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.

EPPA: An Efficient and Privacy-Preserving Aggregation Scheme for Secure Smart Grid Communications

Rongxing Lu, *Member, IEEE,* Xiaohui Liang, *Student Member, IEEE,* Xu Li, *Member, IEEE,* Xiaodong Lin, *Member, IEEE,* and Xuemin (Sherman) Shen, *Fellow, IEEE*

Abstract—The concept of smart grid has emerged as a convergence of traditional power system engineering and information and communication technology. It is vital to the success of next generation of power grid, which is expected to be featuring reliable, efficient, flexible, clean, friendly and secure characteristics. In this paper, we propose an efficient and privacy-preserving aggregation scheme, named EPPA, for smart grid communications. EPPA uses a super-increasing sequence to structure multi-dimensional data and encrypt the structured data by the homomorphic Paillier cryptosystem technique. For data communications from user to smart grid operation center, data aggregation is performed directly on ciphertext at local gateways without decryption, and the aggregation result of the original data can be obtained at the operation center. EPPA also adopts the batch verification technique to reduce authentication cost. Through extensive analysis, we demonstrate that EPPA resists various security threats and preserve user privacy, and has significantly less computation and communication overhead than existing competing approaches.

Index Terms—Smart grid, Security, Privacy-preserving, Multi-dimensional aggregation.

1 Introduction

The August 2003 electrical blackout in North America affected over 100 power plants and paralyzed tens of millions of people's lives [1]. Investigations revealed that the failure was due to load imbalance and lack of effective real-time diagnosis, among others. Indeed, because electricity cannot be easily stocked, load must be matched by the power supply and transmission capacity in the electric power grid. While swift advances in science and technology are triggering radical innovations in many fields, today's power grid is surprisingly still grounded on a design more than 100 years old [2]. With the ubiquitous adoption of electronic devices, it is undoubtedly outdated and no longer meets our growing demand for continuous stable electricity distribution. Modernizing the aging power system is currently a strategic plan in many countries.

Recently, the concept of smart grid has emerged and been recognized as the next generation of power grid [3], [4], [5], [6], [7]. Traditional grid is featured with centralized one-way transmission (from generation plants to customers) and demand-driven response. Smart grid combines traditional grid and information and control technologies. It allows decentralized two-way transmission and reliability- and efficiency-driven response, and aims to provide improved reliability (e.g., self-healing, self-activating, automated outage management), efficiency (e.g., cost-effective power generation, transmission and distribution), sustainability (e.g., accommodation of future alternative and renewable power sources), consumer involvement, and security (physical and cyber).

Smart meters are important components of smart grid. They are two-way communication devices deployed at consumers premise, records power consumption periodically. With smart meters, smart grid is able to collect real-time information about grid operations and status at an operation center, through a reliable communications network deployed in parallel to the power transmission and distribution grid, as shown in Fig. 1. The operation center may be implemented in a distributed way and span different geographic regions. It is responsible for dynamically adjusting power supply to meet demand, and detecting and responding to weaknesses or failures in the power system in real time. Smart grid also automates reliable power distribution by engaging and empowering customers in utility management. It exposes customers' detailed real-time electricity use information (through smart meters) to utility companies, which may then change electricity price accordingly or even adjust customers' usage by pre-installed load control switches in order to help flatten demand peaks. Customers are allowed to access their own real-time use information through smart grid services. In order to lower their own energy costs and enjoy uninterrupted activities, they will be willing to use energy-efficient appliances and tend to shift power use from peak times to non-peak times.

Cyber security is of paramount importance in smart grid as communications are deeply involved in its operations [8], [9], [10], [11], [12], [13]. All the data transmitted in the grid must be authenticated and secured against malicious modification. Privacy (i.e., data confidentiality) is a primary concern from

- *R. Lu, X. Liang and X. Shen are with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1 (e-mail: {rxlu, x27liang, xshen}@bbcr.uwaterloo.ca).*
- *X. Lin is with the Faculty of Business and Information Technology, University of Ontario Institute of Technology, Oshawa, ON, Canada L1H 7K4 (e-mail: xiaodong.lin@uoit.ca).*
- *X. Li is with INRIA Lille - Nord Europe, France. (e-mail: xu.li@inria.fr).*

Digital Object Indentifier 10.1109/TPDS.2012.86 1045-9219/12/$31.00 © 2012 IEEE

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.

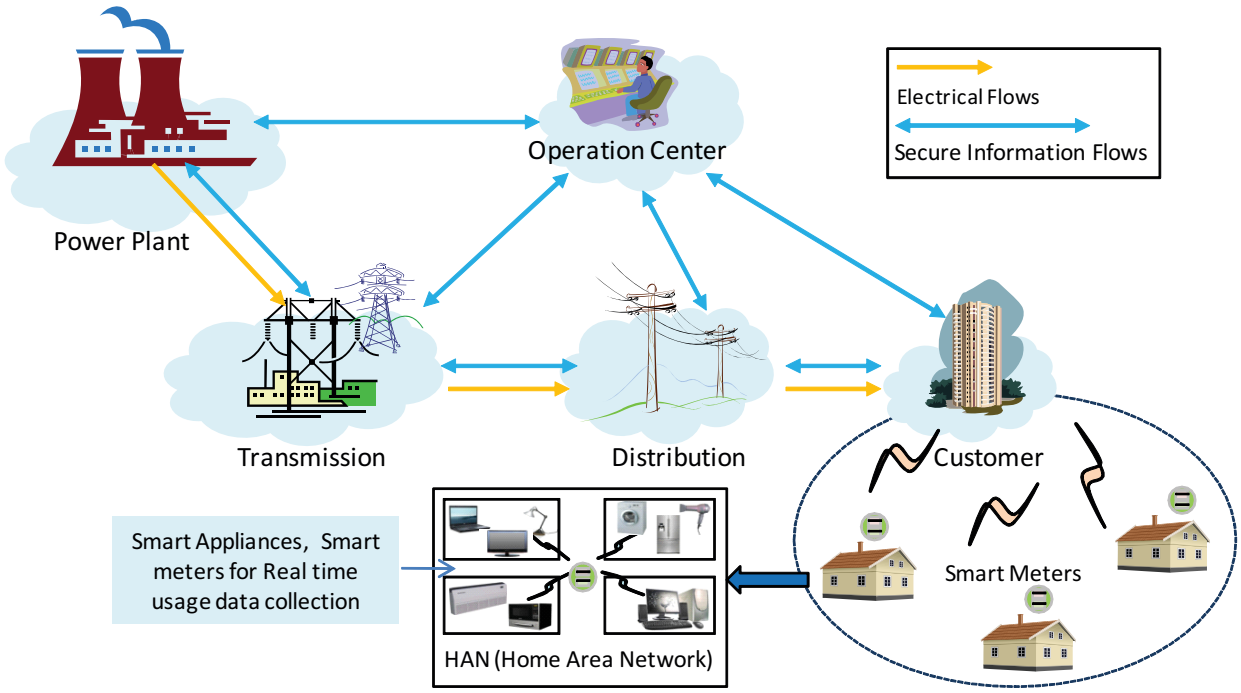


Fig. 1. The conceptual architecture of smart grid

customers point of view as power use information may reveal their physical activities. For example, unusually low daily power consumption of a household and continuous lack of power use for stove and microwave indicate that the home owners are probably away from their home. Such privacy-sensitive information must be protected from unauthorized access. Data confidentiality can be achieved by simple end-to-end encryption. While hiding communication content and protecting privacy, encryption increases data size, and may cause unacceptable communication overhead when power use information is collected at high frequency. Considering that the operation center is concerned only with the overall information in a region, the data of individual consumers in the region can be aggregated at a local gateway and forwarded in a compact form to the operation center in order to save communication bandwidth.

To preserve user privacy, local gateways should not be able to access the content of consumers data. To enable them to perform data aggregation, homomorphic encryption techniques [14] may be applied for encrypting consumers data. In this technique, a specific linear algebraic manipulation toward the plaintext is equivalent to another one conducted on the ciphertext. This unique feature allows the local gateway to perform summation and multiplication based aggregation on received consumer data without decrypting them. Existing data aggregation schemes [15], [16], [17] regards power use information as one-dimensional information. With smart meters being used, it is however multi-dimensional in nature, for example, including the amount of energy consumed, at what time and for what purpose the consumption was, and so on. Taking into account all the dimensions allows finer-grained control and optimization. When multiple dimensions are present, the existing schemes [15], [16], [17] will have to process every dimension separately. We further notice that power usage information is often small in size, smaller than the plain text space of the encryption algorithm used. Each time when it is encrypted, its size will be increased to occupy the entire plain text space. Considering the high data collection frequency, multi-dimensional use information and massive number of consumers, the existing data aggregation schemes generate not only huge communication cost but also impose overwhelming process load on local gateways.

To save communication and computation resources, in this paper, we process all the dimension data as a whole rather than separately, and propose a novel Efficient and Privacy-Preserving Aggregation (EPPA) scheme. This scheme expresses multi-dimensional power use data in a single-dimensional form and supports privacy-preserving aggregation operations on the reformatted data. As a result, data can be efficiently reported to smart grid operation center at a high frequency for real-time monitoring and control. The main contributions of this paper are two-fold.

- Firstly, inspired by the fact that electricity usage data is small in size and multi-dimensional in nature, we present the novel EPPA scheme that utilizes the homomorphic Paillier cryptosystem [14] to achieve privacy-preserving multi-dimensional data aggregation and efficient smart grid communications. Compared with traditional one-dimensional aggregation schemes [15], [16], [17], it leads to dramatically reduced the computation and communication cost.
- Secondly, we analyze the security strength and privacy-preservation ability of EPPA. In particular, we apply the provable security technique to formally prove that the smart grid operation center's response is semantic secure under the chosen plaintext attack. Through com-

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.

parative performance analysis, we demonstrate that EPPA is indeed significantly more efficient than existing one-dimensional aggregation schemes [15], [16], [17].

The remainder of this paper is organized as follows. In Section 2, we introduce our system model, security requirements and our design goal. In Section 3, we recall the bilinear pairings [18] and Paillier cryptosystem as the preliminaries. Then, we present our EPPA scheme in Section 4, followed by its security analysis and performance evaluation in Section 5 and Section 6, respectively. We also discuss the related work in Section 7. Finally, we draw our conclusions in Section 8.

2 System Model, Security Requirements and Design Goal

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

2.1 System Model

In our system model, we mainly focus on how to report residential users' privacy-preserving electricity usage data to the operation center in smart grid communications. Specifically, we consider a typical residential area (RA), which comprises a local gateway (GW) connected with smart grid operation center, and a large number of residential users $\mathbb{U} = \{U_1, U_2, \cdots, U_w\}$, as shown in Fig. 2. The GW is a powerful workshop, which mainly performs two functions: aggregation and relaying. The aggregation component is responsible for aggregating residential users' electricity usage data into a compressed one, while the relaying component helps residential users with forwarding data to the operation center, i.e., to a trusted operation authority (OA) located at operation center, and also helps the OA with relaying the responses back to the residential users in the RA as well. In the process of the aggregation and relaying, the GW will also perform some authentication operations to guarantee the data's authenticity and integrity.

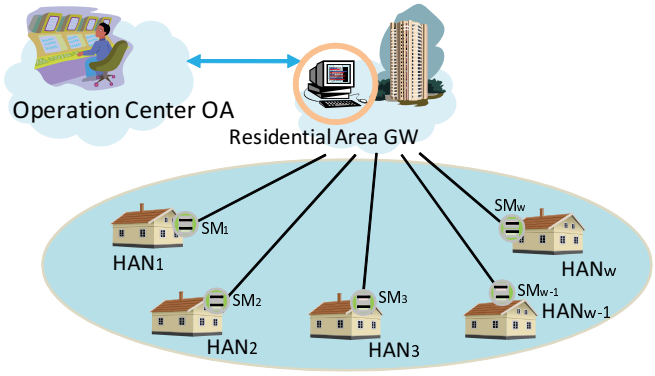


Fig. 2. System model under consideration

Each user $U_i \in \mathbb{U}$ is equipped with various smart meters (SMs), which form a Home Area Network (HAN), and can electronically record the real-time data about electricity use. These near real-time data will then be reported to the OA every a certain period with the relay of the GW. On receiving the reports from residential users, the OA can get the real-time situational awareness so as to make the electricity use more efficient by either carrying out the dynamic price or directly controlling to reduce consumption during peak periods and shift some demands to off-peak hours.

Communication model. In the residential area RA, the communication between each user $U_i \in \mathbb{U}$ and the local GW is through relatively inexpensive WiFi technology. In other words, within the WiFi coverage of the GW, each $U_i \in \mathbb{U}$ can directly/indirectly communicate with the GW. On the other hand, since the distance between the residential area and the operation center is far away, the communication between the GW and the OA is through either wired links or any other links with high bandwidth and low delay. However, although the communication in smart grid is featured with high bandwidth and low delay, since hundred and thousand of residential users scattered at different residential areas in a region will report their electricity usage data almost at the same time, the communication efficiency of the GW-to-OA communication is still a challenging issue.

2.2 Security Requirements

Security is crucial for the success of secure smart grid communications. In our security model, we consider the OA and the GW are trustable, and the residential users $\mathbb{U} = \{U_1, U_2, \cdots, U_w\}$ are honest as well. However, there exists an adversary $\mathcal{A}$ residing in the RA to eavesdrop the residential users' reports. More seriously, the adversary $\mathcal{A}$ could also intrude in the database of the GW and the smart grid operation center to steal the individual user reports. In addition, the adversary $\mathcal{A}$ could also launch some active attacks to threaten the data integrity. Therefore, in order to prevent the adversary $\mathcal{A}$ from learning the users' reports and to detect the adversary $\mathcal{A}$'s malicious actions, the following security requirements should be satisfied in secure smart grid communications.

- *Confidentiality.* Protect individual residential user's reports from the adversary $\mathcal{A}$, i.e., even if $\mathcal{A}$ eavesdrops the WiFi communication in the RA, it cannot identify the contents of the reports; and even if $\mathcal{A}$ steals the data from the operation center's and/or the GW's databases, it can also not identify each individual user's data. In such a way, each individual user's electricity usage data can achieve the privacy-preserving requirement. In addition, the confidentiality requirement also includes the OA's responses should be privacy-preserving, i.e., only the legal residential users in the RA can read them.
- *Authentication and Data Integrity.* Authenticating an encrypted report that is really sent by a legal residential user and has not been altered during the transmission, i.e., if the adversary $\mathcal{A}$ forges and/or modifies a report, the malicious operations should be detected. Then, only the correct reports can be received by the OA and helpful for the electricity use monitoring. Meanwhile, the responses from the OA should also be authenticated so that the residential users can receive the authentic and reliable information.

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.

2.3 Design Goal

Under the aforementioned system model and security requirements, our design goal is to develop an efficient and privacy-preserving aggregation scheme for secure smart grid communications. Specifically, the following two objectives should be achieved.

- *The security requirements should be guaranteed in the proposed scheme.* As stated above, if the smart grid does not consider the security, the residential users' privacy could be disclosed, and the real-time electricity use reports could be altered. Then, the smart grid cannot step into its flourish. Therefore, the proposed scheme should achieve the confidentiality, authentication and data integrity requirements simultaneously.
- *The communication-effectiveness should be achieved in the proposed scheme.* Although the communication between the OA and the GW is featured with high-bandwidth and low-delay, to support hundred and thousand residential users' reports to the OA at almost the same time, the proposed scheme should also consider the communication-effectiveness, so that the near real-time user reports can be fast transmitted to the OA.

3 Preliminaries

In this section, we outline the bilinear pairing technique [18] and review the Paillier Cryptosystem [14], which will serve as the basis of the proposed EPPA scheme.

3.1 Bilinear Pairing

Let $\mathbb{G}$, $\mathbb{G}_T$ be two cyclic groups of the same prime order q, and P be a generator of group $\mathbb{G}$. Suppose $\mathbb{G}$ and $\mathbb{G}_T$ are equipped with a pairing, i.e., a non-degenerated and efficiently computable bilinear map $e : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}_T$ such that $e(P,P) \neq 1_{\mathbb{G}_T}$ and $e(aP_1, bQ_1) = e(P_1, Q_1)^{ab} \in \mathbb{G}_T$ for all $a, b \in \mathbb{Z}_q^*$ and any $P_1, Q_1 \in \mathbb{G}$. We refer to [18] for a more comprehensive description of pairing technique, and complexity assumptions.

Definition 1: A bilinear parameter generator $\mathcal{G}en$ is a probabilistic algorithm that takes a security parameter κ as input, and outputs a 5-tuple $(q, P, \mathbb{G}, \mathbb{G}_T, e)$ where q is a κ-bit prime number, $\mathbb{G}, \mathbb{G}_T$ are two groups with the same order q, $P \in \mathbb{G}$ is a generator, and $e : \mathbb{G} \times \mathbb{G} \rightarrow \mathbb{G}_T$ is a non-degenerated and efficiently computable bilinear map.

Definition 2: (Computational Diffie-Hellman (CDH) Problem) The CDH problem is stated as follows: Given the elements $(P, aP, bP) \in \mathbb{G}$ for unknown $a, b \in \mathbb{Z}_q^*$, to compute $abP \in \mathbb{G}$.

Definition 3: (Bilinear Diffie-Hellman (BDH) Problem) The BDH problem is stated as follows: Given the elements $(P, aP, bP, cP) \in \mathbb{G}$ for unknown $a, b, c \in \mathbb{Z}_q^*$, to compute $e(P,P)^{abc} \in \mathbb{G}_T$.

Definition 4: (Decisional BDH (DBDH) Problem) The DBDH problem in $(\mathbb{G}, \mathbb{G}_T)$ is stated as follows: Given the elements $(P, aP, bP, cP) \in \mathbb{G}$ for unknown $a, b, c \in \mathbb{Z}_q^*$ and $W \in \mathbb{G}_T$, decide whether $W = e(P,P)^{abc} \in \mathbb{G}_T$ or a random element R drawn from $\mathbb{G}_T$.

3.2 Paillier Cryptosystem

The Paillier Cryptosystem can achieve the homomorphic properties, which is widely desirable in many privacy-preserving applications [19], [20]. Concretely, the Paillier Cryptosystem is comprised of three algorithms: key generation, encryption and decryption.

• *Key Generation:* Given the security parameter κ_1, two large prime numbers p_1, q_1 are first chosen, where $|p_1| = |q_1| = \kappa_1$. Then, the RSA modulus $n = p_1 q_1$ and $\lambda = lcm(p_1 - 1, q_1 - 1)$ are computed. Define a function $L(u) = \frac{u-1}{n}$, after choosing a generator $g \in \mathbb{Z}_{n^2}^*$, $\mu = (L(g^\lambda \bmod n^2))^{-1} \bmod n$ is further calculated. Then, the public key is $pk = (n, g)$, and the corresponding private key is $sk = (\lambda, \mu)$.

• *Encryption:* Given a message $m \in \mathbb{Z}_n$, choose a random number $r \in \mathbb{Z}_n^*$, and the ciphertext can be calculated as $c = E(m) = g^m \cdot r^n \bmod n^2$.

• *Decryption:* Given the ciphertext $c \in \mathbb{Z}_{n^2}^*$, the corresponding message can be recovered as $m = D(c) = L(c^{\lambda \bmod n^2}) \cdot \mu \bmod n$. Note that, the Paillier Cryptosystem is provably secure against chosen plaintext attack, and the correctness and security can be referred to [14].

4 Proposed EPPA Scheme

In this section, we propose the efficient and privacy-preserving aggregation scheme (EPPA) for secure smart grid communications, which mainly consists of the following four parts: system initialization, user report generation, privacy-preserving report aggregation, and secure report reading and response.

4.1 System Initialization

For a single-authority smart grid system under consideration, it is reasonable to assume a trusted operation authority (OA) can bootstrap the whole system. Specifically, in the system initialization phase, given the security parameters κ, κ_1, OA first generates $(q, P, \mathbb{G}, \mathbb{G}_T, e)$ by running $\mathcal{G}en(\kappa)$, and then calculates the Paillier Cryptosystem's public key $(n = p_1 q_1, g)$, and the corresponding private key (λ, μ), where p_1, q_1 are two large primes with $|p_1| = |q_1| = \kappa_1$. Assume that the maximum number of households in a residential area is no more than a constant w, and there are total l types of electricity usage data $(T_1, T_2, \cdots, T_l)$ to be reported in smart grid communications, the value of each type T_i is less than a constant d. Then, OA chooses a super-increasing sequence $\vec{\mathbf{a}} = (a_1 = 1, a_2, \cdots, a_l)$, where $a_2, \cdots, a_l$ are large primes such that the length $|a_i| \geq \kappa$, $\sum_{j=1}^{i-1} a_j \cdot w \cdot d < a_i$ for $i = 2, \cdots, l$, and $\sum_{i=1}^{l} a_i \cdot w \cdot d < n$. After that, OA computes $(g_1, g_2, \cdots, g_l)$, where

$$g_i = g^{a_i}, \text{ for } i = 1, 2, \cdots, l \tag{1}$$

OA also chooses two random elements $Q_1, Q_2 \in \mathbb{G}$, two random numbers $\alpha, x \in \mathbb{Z}_q^*$, and computes $e(P,P)^\alpha$, $Y = xP$. In addition, OA chooses two secure cryptographic hash functions H and H_1, where $H : \{0,1\}^* \rightarrow \mathbb{G}$ and $H_1 : \{0,1\}^* \rightarrow \mathbb{Z}_q^*$. In the end, OA publishes the system parameters as

$$\mathsf{pubs} = \left\{ \begin{array}{c} q, P, \mathbb{G}, \mathbb{G}_T, e, n, g_1, \cdots, g_l, \\ Q_1, Q_2, e(P,P)^\alpha, Y, H, H_1 \end{array} \right\} \tag{2}$$

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.

and keeps the master keys $(\lambda, \mu, \vec{\mathbf{a}}, \alpha, x)$ secretly.

When a local gateway (GW) of the residential area (RA) registers itself in the system, it first chooses a random number $x_g \in \mathbb{Z}_q^*$ as the private key, and computes the corresponding public key $Y_g = x_g P$. While when a HAN user $U_i \in \mathbb{U}$ of the RA joins in the system, U_i chooses a random number $x_i \in \mathbb{Z}_q^*$ as the private key, and computes the corresponding public key $Y_i = x_i P$. In addition, the OA uses the master key (α, x) to compute

$$t_{i1} = H_1(U_i||RA||\alpha), t_{i2} = H_1(U_i||RA||x) \tag{3}$$

and generates the authorized RA-related key ak_i to U_i, where

$$ak_i = (\alpha P + t_{i1}Y, t_{i1}P, t_{i2}P, t_{i1}Q_1 + t_{i2}Q_2) \tag{4}$$

With the authorized key ak_i, U_i can securely receive the response sent by the OA in smart grid communication system.

4.2 User Report Generation

In order to achieve the nearly real-time residential users' electricity usage data every η minutes, e.g., $\eta = 15$ minutes, each HAN user $U_i \in \mathbb{U}$ uses the smart meters to collect l types of data $(d_{i1}, d_{i2}, \cdots, d_{il})$, where each $d_{ij} \leq d$, and performs the following steps:

- *Step-1:* Choose a random number $r_i \in \mathbb{Z}_n^*$, and compute

$$C_i = g_1^{d_{i1}} \cdot g_2^{d_{i2}} \cdot \cdots \cdot g_l^{d_{il}} \cdot r_i^n \bmod n^2 \tag{5}$$

- *Step-2:* Use the private key x_i to make a signature σ_i as

$$\sigma_i = x_i H(C_i||RA||U_i||TS) \tag{6}$$

where TS is the current timestamp, which can resist the potential replay attack.

- *Step-3:* Report the encrypted electricity usage data $C_i||RA||U_i||TS||\sigma_i$ to the local gateway GW in the residential area RA.

4.3 Privacy-preserving Report Aggregation

After receiving total w encrypted electricity usage data $C_i||RA||U_i||TS||\sigma_i$, for $i = 1, 2, \cdots, w$, the local GW first checks the timestamp TS and the signature σ_i to verify its validity, i.e., verify whether $e(P, \sigma_i) \stackrel{?}{=} e(Y_i, H(C_i||RA||U_i||TS))$. If it does hold, the signature is accepted, since $e(P, \sigma_i) = e(P, x_i H(C_i||RA||U_i||TS)) = e(Y_i, H(C_i||RA||U_i||TS))$. In order to make the verification efficiently, the GW can perform the batch verification as

$$\begin{aligned} e\left(P, \sum_{i=1}^{w} \sigma_i\right) &= e\left(P, \sum_{i=1}^{w} x_i H(C_i||RA||U_i||TS)\right) \\ &= \prod_{i=1}^{w} e\left(P, x_i H(C_i||RA||U_i||TS)\right) \\ &= \prod_{i=1}^{w} e\left(Y_i, H(C_i||RA||U_i||TS)\right) \end{aligned} \tag{7}$$

Then, the time-consuming pairing operations $e(\cdot, \cdot)$ can be reduced from $2w$ to $w + 1$ times.

After the validity checking, the GW performs the following steps for privacy-preserving report aggregation:

- *Step-1:* Compute the aggregated and encrypted data C on $C_1, C_2, \cdots, C_w$ as

$$C = \prod_{i=1}^{w} C_i \bmod n^2 \tag{8}$$

- *Step-2:* Use the private key x_g to make a signature σ_g as

$$\sigma_g = x_g H(C||RA||GW||TS) \tag{9}$$

where TS is the current timestamp.

- *Step-3:* Report the aggregated and encrypted data $C||RA||GW||TS||\sigma_g$ to the operation authority OA.

4.4 Secure Report Reading and Response

Upon receiving $C||RA||GW||TS||\sigma_g$, the OA first verifies the validity by checking $e(P, \sigma_g) = e(Y_g, H(C||RA||GW||TS))$, and then performs the following steps to read the aggregated and encrypted report C, where C is implicitly formed by

$$\begin{aligned} C &= \prod_{i=1}^{w} C_i \bmod n^2 \\ &= \prod_{i=1}^{w} g_1^{d_{i1}} \cdot g_2^{d_{i2}} \cdot \cdots \cdot g_l^{d_{il}} \cdot r_i^n \bmod n^2 \\ &= g_1^{\sum_{i=1}^{w} d_{i1}} \cdot g_2^{\sum_{i=1}^{w} d_{i2}} \cdots g_l^{\sum_{i=1}^{w} d_{il}} \cdot \left(\prod_{i=1}^{w} r_i\right)^n \bmod n^2 \\ &= g^{a_1 \sum_{i=1}^{w} d_{i1}} \cdot g^{a_2 \sum_{i=1}^{w} d_{i2}} \cdots g^{a_l \sum_{i=1}^{w} d_{il}} \cdot \left(\prod_{i=1}^{w} r_i\right)^n \bmod n^2 \\ &= g^{a_1 \sum_{i=1}^{w} d_{i1} + a_2 \sum_{i=1}^{w} d_{i2} + \cdots + a_l \sum_{i=1}^{w} d_{il}} \cdot \left(\prod_{i=1}^{w} r_i\right)^n \bmod n^2 \end{aligned} \tag{10}$$

- *Step-1:* By taking $M = a_1 \sum_{i=1}^{w} d_{i1} + a_2 \sum_{i=1}^{w} d_{i2} + \cdots + a_l \sum_{i=1}^{w} d_{il}$ and $R = \prod_{i=1}^{w} r_i$, the report $C = g^M \cdot R^n \bmod n^2$ is still a ciphertext of Paillier Cryptosystem. Therefore, the OA can use the master key (λ, μ) to recover M as

$$M = a_1 \sum_{i=1}^{w} d_{i1} + a_2 \sum_{i=1}^{w} d_{i2} + \cdots + a_l \sum_{i=1}^{w} d_{il} \bmod n \tag{11}$$

- *Step-2:* By invoking the Algorithm 1, the OA can recover and store the aggregated data $(D_1, D_2, \cdots, D_l)$, where each $D_j = \sum_{i=1}^{w} d_{ij}$.

Algorithm 1 Recover the aggregated report

```
procedure RECOVER THE AGGREGATED REPORT
  Input: a⃗ = (a_1 = 1, a_2, ··· , a_l) and M
  Output: (D_1, D_2, ··· , D_l)
    Set X_l = M
    for j = l to 2 do
      X_{j-1} = X_j mod a_j
      D_j = (X_j − X_{j-1}) / a_j = Σ_{i=1}^{w} d_ij
    end for
    D_1 = X_1 = Σ_{i=1}^{w} d_i1
    return (D_1, D_2, ··· , D_l)
end procedure
```

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.

The correctness of Algorithm 1. In Algorithm 1, $X_l = M$, i.e., $X_l = a_1\sum_{i=1}^{w} d_{i1} + a_2\sum_{i=1}^{w} d_{i2} + \cdots + a_{l-1}\sum_{i=1}^{w} d_{i(l-1)} + a_l\sum_{i=1}^{w} d_{il} \bmod n$. Since any type of data is less than a constant d, we have

$$\begin{aligned} &a_1\sum_{i=1}^{w} d_{i1} + a_2\sum_{i=1}^{w} d_{i2} + \cdots + a_{l-1}\sum_{i=1}^{w} d_{i(l-1)} \\ <&a_1\sum_{i=1}^{w} d + a_2\sum_{i=1}^{w} d + \cdots + a_{l-1}\sum_{i=1}^{w} d \\ =&\sum_{j=1}^{l-1} a_j w d < a_l \end{aligned} \tag{12}$$

Therefore, $X_{l-1} = X_l \bmod a_l = a_1\sum_{i=1}^{w} d_{i1} + a_2\sum_{i=1}^{w} d_{i2} + \cdots + a_{l-1}\sum_{i=1}^{w} d_{i(l-1)}$, and

$$\frac{X_l - X_{l-1}}{a_l} = \frac{a_l\sum_{i=1}^{w} d_{il}}{a_l} = \sum_{i=1}^{w} d_{il} = D_l \tag{13}$$

With the similar procedure, we can also prove each $D_j = \sum_{i=1}^{w} d_{ij}$, for $j = 1, 2, \cdots, l-1$. As a result, the correctness of Algorithm 1 is shown. □

After analyzing the near real-time electricity usage data $(D_1, D_2, \cdots, D_l)$, the OA responds a message $m \in \mathbb{G}_T$ to inform HAN users $\mathbb{U} = \{U_1, U_2, \cdots, U_w\}$ in the residential area RA about electricity use and control their cost. The concrete steps are performed as follows.

• *Step-1:* The OA first chooses a random number $s \in \mathbb{Z}_q^*$, and computes $\overline{C} = (\overline{C}_1, \overline{C}_2, \overline{C}_3, \overline{C}_4)$, where

$$\begin{cases} \overline{C}_1 = m \cdot e(P,P)^{\alpha s} \bmod q \\ \overline{C}_2 = sP, \overline{C}_3 = sY - sQ_1, \overline{C}_4 = -sQ_2 \end{cases} \tag{14}$$

Then, the OA makes the signature $\sigma = xH(\overline{C}||RA||OA||TS)$, where TS is the current timestamp, and sends back $\overline{C}||\sigma$ to the local GW at the residential area RA.

• *Step-2:* Upon receiving $\overline{C}||\sigma$, the GW verifies the validity of $\overline{C}$ by checking whether $e(P,\sigma) = e(Y, H(\overline{C}||RA||OA||TS))$. If it does hold, the GW broadcasts $\overline{C}$ in the residential area RA.

• *Step-3:* After receiving the authenticated $\overline{C}$ from the GW, each HAN user $U_i \in \mathbb{U}$ uses the authorized key $ak_i = (\alpha P + t_{i1}Y, t_{i1}P, t_{i2}P, t_{i1}Q_1 + t_{i2}Q_2)$ to recover m from $\overline{C}$ in the followings.

$$\begin{aligned} &\frac{e(\overline{C}_2, \alpha P + t_{i1}Y)}{e(t_{i1}P, \overline{C}_3)e(t_{i2}P, \overline{C}_4)e(t_{i1}Q_1 + t_{i2}Q_2, \overline{C}_2)} \\ =&\frac{e(sP, \alpha P + t_{i1}Y)}{e(t_{i1}P, sY - sQ_1)e(t_{i2}P, -sQ_2)e(t_{i1}Q_1 + t_{i2}Q_2, sP)} \\ =&\frac{e(sP, \alpha P + t_{i1}Y)}{e(t_{i1}P, sY)} = e(P,P)^{\alpha s} \end{aligned} \tag{15}$$

$$\frac{\overline{C}_1}{e(P,P)^{\alpha s}} = \frac{m \cdot e(P,P)^{\alpha s}}{e(P,P)^{\alpha s}} = m \tag{16}$$

With the recovered information m, U_i can determine to shift power use from peak times to non-peak times for electricity use efficiency.

5 Security Analysis

In this section, we analyze the security properties of the proposed EPPA scheme. In particular, following the security requirements discussed earlier, our analysis will focus on how the proposed EPPA scheme can achieve the individual residential user's report privacy preservation, the report's source authentication and data integrity, and the confidentiality of the OA's response.

• *The individual user's report is privacy-preserving in the proposed EPPA scheme.* In the proposed EPPA scheme, user U_i's data $(d_{i1}, d_{i2}, \cdots, d_{il})$ sensed by smart meters are formed as $C_i = g_1^{d_{i1}} \cdot g_2^{d_{i2}} \cdot \cdots \cdot g_l^{d_{il}} \cdot r_i^n \bmod n^2$, which can be implicitly expressed as

$$\begin{aligned} C_i &= g_1^{d_{i1}} \cdot g_2^{d_{i2}} \cdot \cdots \cdot g_l^{d_{il}} \cdot r_i^n \bmod n^2 \\ &= g^{a_1 d_{i1}} \cdot g^{a_2 d_{i2}} \cdot \cdots \cdot g^{a_l d_{il}} \cdot r_i^n \bmod n^2 \\ &= g^{a_1 d_{i1} + a_2 d_{i2} + \cdots + a_l d_{il}} \cdot r_i^n \bmod n^2 \end{aligned} \tag{17}$$

Obviously, let m_i be $a_1 d_{i1} + a_2 d_{i2} + \cdots + a_l d_{il}$, then the ciphertext $C_i = g^{m_i} \cdot r_i^n \bmod n^2$ is still a valid ciphertext of Paillier Cryptosystem. Since Paillier Cryptosystem is semantic secure against the chosen plaintext attack, the data $(d_{i1}, d_{i2}, \cdots, d_{il})$ in m_i is also semantic secure and privacy-preserving. Therefore, even though the adversary $\mathcal{A}$ eavesdrops C_i, he still cannot identify the corresponding contents. After collecting all reports $C_1, C_2, \cdots, C_w$ from the residential users, the GW will not recover each reports, instead, it just computes $C = \prod_{i=1}^{w} C_i \bmod n^2$ to perform report aggregation. Therefore, even if the adversary $\mathcal{A}$ intrudes in the GW's database, he cannot get the individual report $(d_{i1}, d_{i2}, \cdots, d_{il})$ either. Finally, after receiving $C = \prod_{i=1}^{w} C_i \bmod n^2$ from the GW, the OA recovers C as $(D_1, D_2, \cdots, D_l)$, where each $D_j = \sum_{j=1}^{w} d_{ij}$, and stores the entry in the database. However, since each $D_j = \sum_{j=1}^{w} d_{ij}$ is an aggregated result, even if the adversary $\mathcal{A}$ steals the data, he still cannot get the individual user U_i's data $(d_{i1}, d_{i2}, \cdots, d_{il})$. Therefore, from the above three aspects, the individual user's report is privacy-preserving in the proposed EPPA scheme.

• *The authentication and data integrity of the individual user's report and the aggregated report are achieved in the proposed EPPA scheme.* In the proposed EPPA scheme, each individual user's report and the aggregated report are signed by BLS short signature [21]. Since the BLS short signature is provably secure under the CDH problem in the random oracle model [22], the source authentication and data integrity can be guaranteed. As a result, the adversary $\mathcal{A}$'s malicious behaviors in the smart grid communications can be detected in the proposed EPPA scheme.

• *The confidentiality of the OA's response is also achieved in the proposed EPPA scheme.* When the OA responds m to the residential users in RA, he encrypts it as $\overline{C} = (\overline{C}_1, \overline{C}_2, \overline{C}_3, \overline{C}_4)$. In order to show the confidentiality of m is satisfied, we use the following theorem to prove that $\overline{C} = (\overline{C}_1, \overline{C}_2, \overline{C}_3, \overline{C}_4)$ is semantic secure against chosen-plaintext attack under the assumption that DBDH problem is hard.

Theorem 1: The ciphertext $\overline{C} = (\overline{C}_1, \overline{C}_2, \overline{C}_3, \overline{C}_4)$ is semantic secure against chosen-plaintext attack under the DBDH

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.

assumption.

Proof: Let $a, b, c \in \mathbb{Z}_q^*$, $\tilde{b} \in \{0, 1\}$. If $\tilde{b} = 0$, set $W = e(P, P)^{abc}$; while if $\tilde{b} = 1$, set W to be a random element in $\mathbb{G}_T$. Given (P, aP, bP, cP, W), the DBDH problem is to guess $\tilde{b}$. Assume that there is an adversary $\mathcal{A}$ which runs in polynomial time and has a non-negligible advantage ε to break the semantic security of $\overline{C} = (\overline{C}_1, \overline{C}_2, \overline{C}_3, \overline{C}_4)$ in the EPPA scheme, then we can construct another adversary $\mathcal{B}$ which has access to $\mathcal{A}$ and achieves a non-negligible advantage to break the DBDH problem.

First, $\mathcal{B}$ is given an DBDH instantiation (P, aP, bP, cP, W) as input, with $W = e(P, P)^{abc}$ when $\tilde{b} = 0$. $\mathcal{B}$ chooses three random numbers $\alpha', \beta_1, \beta_2 \in \mathbb{Z}_q^*$ and sets the system parameters

$$\begin{cases} Y = xP = \underline{aP} \\ Q_1 = \underline{xP + \beta_1 P = aP + \beta_1 P}, Q_2 = \underline{\beta_2 P} \\ e(P, P)^{\alpha} = \underline{e(aP, bP) \cdot e(P, P)^{\alpha'}} \end{cases} \tag{18}$$

which implicitly denotes $\alpha = ab + \alpha'$. Because of the randomness of $\alpha', \beta_1, \beta_2$ and aP, bP, the distributions of $(Y, Q_1, Q_2, e(P, P)^{\alpha})$ are unchanged, then the simulated parameters $(Y, Q_1, Q_2, e(P, P)^{\alpha})$ are indistinguishable from the real environment in the eye of $\mathcal{A}$.

Upon receiving $(Y, Q_1, Q_2, e(P, P)^{\alpha})$, $\mathcal{A}$ chooses two messages m_0 and m_1 in $\mathbb{G}_T$ and returns them to $\mathcal{B}$. At this moment, $\mathcal{B}$ flips a bit $b^* \in \{0, 1\}$ and generates a ciphertext $\overline{C} = (\overline{C}_1, \overline{C}_2, \overline{C}_3, \overline{C}_4)$, where

$$\begin{cases} \overline{C_1} = \underline{m_b^* \cdot W \cdot e(cP, \alpha' P)} \\ \overline{C_2} = \underline{cP} \\ \overline{C_3} = cY - cQ_1 = caP - c(aP + \beta_1 P) = \underline{\beta_1 cP} \\ \overline{C_4} = -cQ_2 = \underline{-\beta_2 cP} \end{cases} \tag{19}$$

In the end, $\mathcal{B}$ sends $\overline{C} = (\overline{C}_1, \overline{C}_2, \overline{C}_3, \overline{C}_4)$ to $\mathcal{A}$. After receiving $\overline{C}$, $\mathcal{A}$ returns $\mathcal{B}$ a bit b' as the guess of b^*. $\mathcal{B}$ then guesses $\tilde{b} = 0$ if $b' = b^*$. Obviously, when $\tilde{b} = 0$, i.e., $W = e(P, P)^{abc}$, we have

$$\begin{aligned} \overline{C_1} &= m_{b^*} \cdot W \cdot e(cP, \alpha' P) \\ &= m_{b^*} \cdot e(P, P)^{abc} \cdot e(cP, \alpha' P) \\ &= m_{b^*} \cdot e(P, P)^{abc + \alpha' c} = m_{b^*} \cdot e(P, P)^{\alpha c} \end{aligned} \tag{20}$$

Then, $\overline{C_1}$ becomes a valid component of the ciphertext. In this case, $\mathcal{A}$ will guess b^* correctly with the probability $\frac{1}{2} + \varepsilon$. Thus, $\Pr[\mathcal{B} \text{ success} | \tilde{b} = 0] = \frac{1}{2} + \varepsilon$. If $\tilde{b} = 1$, $\overline{C_1} = m_{b^*} \cdot W \cdot e(cP, \alpha' P)$ is independent with b^* due to the randomness of W. Therefore, $\Pr[\mathcal{B} \text{ success} | \tilde{b} = 1] = \frac{1}{2}$. Summarizing the above two cases, we have

$$\Pr[\mathcal{B} \text{ success}] = \frac{1}{2}\left(\frac{1}{2} + \varepsilon\right) + \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{2} + \frac{\varepsilon}{2}$$

Since ε is non-negligible, the above result contradicts with the assumption that DBDH problem is hard. As a result, the ciphertext $\overline{C} = (\overline{C}_1, \overline{C}_2, \overline{C}_3, \overline{C}_4)$ is semantic secure under the chosen plaintext attack, i.e., the OA's response also achieves the confidentiality in the proposed EPPA scheme. □

From the above analysis, we can see that the proposed EPPA scheme is secure and privacy-preserving, and can achieve our security design goal.

6 Performance Evaluation

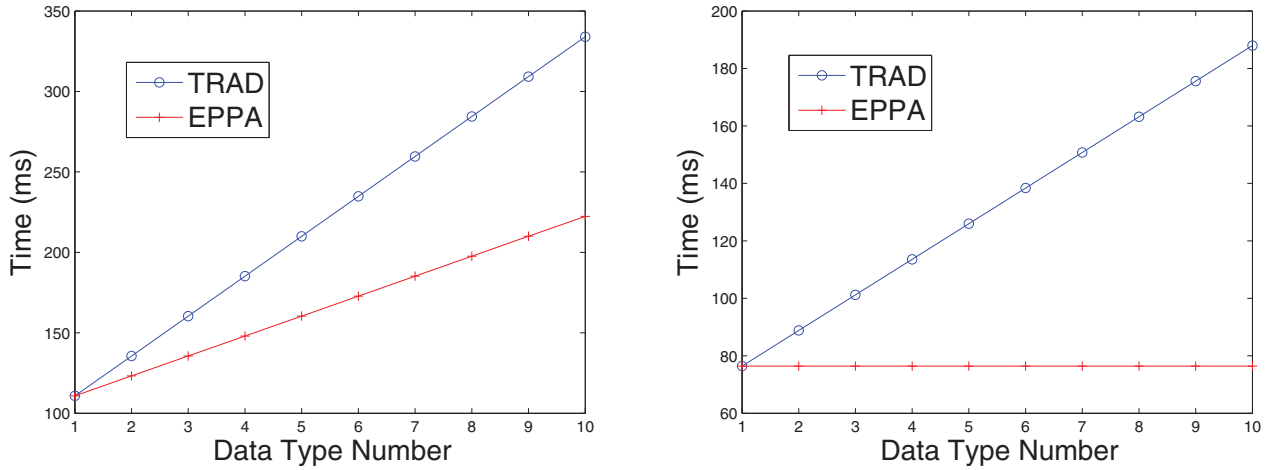
In this section, we evaluate the performance of the proposed EPPA scheme in terms of the computation complexity of residential user, local GW and OA, and the communication overhead of user-to-GW and GW-to-OA communications.

6.1 Computation Complexity

For the proposed EPPA scheme, when a residential user U_i generates an encrypted electricity usage data $C_i||RA||U_i||TS||\sigma_i$, it requires $l+1$ exponentiation operations in $\mathbb{Z}_{n^2}$ to generate C_i, and 1 multiplication operation in $\mathbb{G}$ for σ_i's generation. After receiving the ciphertext from w users, the local GW first verifies the received data by performing a batch verification which includes $w+1$ pairing operations. In addition, the GW should aggregate the reports from different users and generate a signature on the aggregated data. Since the multiplication in $\mathbb{Z}_{n^2}$ is considered negligible compared to exponentiation and pairing operations, the computational cost of aggregation is negligible, and the generation of the signature only includes 1 multiplication operation in $\mathbb{G}$. As for the OA, it verifies the aggregated data from the GW with 2 pairing operations and obtains the data by Paillier decryption which includes 1 exponentiation operation in $\mathbb{Z}_{n^2}$ as shown in Section 3.2. The OA further sends a response to the GW and in turn to residential users. The generation of the response m costs OA for 4 multiplication operations in $\mathbb{G}$ and 1 exponentiation operation in $\mathbb{G}_T$. In order to deliver the response m to users, the extra computational costs for the GW are 2 pairing operations. After obtaining the response, the extra computational costs for users are from 4 pairing operations. Denote the computational costs of an exponentiation operation in $\mathbb{Z}_{n^2}$, a multiplication operation in $\mathbb{G}$, an exponentiation operation in $\mathbb{G}_T$ and a pairing operation by C_e, C_m, C_{et} and C_p, respectively. Then, totally for the user, the GW and the OA, the computational cost will be $(l+1)*C_e+C_m+4*C_p$, $(w+3)*C_p+C_m$, and $2*C_p+C_e+4*C_m+C_{et}$ in the proposed EPPA scheme.

The proposed EPPA scheme enables a residential user to embed multi-dimensional data into one compressed data. It largely reduces the encryption times for users. In the following, for the comparison with EPPA, we consider a traditional approach (denoted by TRAD), where each user generates a ciphertext for one dimensional data. Under this setting, for l-dimensional data, a user has to generate l ciphertexts and consumes totally $2 \cdot l$ exponentiation operations in $\mathbb{Z}_{n^2}$ for the encryption. Including 1 multiplication operation in $\mathbb{G}_T$ for the signature, the total computational costs for a user would be $2l * C_e + C_m$. After receiving the reports from w users, the GW takes $w+1$ pairing operations for verifying the signatures, several negligible-cost multiplication operations in $\mathbb{Z}_{n^2}$ for aggregating the ciphertexts, and 1 multiplication operation in $\mathbb{G}$ for generating the signature. The GW then forwards l ciphertexts to the OA, where each ciphertext contains the sum of one dimensional data of all users. The verification by OA will consume 2 pairing operations. The number of decryptions executed by OA is l and thus l exponentiation operations in $\mathbb{Z}_{n^2}$ are spent on the decryption. For the response phase, we

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.



(a) Computation cost of each user

(b) Computation cost of the OA

Fig. 3. Computation Costs of Users and OA

assume TRAD works exactly the same as the EPPA scheme. Therefore, the computational costs of a residential user, the GW, and the OA will be $2l*C_e+C_m+4*C_p$, $(w+3)*C_p+C_m$, and $2*C_p+l*C_e+4*C_m+C_{et}$, respectively.

TABLE 1
Comparison of Computation Complexity

	EPPA	TRAD
User	$(l+1)*C_e+C_m+4*C_p$	$2l*C_e+C_m+4*C_p$
GW	$(w+3)*C_p+C_m$	$(w+3)*C_p+C_m$
OA	$2*C_p+C_e+4*C_m+C_{et}$	$2*C_p+l*C_e+4*C_m+C_{et}$

We present the computation complexity comparison of EPPA and TRAD in Table 1. Furthermore, we conduct the experiments with PBC [23] and MIRACL [24] libraries running on a 3.0GHz-processor 512MB-memory computing machine to study the operation costs. The experimental results indicate that a single exponentiation operation in $\mathbb{Z}_{n^2}$ ($|n^2| = 2048$) almost costs 12.4 ms, a single multiplication operation in $\mathbb{G}$ with 160 bits costs 6.4 ms and the corresponding pairing operation costs 20 ms. With the exact operation costs, we depict the variation of computation costs in terms of l in Fig. 3. From the figure, it can be obviously shown that the EPPA scheme largely reduces the computation complexity for both users and the OA.

6.2 Communication Overhead

The communications of the proposed EPPA scheme can be divided into two parts, user-to-GW communication and GW-to-OA communication. We first consider the user-to-GW communication, where users generate their data reports and deliver these reports to the local GW. The data report is in the form of $C_i||RA||U_i||TS||\sigma_i$ for user U_i and its size should be $Sz = 2048+|RA|+|U_i|+|TS|+160$ if we choose 1024-bit n and 160-bit $\mathbb{G}$. The GW collects the reports from total w users, indicating that the overall communication overhead between users and the GW is $S_{eppa} = w * Sz$. Alternatively, if the traditional TRAD scheme is adopted, each user has to generate a 2048-bit ciphertext for each dimensional data. In this case, the communication overhead of user-to-GW will increase to $S_{trad} = (2048 * l + |U_i| + |RA| + |TS| + 160) * w$. We plot the communication overhead of both schemes in terms of user number w and data types l, as shown in Fig. 4, where we set $|RA| + |U_i| + |TS|$ as 100-bit length. It can be seen that the proposed EPPA scheme always achieves lower communication overhead compared to the TRAD.

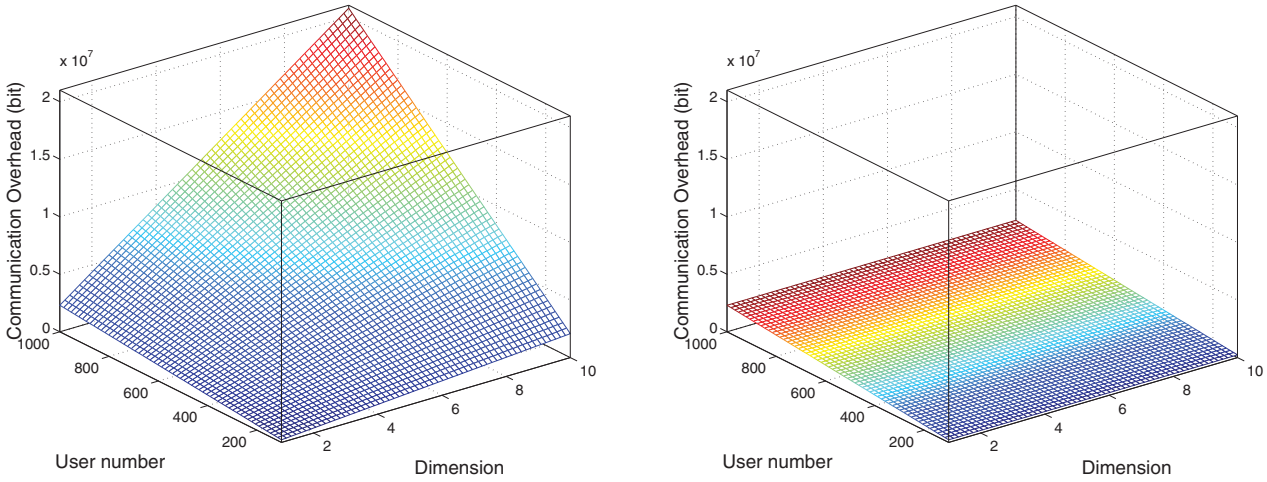
Next, we consider the GW-to-OA communication of both EPPA and TRAD. In EPPA, the communication is off started by the GW who aims to deliver the aggregated report to the OA. The report is in the form of "$C||RA||GW||TS||\sigma_g$" and with $2048 + |RA| + |GW| + |TS| + 160$ in length. In comparison, in TRAD, different dimensional data has to be aggregated separately, and thus the size of the aggregated reports will be $2048 * l + |RA| + |GW| + |TS| + 160$. In Fig. 5, we further plot the communication overhead in terms of user number w and data type l. It is shown that the EPPA scheme significantly reduces the communication overhead of the GW-to-OA communication.

From the above analysis, the proposed EPPA scheme is indeed efficient in terms of computation and communication cost, which is suitable for the real-time high-frequency data collection in smart grid communications.

7 Related Works

Since large volumes of data from users are to be reported to the OA, it is essential to aggregate individual users' data at intermediate nodes for reducing communication overhead. In most existing secure data aggregation schemes, an intermediate needs to decrypt the received data, aggregate them using aggregation functions, and then encrypt the aggregated result before forwarding it. This process is fairly expensive and risky when intermediate nodes are not trusted. Castelluccia et al. [25] adopted homomorphic encryption techniques to enable efficient aggregation of encrypted data without decryption at

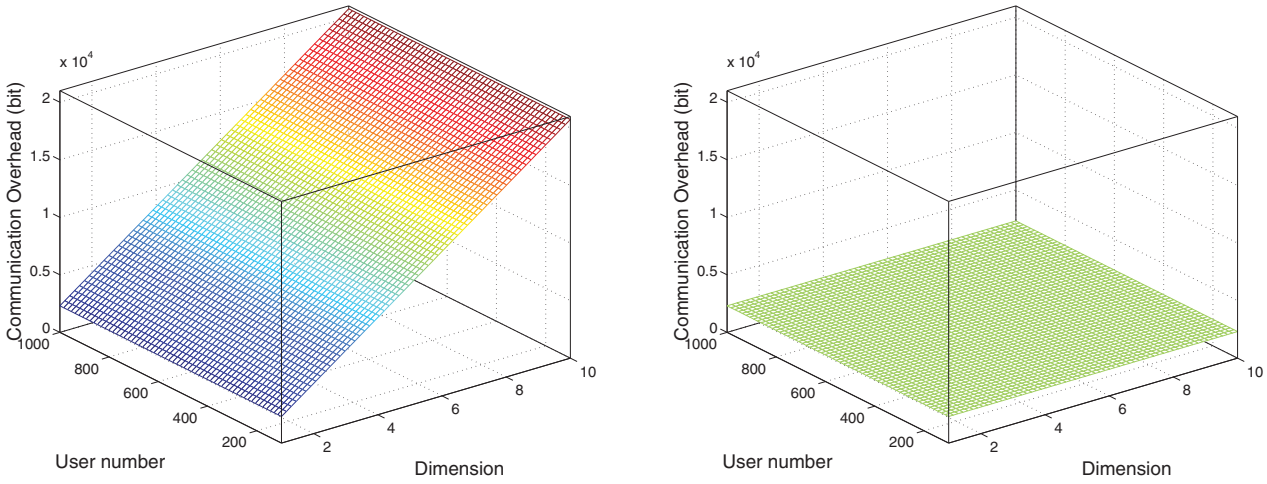
This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication.



(a) TRAD user-to-GW

(b) EPPA user-to-GW

Fig. 4. User-to-GW communication overhead



(a) TRAD GW-to-OA

(b) EPPA GW-to-OA

Fig. 5. GW-to-OA communication overhead

intermediate nodes. The proposed scheme is very promising and has triggered considerable followup research [15], [16], [17], as discussed below.

Castelluccia et al. [15] extended [25] and presented a provable secure and efficient encryption and aggregation scheme. The scheme requires a small number of single-precision additions. It expands packet size only by a small number of bits in the encryption operation and therefore improves computational and communication efficiency. Westhoff et al. [16] indicated that [25] uses different keys per node at the cost of mandatory transmitting of the ID list of the encrypting nodes, resulting in an increased message overhead per monitoring node. They proposed a key pre-distribution scheme that suits the end-to-end encryption of reverse multicast traffic in sensor networks. By using their scheme, a symmetric homomorphic encryption can be applied to increase the efficiency, robustness and flexibility of data aggregation in sensor networks. Shi et al. [17] adopted [25] as a tool, and required users to slice their data into pieces and then to collaboratively aggregate the pieces with others' to preserve user privacy. These previous research works focus on one-dimensional data. How to aggregate multi-dimensional data remains to be a challenging issue.

In smart grid, each individual user has multiple dimensional electricity usage data which is small in size. Performing homomorphic encryption on each dimension of the data requires large computational efforts, and results in large-sized ciphertext and then unaffordable communication cost. Lin et al. [26] introduced super-increasing sequence and perturbation techniques into *compressed data aggregation*. In the aggregation process, a super-increasing sequence is initialized and used

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.

to integrate multi-dimensional data as one single piece in the plaintext space. The operations of homomorphic encryption will not corrupt the data structure, i.e., the data in different dimensions are mixed. After receiving the aggregated data, the receiver performs a single decryption and takes several multiplications to recovery the data from the plaintext. This scheme is suitable for the scenario that the data to be encrypted is much smaller than the plaintext space.

The algorithm from [26] assumes that a symmetric key is shared between a sender and a receiver at the initialization phase. In smart grid, thousands of users may communicate with a single gateway. It is not practical to deploy and manage thousands of keys for users and the gateway. In addition, once some residential users stop working and cannot report their data during some periods, the OA cannot use the proper shared keys to recover the correct data since he doesn't know the participants due to user privacy. As a result, the reliability becomes a big challenge, and this solution cannot be directly be applied. In this paper, we have designed a compressed data aggregation scheme under the public key infrastructure, where users encrypt their reports only with the public key of the OA, and the OA can decrypt the reports only with his private key. Therefore, compared to the symmetric key algorithm in [26], our approach can not only reduce considerable initialization efforts but also achieve high reliability.

8 Conclusions

In this paper, we have proposed an efficient and privacy-preserving aggregation scheme (EPPA) for secure smart grid communications. It realizes a multi-dimensional data aggregation approach based on the homomorphic Paillier cryptosystem. Compared with the traditional one-dimensional data aggregation methods, EPPA can significantly reduce computational cost and significantly improve communication efficiency, satisfying the real-time high-frequency data collection requirements in smart grid communications. We have also provided security analysis to demonstrate its security strength and privacy-preserving ability, and performance analysis to show the efficiency improvement. For the future work, we will study the possible behavior by internal attackers and extend the EPPA scheme to effectively resist such attacks.

References

[1] "Blackout 2003," http://www.ieso.ca/imoweb/EmergencyPrep/blackout2003.

[2] G. W. Arnold, "Challenges and opportunities in smart grid: A position article," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 922–927, 2011.

[3] F. Li, W. Qiao, H. Sun, H. Wan, J. Wang, Y. Xia, Z. Xu, and P. Zhang, "Smart transmission grid: Vision and framework," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 168–177, 2010.

[4] K. Moslehi and R. Kumar, "A reliability perspective of the smart grid," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 57–64, 2010.

[5] D. Niyato, L. Xiao, and P. Wang, "Machine-to-machine communications for home energy management system in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 53–59, 2011.

[6] Z. M. Fadlullah, M. M. Fouda, N. Kato, A. Takeuchi, N. Iwasaki, and Y. Nozaki, "Toward intelligent machine-to-machine communications in smart grid," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 60–65, 2011.

[7] H. Liang, B. Choi, W. Zhuang, and X. Shen, "Towards optimal energy store-carry-and-deliver for phevs via v2g system," in *Proc. IEEE INFOCOM'12*, Orlando, Florida, USA, March 25-30 2012.

[8] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," *IEEE Security & Privacy*, vol. 7, no. 3, pp. 75–77, 2009.

[9] Z. M. Fadlullah, M. M. Fouda, X. Shen, Y. Nozaki, and N. Kato, "An early warning system against malicious activities for smart grid communications," *IEEE Network Magazine*, to appear.

[10] M. M. Fouda, Z. M. Fadlullah, N. Kato, R. Lu, and X. Shen, "A lightweight message authentication scheme for smart grid communications," *IEEE Transactions on Smart Grid*, to appear.

[11] M. He and J. Zhang, "A dependency graph approach for fault detection and localization towards secure smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 342–351, 2011.

[12] Y. Yuan, Z. Li, and K. Ren, "Modeling load redistribution attacks in power systems," *IEEE Transactions on Smart Grid*, vol. 2, no. 2, pp. 382–390, 2011.

[13] R. Lu, X. Li, X. Lin, X. Liang, and X. Shen, "GRS: The green, reliability, and security of emerging machine to machine communications," *IEEE Communications Magazine*, vol. 49, no. 4, pp. 28–35, 2011.

[14] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *EUROCRYPT*, 1999, pp. 223–238.

[15] C. Castelluccia, A. C.-F. Chan, E. Mykletun, and G. Tsudik, "Efficient and provably secure aggregation of encrypted data in wireless sensor networks," *TOSN*, vol. 5, no. 3, 2009.

[16] D. Westhoff, J. Girão, and M. Acharya, "Concealed data aggregation for reverse multicast traffic in sensor networks: Encryption, key distribution, and routing adaptation," *IEEE Transactions on Mobile Computing*, vol. 5, no. 10, pp. 1417–1431, 2006.

[17] J. Shi, R. Zhang, Y. Liu, and Y. Zhang, "Prisense: Privacy-preserving data aggregation in people-centric urban sensing systems," in *Infocom*, 2010, pp. 758–766.

[18] D. Boneh and M. K. Franklin, "Identity-based encryption from the weil pairing," in *CRYPTO*, 2001, pp. 213–229.

[19] Y. Sang, H. Shen, and H. Tian, "Privacy-preserving tuple matching in distributed databases," *IEEE Transactions on Knowledge and Data Engineering*, vol. 21, no. 12, pp. 1767–1782, 2009.

[20] S. Zhong, "Privacy-preserving algorithms for distributed mining of frequent itemsets," *Inf. Sci.*, vol. 177, no. 2, pp. 490–503, 2007.

[21] D. Boneh, B. Lynn, and H. Shacham, "Short signatures from the weil pairing," *Journal of Cryptology*, vol. 17, no. 4, pp. 297–319, 2004.

[22] M. Bellare and P. Rogaway, "Random oracles are practical: A paradigm for designing efficient protocols," in *ACM Conference on Computer and Communications Security*, 1993, pp. 62–73.

[23] B. Lynn, "PBC library," http://crypto.stanford.edu/pbc/.

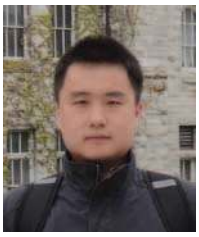
[24] "Multiprecision integer and rational arithmetic c/c++ library," http://www.shamus.ie/.

[25] C. Castelluccia, E. Mykletun, and G. Tsudik, "Efficient aggregation of encrypted data in wireless sensor networks," in *MobiQuitous*, 2005, pp. 109–117.

[26] X. Lin, R. Lu, and X. Shen, "MDPA: Multidimensional privacy-preserving aggregation scheme for wireless sensor networks," *Wireless Communications and Mobile Computing (Wiley)*, vol. 10, no. 6, pp. 843–856, 2010.

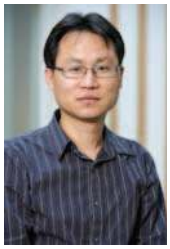
Rongxing Lu (S'09-M'11) is currently working toward a Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He is currently a research assistant with the Broadband Communications Research (BBCR) Group, University of Waterloo. His research interests include wireless network security, applied cryptography, and trusted computing.

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.



Xiaohui Liang (S'10) is currently working toward a Ph.D. degree with the Department of Electrical and Computer Engineering, University of Waterloo, Canada. He is currently a research assistant with the Broadband Communications Research (BBCR) Group, University of Waterloo. His research interests include wireless network security, applied cryptography, and e-healthcare system.

Xu Li received a PhD (2008) degree from Carleton University, Canada, an MSc (2005) degree from the University of Ottawa, Canada, and a BSc (1998) degree from Jilin University, China, all in computer science. Before joining Inria as a full researcher in 2011, he held PDF positions at several locations: the University of Waterloo, Canada; Inria/CNRS, France; and the University of Ottawa, Canada. He has published nearly five dozen refereed papers at premium venues such as IEEE JSAC, TPDS, TMC, TC, TAC, ComMag, Infocom, MASS, SECON, etc. Part of these works was directly funded by the Natural Sciences and Engineering Research Council of Canada (NSERC) through doctoral scholarships and post-doctoral fellowships programs. His current research interests are in the broad area of machine-to-machine communications, covering a variety of issues such as data communications, mobility management, QoS provisioning, network security and trust management. He is on the editorial board of the European Transactions on Telecommunications, Ad Hoc & Sensor Wireless Networks, and Parallel and Distributed computing and Networks. He is/was a guest editor of a few international journals.



Xiaodong Lin (S'07-M'09) received the Ph.D. degree in information engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 1998 and the Ph.D. degree (with Outstanding Achievement in Graduate Studies Award) in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2008. He is currently an assistant professor of information security with the Faculty of Business and Information Technology, University of Ontario Institute of Technology, Oshawa, ON, Canada. His research interests include wireless network security, applied cryptography, computer forensics, and software security. Dr. Lin was the recipient of a Natural Sciences and Engineering Research Council of Canada (NSERC) Canada Graduate Scholarships (CGS) Doctoral and the Best Paper Awards of the IEEE International Conference on Computer Communications and Networks (ICCCN 2009) and the IEEE International Conference on Communications (ICC 2007) - Computer and Communications Security Symposium.

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. He is a Professor and University Research Chair, Department of Electrical and Computer Engineering, University of Waterloo, Canada. Dr. Shen's research focuses on resource management in interconnected wireless/wired networks, UWB wireless communications networks, wireless network security, wireless body area networks and vehicular ad hoc and sensor networks. He is a co-author of three books, and has published more than 400 papers and book chapters in wireless communications and networks, control and filtering. Dr. Shen has served as the Technical Program Committee Chair for IEEE VTC'10, the Tutorial Chair for IEEE ICC'08, the Technical Program Committee Chair for IEEE Globecom'07, the General Co-Chair for Chinacom'07 and QShine'06, the Founding Chair for IEEE Communications Society Technical Committee on P2P Communications and Networking. He has also served as a Founding Area Editor for IEEE Transactions on Wireless Communications; Editor-in-Chief for Peer-to-Peer Networking and Application; Associate Editor for IEEE Transactions on Vehicular Technology; Computer Networks; and ACM/Wireless Networks, Guest Editor for IEEE JSAC, IEEE Wireless Communications, IEEE Communications Magazine, and ACM Mobile Networks and Applications, etc. Dr. Shen received the Excellent Graduate Supervision Award in 2006, and the Outstanding Performance Award in 2004 and 2008 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, and a Distinguished Lecturer of IEEE Communications Society.