( R_{i+1} \) or the CC when \( R_i \) is \( R_m \).

5) Aggregation Report Reading: Upon receiving all the reports from relay \( R_m \), the CC will perform the following steps to estimate RA’s overall electricity usage:

- \textbf{Step 1:} For each block \( B \), compute

\[
V_B = H(t)^{sk_{a-1,B}} \cdot C_{m,B}
\]

- \textbf{Step 2:} For each \( V_B \) found in Step 1, it is computationally easy to retrieve the noisy sum by brute-force search. CC calculates all the retrieved noisy sum as the final estimate of RA’s electricity usage.

Note the accumulated random noise from dilute geometric distribution in the final result is negligible compared with the original precise sum, which means that our aggregation scheme will not block CC’s proper power allocation of the served RA.

IV. SECURITY ANALYSIS

In this section, following the security model discussed earlier, we analyze the security properties of the proposed PPPA scheme by investigating how the proposed PPPA scheme can achieve the authentication, data integrity and privacy of an individual user’s electricity usage report.

- **PPPA can provide authentication and guarantee the integrity of individual user’s electricity usage report.** In the proposed scheme, the key to successful decryption of the noisy sum for a block \( B \) corresponding to users \( U_a, U_{a+1}, \cdots, U_b \) is the condition that \( sk_{a-1,B} + sk_{a,B} + \cdots + sk_{a+|B|+m,B} = 0 \) mod \( p \). As a result, if a malicious adversary \( A \) forges a secret key to encrypt the illegal electricity usage data to participate in the report process, the above condition can not be fulfilled to obtain a meaningful result. Meanwhile, since electricity usage report is encrypted as \( c_{i,B} = g^{ri,B} \cdot H(t)^{sk_{i,B}} \), it is hard to obtain or undue delay electricity usage report. In addition, by requiring each relay to add a path signature, it can ensure that the data is from a legal path avoiding attacks such as injecting falsified data along transmission path.

- **The individual user’s report is privacy-preserving in the proposed PPPA scheme.** Following the computational differential privacy analysis [15], we prove that our proposed scheme can strongly preserve an uncompromised user’s privacy. Let \( x = \{x_a, x_{a+1}, \cdots, x_b\} \), \( y = \{y_a, y_{a+1}, \cdots, y_b\} \) be two vectors of users electricity usage that differ only at position \( i \), where each \( x_i, y_i \in \{0, 1, \cdots, \Delta\} \) represent user \( U_i \)’s electricity usage data. Let \( Sum(x) = \sum_{i} x_i + r_{i,B} = s \) and \( Sum(y) = \sum_{i} y_i + r_{i,B} = s \) be the summation of the noisy values of users in block \( B \), where \( r_{i,B} \) is a random noise from the diluted geometric distribution defined in our scheme. Let \( C_i \) denote the event that user \( U_i \) remains uncompromised. Let \( \rho \) be the event that \( \exists i \in [a : b], r_{i,B} \) is sampled from \( Geom(\alpha) \) ( \( \alpha = \exp(\frac{\rho}{\beta}) \)). Let \( \Pi \) be the
randomized protocol defined in our scheme. Note that the result of $\Pi$ largely depends on security parameter $\lambda$. Let $A$ be a probabilistic polynomial-time algorithm that outputs $b \in \{0, 1\}$ given the output of $\Pi$ and the information $C$ obtained from compromised users. Let $\eta$ be a negligible function $\eta : \mathbb{N} \rightarrow \mathbb{R}^+$. Observe that $Pr[\text{Sum}(x) = s|\rho] \leq \exp(\epsilon) \cdot Pr[\text{Sum}(y) = s|\rho]$ due to the following property regarding geometric distribution: $Pr[u + r = k] \leq e^\epsilon \cdot Pr[v + r = k]$ for any integer $k$, if two integers $u$ and $v$ have the relationship that $|u - v| \leq \Delta$, $r$ is a random variable sampled from $\text{Geom}(\exp(\frac{1}{\epsilon}))$ and $\epsilon > 0$. Observe that $Pr[C; \eta] = (1 - \beta) \cdot (1 - \beta\gamma)^{|B| - 1} \leq (1 - \beta\gamma)^{|B|} \leq \delta$. (The above result utilizes the inequality that $1 - f \leq \exp(-f)$ for all reals $f$.) Observe that for all $\lambda \in \mathbb{N}$, the following equation holds for some function $\eta$:

$$E_{\Pi, C; |\eta|}; \{Pr[A(\Pi(x, y), C) = b|\text{Sum}(x) = s, \rho, \Pi, C_i] \leq \eta(\lambda) + Pr[A(\Pi(y, C) = b|\text{Sum}(y) = s, \rho, \Pi, C_i] \}
$$

Then we have:

$$Pr[A(\Pi(x, y), C) = b|\text{Sum}(x) = s, \rho] \leq \eta(\lambda) + Pr[A(\Pi(y, C) = b|\text{Sum}(y) = s, \rho]$$

Furthermore, we have the following deduction:

$$Pr[A(\Pi(x, y), C) = b, \text{Sum}(x) = s|\rho] = Pr[A(\Pi(x, y), C) = b|\text{Sum}(x) = s, \rho] 
\times Pr[\text{Sum}(x) = s|\rho] \leq Pr[A(\Pi(y, C) = b|\text{Sum}(y) = s, \rho] 
\times Pr[\text{Sum}(x) = s|\rho] + \eta(\lambda) Pr[\text{Sum}(x) = s|\rho]$$

The following results can be obtained:

$$Pr[A(\Pi(x, y), C) = b] = Pr[A(\Pi(x, y), C) 
\leq \sum_s Pr[A(\Pi(x, y), C) = b, \text{Sum}(y) = s|\rho] \cdot Pr[\rho] + \delta
\leq \sum_s \{e^\epsilon \cdot Pr[A(\Pi(y, C) = b|\text{Sum}(y) = s|\rho] + \eta(\lambda) Pr[\text{Sum}(x) = s|\rho] \cdot Pr[\rho] 
+ \eta(\lambda) Pr[\text{Sum}(x) = s|\rho] \cdot Pr[\rho]\} + \delta
\leq e^\epsilon \cdot Pr[A(\Pi(y, C) = b|\rho] \cdot Pr[\rho] + \eta(\lambda) + \delta
\leq e^\epsilon \cdot Pr[A(\Pi(y, C) = b|\rho] + \delta + \eta(\lambda)$$

Therefore, each uncompromised user in Block $B$ in our proposed scheme is guaranteed the computational $(\epsilon, \delta)$-differential privacy against any compromise process defined in our security model.

V. PERFORMANCE EVALUATION

In this section, we analyze the performance of the proposed PPPA scheme from three different aspects: computation cost, communication overhead and failure tolerance, and show that our scheme is feasible in terms of performance.

Computation Cost: Smart meters, which are deployed at user sides, are usually characterized with limited computational ability. Therefore, computation should not be a burden at user side in smart grid. To evaluate the efficiency of the proposed PPPA scheme, we first compare it with Binary protocol [15] in terms of each user’s encryption. Since both of the two schemes utilize cryptographic construction in PSA Scheme [16] to encrypt data, it is reasonable to adopt the performance estimates in [16] which claims that each encryption takes about 0.6 ms on a modern computer using high-speed elliptic curves such as “curve25519”. Observe that each user in either of the two schemes needs to encrypt its data for each user block including itself (Encryption times is equal to the depth of the tree), since both of the schemes utilize tree-structure construction to handle user failures. Fig. 4(a) plots the computation cost varying with the number of users in Binary Protocol and PPPA scheme. From the figure, we can easily see that our proposed PPPA scheme is much more computation-efficient.

Communication Overhead: From [7], we know that conventional wireline communication technologies, such as fiber optical technologies, can be used for high-speed data and bulk information delivery between GW and CC while wireless technologies such as WIFI and Zigbee are expected to connect
each user with GW and thousands of residential users. In smart grid, since huge number of users need to upload user report at the same time in order to get real-time situation awareness, communication overhead between GW and users may be the bottleneck of the whole system’s performance. To evaluate the communication-effectiveness of the proposed PPPA scheme, we also make a comparison between it and Binary Protocol. Since both of them employs the cryptographic construction of PSA scheme [16], it is reasonable to assume that the encrypted data’s length is one unit. Fig. 4(b) plots the communication overhead varying with the number of residential users in Binary Protocol and PPPA. From the figure, we can easily see that our proposed PPPA scheme is more efficient in terms of communication overhead between the users and GW.

Failure tolerance: User failures are inevitable in smart grid due to device breakdown or user’s maloperation. When facing such challenges, the proposed PPPA scheme can remain operative. As described before, it is not hard for CC to detect malfunctioning users and then find a set of user blocks to uniquely cover the functioning users. In this way, CC can still estimate the overall electricity usage of RA.

VI. RELATED WORKS

Recently, there have appeared many research works studying privacy-preserving data aggregation in smart grid. However, some common techniques such as anonymizing or blind signing can not provide users with the real assurances of privacy. One of the promising solutions is the differential privacy technique, which is formulated by Dwork et al. [13] [14]. Intuitively, such notion is based on the idea that the presence or the absence of an individual in the database, or its particular value, should not affect in a significant way the probability of obtaining a certain answer for a given query [18]. Therefore, target statistic is randomized before releasing. Compared with the traditional scenario where aggregator is trusted with responsibility of randomization, recent works on differential privacy consider the existence of a malicious aggregator. In [16], Shi, et al. use applied cryptographic technique and distributed data randomization technique to guarantee the differential privacy of the outcome statistic. Then, Chan, et al. [15] introduce binary tree construction idea to achieve fault tolerance, a problem left by [16]. Greatly inspired by Finkel, et. al’s generalization of the binary tree to quad tree for the treatment of data with inherently two-dimensional structure [19], we propose to use quad tree to further reduce computation and communication cost while still achieving fault tolerance in smart grid. In addition, we take into account the architecture of smart grid when designing the privacy-preserving aggregation scheme contrasting with the previous differential privacy literature.

VII. CONCLUSIONS

In this paper, we have proposed a practical privacy-preserving aggregation scheme, named PPPA, for smart grid. It is characterized by using applied cryptographic technique to achieve provable security guarantee, differential technique to achieve privacy-preserving of each individual user, and quad tree structure to achieve failure tolerance. Formal security analysis have shown that the proposed PPPA scheme can provide strong security and privacy guarantee. In addition, extensive performance evaluation have demonstrated its practical feasibility. In our future work, we will consider the more challenging scenario where data pollution attack is launched to smart grid.

ACKNOWLEDGMENT

This work was supported in part by NSFC (No. 61171149), the Fundamental Research Funds for the Chinese Central Universities (No. 2013zzzx008-2), and ORF-RE, Ontario, Canada.

REFERENCES