An Elliptic Curve-Based Scalable Data Aggregation Scheme for Smart Grid

Yuwen Chen, José-Fernán Martínez-Ortega, Pedro Castillejo, and Lourdes López

***Abstract*—In a smart grid, a smart meter reports its real-time electricity consumption data to utility supplier periodically, the utility supplier adjusts its supplement accordingly. However, real-time electricity consumption data may reveal the inhabitants' behaviors, thus smart grid data aggregation schemes are introduced to protect user's privacy, which enables the utility supplier to get the total consumption data rather than a single meter's electricity consumption data. However, existing solutions do not solve all the problems, some schemes are not scalable, once the system is deployed, it is unable to add new meters to the system, in some schemes, the electricity consumption data are encrypted using the utility supplier's public key, if the utility supplier accidentally obtains the encrypted data, it is able to decrypt the data. For these reasons, we designed a scalable elliptic-based multidimensional data aggregation scheme for the smart grid, an independent key is assigned to a meter, even if the utility supplier is unable to decrypt a single meter's encrypted electricity consumption data. Besides, the proposed scheme is scalable, it is easy to add or remove a meter from the system, and the proposed scheme enables a meter to report multiple types of data at a time.**

***Index Terms*—Data privacy, data security, elliptic curves, scalability, smart grids.**

I. Introduction

SMART meters for electricity and for gas have been largely deployed in Europe [1], [2], which brings two-way communication between smart meters and utility supplier. Smart meters report their real-time consumption data and other information to the utility supplier. However, some advanced techniques have shown the possibility for an adversary to get people's private information by analyzing the real-time electricity consumption data [3]–[5]. As real-time electricity consumption data can leak the owner's personal behaviors, it is necessary to protect a user's real-time electricity consumption data, thus smart grid data aggregation schemes are proposed, they enable the utility supplier to get the total consumption data, however, the utility supplier is unable to get a single smart meter's consumption data. In this way, the utility supplier can adjust its supplement dynamically according to the total consumption data, and users' privacy is protected.

Even though existing schemes can prevent users' real-time electricity consumption data from being disclosed to the adversary or the utility supplier. However, there are problems that have not been solved. First, in some of these schemes, users' real-time electricity consumption data are encrypted by the utility supplier's public key, if the aggregator and the utility supplier collude, the aggregator sends a user's encrypted electricity consumption data to the utility supplier, the utility supplier can get a user's real-time electricity consumption data by decrypting this data using its private key. Besides, as the encrypted electricity consumption data are sent in a public channel, if the utility supplier gets the data by eavesdropping, the utility supplier can decrypt the encrypted electricity consumption data using his private key.

Second, in some schemes, a trusted third party is introduced, which generates a series of keys for smart meters to encrypt their real-time electricity consumption data, thus the utility supplier cannot get a single smart meter's consumption data. However, these schemes have a scalability problem. After the system is deployed, it is impossible to add a new smart meter to the system, or if some of the deployed smart meters are broken, the whole system cannot work correctly.

Third, some schemes only enable meters to report one type of data, in fact, data of smart grid are likely to be multidimensional (time, electricity consumption purpose, and so on), if the utility supplier wants to have an in-depth analysis of these data, he needs as much data as he can.

For these reasons, we come up with an elliptic curve-based scalable data aggregation scheme. The proposed scheme is inspired by the study of Hao *et al.* [6], [7], their methods were initially designed for the anonymous voting system. Our contributions are mainly reflected in the following aspects.

1) Smart meters encrypt their electricity consumption data using an independent key, the utility supplier is not able to get a single meter's consumption data, thus the proposed scheme can withstand the collude attack or the eavesdropping attack. The security of the proposed scheme is based on the computational hardness of the bilinear computational Diffie–Hellman problem.
2) The proposed scheme is scalable, even if the system is deployed, it is easy to add new meters to the system or to remove meters out of the system dynamically, the system works normally after the entities in the system conduct a series of actions.
3) The proposed scheme enables meters to report multiple types of data to the utility supplier, as a result, the utility supplier can conduct an in-depth analysis of these data. Besides, the proposed scheme is more efficient in

Manuscript received March 18, 2019; revised September 5, 2019; accepted November 12, 2019. Date of publication December 4, 2019; date of current version June 3, 2020. This work was supported in part by the I3RES (ICT-based Intelligent management of Integrated RES for the Smart Grid optimal operation), in part by the e-GOTHAM (Sustainable-Smart Grid Open System for the Aggregated Control, Monitoring, and Management of Energy), and in part by the Chinese Scholarship Council with File No: 201507040027. *(Corresponding author: Yuwen Chen.)*

The authors are with the Departamento de Ingeniería Telemática y Electrónica, Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación, Universidad Politécnica de Madrid, Madrid 28031, Spain (e-mail: yuwen.chen@upm.es; jf.martinez@upm.es; pedro.castillejo@upm.es; lourdes.lopez@upm.es).

Digital Object Identifier 10.1109/JSYST.2019.2954080

computation cost and communication cost compared to related works.

II. Related Work

Lu *et al.* proposed a data aggregation scheme: EPPA [8], which is based on the homomorphic Paillier cryptosystem [9], a super-increasing sequence is used to enable a meter to report multiple types of data to the utility supplier. Li *et al.* proposed a similar scheme, their scheme enables the aggregations of electricity consumption data of different ranges, which makes it easy to meet the fine-grained demands [10]. The scheme of Lu *et al.* enables the aggregations of electricity consumption data of two ranges [11]. Lu *et al.* used the Chinese remainder theory to ensure the reporting of multiple types of data [12]. Shen *et al.* used the Horner's rule method to ensure a meter to report multiple types of data [13]. The MuDA methods [14] by Chen *et al.* enables meters to report multiple types of data, and statistical analysis of these data is enabled. The scheme of Ge *et al.*enables a utility supplier to learn the statistical of the consumption data in a more efficient way [15].

Boudia *et al.* [16] proposed an elliptic curve-based secure multidimensional aggregation scheme, the control center generates a series of public keys, meters using these public keys to encrypt their consumption data. Fan *et al.* [17] proposed a privacy-enhanced data aggregation scheme, a trusted third party is introduced, this trusted third party will generate a series of "blinding factors": $\pi_0, \pi_1, \ldots, \pi_n$, π_0 is sent to the aggregator, and π_n is sent to the corresponding smart meter, a meter encrypts its consumption data using π_n, the aggregator decrypts the total consumption data using π_0. A drawback of this scheme is that it is not scalable, it is unable to add a new meter into the system or remove a meter from the system. The scheme of He *et al.* has the similar problem [18]. Wang [19] proposed an identity-based data aggregation protocol, their protocol is based on an identity-based encryption scheme and an identity-based signature scheme, as these two schemes share the same private and public parameters, the computation cost is reduced. However, in their scheme, the private keys of all the entitites are generated by the system, which means the system can decrypt the information which are encrypted using these public keys, the information encrypted using these public keys are transparent to the system, besides, as a smart meter's data are encrypted using the electricity service provider's public key, if the electricity service provider accidently get a meter's encrypted electricity consumption, it is able to know the meter's electricity consumption data. In the scheme of Jia *et al.* [20], they permit a meter itself to randomly divide its readings into *m* shares, and these shares are encrypted using its private keys, only the aggregator with the right key can recover the consumption data, however, their scheme will fail to aggregate in present of malfunctioning meters. One of the methods of Kursawe *et al.* [21] is similar to the proposed scheme. However, in their scheme, for meters to report their electricity consumption data, they have to negotiate a key with the rest of the meters. It requires too much computation cost and communication cost to generate the round specific key, which is not worthy.

The "PDAFT" method of Chen *et al.* [22] is based on the Paillier homomorphic cryptosystem, the system is able to solve the meter failure problem to some extent, a trusted authority is required to generate keys for meters to encrypt their consumption data, these keys are stored at the trusted authority. When some meters fail to work correctly, the trusted authority is required to provide dummy ciphertexts related to meters, as the trusted authority knows the private keys of the meters, so it can accomplish this task. In this way, their scheme works correctly when some meters fail, however, their scheme does not solve the privacy protection problem completely, as the trusted authority knows the private keys of the meters. In the scheme of Ni *et al.*, a trusted authority generates the private keys for meters, when some meters fail, the utility supplier is able to recover the sum of the rest of the smart meters [23], however, the meters do not have private key privacy. In the work of Bao and Lu [24], a control center stores all the public keys of the meters, thus if some meters fail to work, the control center can still get the total electricity consumption data, their solution does not solve the scalability problem completely, as it is unable to add new smart meters into the system. In their scheme, the electricity consumption data are encrypted using the gateway's public key and the control center's public key together, thus even if the control center get the encrypted data, it is unable to decrypt the data. The scheme of Shi *et al.* is a group-based aggregation scheme [25], their scheme alleviates the meter failure problem to some extent, as they just abandon the data from the group with broken meters, there is an error in the final aggregation data.

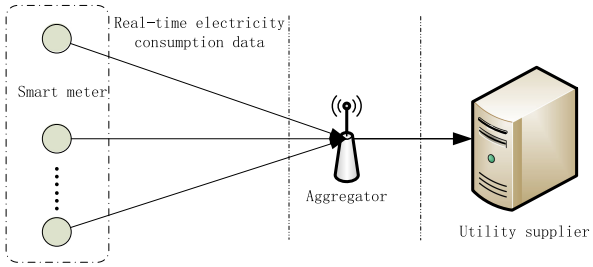


Fig. 1. System model.

The works of Erkin *et al.* [26] and Si *et al.* [27] provide a good summary of the most popular methods to protect a user's privacy in the smart grid data aggregation scenario. Some other approaches have been discussed, too, for example, distributed load scheduling method of Liu *et al.* [28], homomorphic cryptosystem-based methods of García and Jacobs [29], Busom *et al.* [30], and Ruj *et al.* [31], lattice-based homomorphic method of Abdallah and Shen [32], [33], and the noise addition method of Bohli *et al.* [34], He *et al.* [35], and Barbosa *et al.* [36].

III. System Model

The system in this article is a three-layer system as depicted in Fig. 1. There are three types of entities in the system: smart meters, aggregators, and utility supplier. Meters are divided into groups, meters in one aggregation area form a group. A meter retrieves its own electricity consumption data, encrypts it using an independent key and sends this encrypted data to the aggregator. When the aggregator collects the encrypted electricity consumption data of all the meters, it adds them together and sends the sum to the utility supplier. Utility supplier can retrieve the total consumption data using its key. In this way, the utility supplier will obtain the total electricity consumption data, but it is unable to know the consumption data of an individual smart meter.

In addition, the proposed scheme can be applied to two-layer systems, in which meters report their consumption data to the utility supplier directly.

TABLE I
SYMBOLS USED IN THE PROPOSED SCHEME

Symbol	Description
G_1, G_T	Multiplicative group
$e: G_1 \times G_1 \to G_T$	A bilinear map pairing
P	A generator of G_1
g	A generator of G_T
(M_i, id_i)	i^{th} meter and its identity
(A_j, id_j)	j^{th} aggregator and its identity
(d_i, R_i)	The public-private key pair of meter M_i
(d_j, R_j)	The public-private key pair of j^{th} aggregator
(d_x, R_x)	The public-private key pair of utility supplier
$m_{i1}, m_{i2},\ldots,m_{ik}$	k^{th} type of data of meter M_i
$\|$	String connection
R_{si}	i^{th} meter's key to encrypt its consumption data
R_{sx}	Utility supplier's key to decrypt the total consumption data
$c_i = (c_{i1}, \ldots, c_{ik})$	c_i is the encrypted report data of meter M_i, c_{ik} is the k^{th} type of data of meter M_i
$c_j = (c_{j1}, \ldots, c_{jk})$	c_j is the sum of the encrypted report data of all the meters, c_{jk} is the sum of the k^{th} type of data
C_{uk}	The decrypted sum of the k^{th} type of data
S_i	A signature of meter M_i
S_j	A signature of j^{th} aggregator

IV. PROPOSED SCHEME

In this section, we introduce the proposed smart meter data aggregation scheme, Table I lists the notions that will be used.

A. System Initialization

The utility supplier generates a multiplicative group G_1, let P be a random generator of G_1, g be a random generator of G_T, $e : G_1 \times G_1 \to G_T$ be a bilinear map pairing. Afterward, it publishes the parameters to all entities in the system.

B. Registration Phase

Suppose the registration messages are sent in private and secure channels, which means an adversary is unable to launch attacks during the registration phase. The registration phase mainly consists of the following steps.

Step 1: Utility supplier selects a random $d_x \in Z_n^+$ as its private key, the corresponding public key is $R_x = d_x \cdot P$, utility supplier sends R_x to the aggregator.

Step 2: Meter M_i selects a random $d_i \in Z_n^+$ as its private key, the public key is set as $R_i = d_i \cdot P$, M_i sends its identity id_i and public key R_i to the aggregator.

Step 3: Aggregator selects a random $d_j \in Z_n^+$ as its private key, the public key is $R_j = d_j \cdot P$. When the aggregator receives all meters' public keys, it calculates $R_{ci} = (\sum_{j=1}^{i-1} R_j - \sum_{j=i+1}^{n} R_j - R_x)$ for meter M_i, and sends R_{ci} to M_i, M_i computes $R_{si} = d_i \cdot R_{ci}$ as its key to encrypt the real-time electricity consumption data.

Step 4: Aggregator calculates $R_A = \sum_{i=1}^{n} R_i$ and sends R_A, id_j, and R_j to the utility supplier. Utility supplier's key to decrypt the total electricity consumption data is $R_{sx} = d_x \cdot R_A = d_x \cdot \sum_{i=1}^{n} R_i$.

Note that, after registration, the entities can build shared keys for two-way communication using the elliptic curve Diffie–Hellman key exchange protocol.

C. Multiple Data Reporting

At the start of each reporting cycle, meter M_i gets the data $m_{i1}, m_{i2}, \ldots, m_{ik}$, encrypts them using its key and generates a signature on the encrypted data. The detailed process is described as follows.

1) Meter M_i extracts its data $m_{i1}, m_{i2} \ldots m_{ik}$.
2) Meter M_i gets the current timestamp t_i.
3) Meter M_i computes a key $g_i = e(H(t_i), R_{si})$.
4) Meter M_i encrypts $m_{i1}, m_{i2}, \ldots, m_{ik}$ as $c_{i1} = g_i \cdot g^{m_{i1}}$, $c_{i2} = g_i \cdot g^{m_{i2}}, \ldots, c_{ik} = g_i \cdot g^{m_{ik}}$, and $c_i = (c_{i1}, c_{i2}, \ldots, c_{ik})$.
5) Meter M_i generates a signature $S_i = d_i \cdot H(c_i, id_i, t_i)$.
6) Meter M_i sends $\{c_i, S_i, t_i, id_i\}$to the aggregator.

When aggregator receives $\{c_i, S_i, t_i, id_i\}$, it first checks the timestamp t_i and the signature S_i. Then, it adds all c_i. Finally, aggregator reports these data to the utility supplier.

1) Aggregator checks the timestamp t_i.
2) Aggregator checks if $e(S_i, P) = e(H(c_i, id_i, t_i), R_i)$, if this equation holds, it accepts the message, otherwise, it abandons this message.
3) Aggregator computes $c_{j1} = \prod_{i=1}^{n} c_{i1}$, $c_{j2} = \prod_{i=1}^{n} c_{i2}$, $\ldots, c_{jk} = \prod_{i=1}^{n} c_{ik}$ and $c_j = (c_{j1}, c_{j2}, \ldots, c_{jk})$.
4) Aggregator generates a signature $S_j = d_j \cdot H(c_j, \mathrm{id}_j, t_i)$.
5) Aggregator sends $\{c_j, S_j, t_i, \mathrm{id}_j\}$ to the utility supplier.

After the utility supplier gets $\{c_j, S_j, t_i, \mathrm{id}_j\}$, it checks the timestamp t_i and the signature S_j to verify the validity of this message. Afterward, it decrypts c_j with its key.

1) Utility supplier checks the timestamp t_i.
2) Utility supplier checks if $e(S_j, P) = e(H(c_j, id_j, t_i), R_j)$, if the equation holds, it accepts the message.
3) Utility supplier computes $g_x = e(H(t_i), R_{sx})$ and $C_{u1} = g_x \cdot c_{j1} = g^{\sum_{i=1}^{n} m_{i1}}$,$C_{u2} = g_x \cdot c_{j2} = g^{\sum_{i=1}^{n} m_{ik}}, \ldots,$ $C_{uk} = g_x \cdot c_{jk} = g^{\sum_{i=1}^{n} m_{ik}}$.
4) Utility supplier uses the Pollard's lambda method to get $\sum_{i=1}^{n} m_{ik} = \log_g C_{uk}$.

D. Proof of the Correctness

First, we prove the equation $e(S_i, P) = e(H(c_i, id_i, t_i), R_i)$, if the equation holds, the signature scheme used in this article is correct

$$\begin{aligned} e(S_i, P) &= e(d_i \cdot H(c_i, id_i, t_i), P) \\ &= e(H(c_i, id_i, t_i), d_i \cdot P) \\ &= e(H(c_i, id_i, t_i), R_i). \end{aligned}$$

To make the verification process more efficient, an aggregator can conduct a batch verification on the signatures, thus the pairing operations can be reduced from 2*n* to (*n* + 1). $\partial_i \in Z_n^+$ are short length random numbers

$$\begin{aligned} e\left(\sum_{i=1}^{n} S_i^{\partial_i}, P\right) &= e\left(\sum_{i=1}^{n} (d_i \cdot H(c_i, id_i, t_i))^{\partial_i}, P\right) \\ &= \prod_{i=1}^{n} e(d_i \cdot H(c_i, id_i, t_i), P)^{\partial_i} \\ &= \prod_{i=1}^{n} e(H(c_i, id_i, t_i), d_i \cdot P)^{\partial_i} \\ &= \prod_{i=1}^{n} e\left(H(c_i, id_i, t_i), R_i^{\partial_i}\right). \end{aligned}$$

TABLE II
EQUALITY OF THE EQUATION

	d_1	d_2	d_3	d_4
d_1		$-d_1 \cdot d_2 \cdot P$	$-d_1 \cdot d_3 \cdot P$	$-d_1 \cdot d_4 \cdot P$
d_2	$+d_2 \cdot d_1 \cdot P$		$-d_2 \cdot d_3 \cdot P$	$-d_2 \cdot d_4 \cdot P$
d_3	$+d_3 \cdot d_1 \cdot P$	$+d_3 \cdot d_2 \cdot P$		$-d_3 \cdot d_4 \cdot P$
d_4	$+d_4 \cdot d_1 \cdot P$	$+d_4 \cdot d_2 \cdot P$	$+d_4 \cdot d_3 \cdot P$	

Second, we prove $\sum_{i=1}^{n}(d_i \cdot \sum_{j=1}^{i-1} R_j - d_i \cdot \sum_{j=i+1}^{n} R_j) = 0$, Table II illustrates the equality in a more intuitive way, suppose $n = 4$

$$\begin{aligned}
&\sum_{i=1}^{n}\left(d_i \cdot \sum_{j=1}^{i-1} R_j - d_i \cdot \sum_{j=i+1}^{n} R_j\right) \\
&\quad = \sum_{i=1}^{n}\left(d_i \cdot \sum_{j=1}^{i-1}(d_j \cdot P) - d_i \cdot \sum_{j=i+1}^{n}(d_j \cdot P)\right) \\
&\quad = \left(\sum_{i=1}^{n}\left(d_i \cdot \sum_{j=1}^{i-1} d_j - d_i \cdot \sum_{j=i+1}^{n} d_j\right)\right) \cdot P \\
&\quad = \left(\sum_{i=1}^{n}\left(d_i \cdot \sum_{j<i} d_j\right) - \sum_{i=1}^{n}\left(d_i \cdot \sum_{j>i} d_j\right)\right) \cdot P = 0.
\end{aligned}$$

Third, we prove $\sum_{i=1}^{n} R_{si} = -\sum_{i=1}^{n}(d_i \cdot d_x \cdot P)$ in the following way:

$$\begin{aligned}
\sum_{i=1}^{n} R_{si} &= \sum_{i=1}^{n}(d_i \cdot R_{ci}) \\
&= \sum_{i=1}^{n}\left(d_i \cdot \left(\sum_{j=1}^{i-1} R_j - \sum_{j=i+1}^{n} R_j - R_x\right)\right) \\
&= \sum_{i=1}^{n}\left(d_i \cdot \left(\sum_{j=1}^{i-1} R_j - \sum_{j=i+1}^{n} R_j\right)\right) - \sum_{i=1}^{n}(d_i \cdot R_x) \\
&= -\sum_{i=1}^{n}(d_i \cdot d_x \cdot P).
\end{aligned}$$

Then, we can prove $d_x \cdot R_A + \sum_{i=1}^{n} R_{si} = 0$ easily

$$\begin{aligned}
d_x \cdot R_A + \sum_{i=1}^{n} R_{si} &= d_x \cdot \sum_{i=1}^{n} R_i - \sum_{i=1}^{n}(d_i \cdot d_x \cdot P) \\
&= \sum_{i=1}^{n}(d_x \cdot d_i \cdot P) - \sum_{i=1}^{n}(d_i \cdot d_x \cdot P) = 0.
\end{aligned}$$

Until now, we can prove $C_{uk} = g^{\sum_{i=1}^{n} m_{ik}}$

$$\begin{aligned}
C_{uk} &= g_x \cdot \prod_{i=1}^{n} c_{ik} = e\left(H\left(t_i\right), R_{sx}\right) \cdot \prod_{i=1}^{n}\left(g_i \cdot g^{m_{ik}}\right) \\
&= e\left(H\left(t_i\right), d_x \cdot R_A\right) \cdot \prod_{i=1}^{n} g_i \cdot \prod_{i=1}^{n} g^{m_{ik}} \\
&= e\left(H\left(t_i\right), d_x \cdot R_A\right) \cdot \prod_{i=1}^{n} e\left(H\left(t_i\right), R_{si}\right) \cdot g^{\sum_{i=1}^{n} m_{ik}} \\
&= e\left(H\left(t_i\right), d_x \cdot R_A\right) \cdot e\left(H\left(t_i\right), \sum_{i=1}^{n} R_{si}\right) \cdot g^{\sum_{i=1}^{n} m_{ik}} \\
&= e\left(H(t_i), d_x \cdot R_A + \sum_{i=1}^{n} R_{si}\right) \cdot g^{\sum_{i=1}^{n} m_{ik}} = g^{\sum_{i=1}^{n} m_{ik}}.
\end{aligned}$$

E. Dynamic Leave

The proposed scheme supports dynamic leave, if m smart meters $M_{r1}, M_{r2}, \ldots, M_{rk}$ with public key $R_{r1}, R_{r2}, \ldots, R_{rk}$ and private key $d_{r1}, d_{r2}, \ldots, d_{rt}$ have to be removed from the system, the system still works correctly after the leaving of these meters. Suppose $\sum^{R_{rk}}$ is the sum of $R_{r1}, R_{r2}, \ldots, R_{rk}$, and $\sum^{R_i}$ is the sum of the public keys of the meters remaining in the system, then $\sum R_i = \sum_{i=1}^{n} R_i - \sum R_{rk}$. Following steps must be conducted to ensure the system to work correctly.

1) Aggregator computes $R'_A = R_A - \sum R_{rk}$ and sends R'_A to the utility supplier, utility supplier updates $R_A = R'_A$.
2) For a meter M_i remains in the system, aggregator computes $\sum_{rk<i} R_{rk} - \sum_{rk>i} R_{rk}$ and sends it to M_i, M_i updates $R'_{ci} = R_{ci} - (\sum_{rk<i} R_{rk} - \sum_{rk>i} R_{rk})$. $\sum_{rk<i} R_{rk}$ is the sum of the public keys of meter M_{rk} whose subscript is smaller than the ith meter in the original system, $\sum_{rk>i} R_{rk}$ is the sum of the public keys of meter M_{rk} whose subscript is larger than ith meter in the original system.

Suppose $\sum_{j=1}^{i-1} R_j - \sum_{j=i+1}^{n} R_j - (\sum_{rk<i} R_{rk} - \sum_{rk>i} R_{rk}) = T$, then we can get $R'_{ci} = R_{ci} - (\sum_{rk<i} R_{rk} -) = \sum_{j=1}^{i-1} R_j - \sum_{j=i+1}^{n} R_j - (\sum_{rk<i} R_{rk} - \sum_{rk>i} R_{rk}) - R_x = T - R_x$. The sum of $d_i \cdot R'_{ci}$ of the meters remain in the system is $\sum(d_i \cdot R'_{ci}) = \sum(d_i \cdot (T - R_x)) = d_i \cdot \sum T - \sum(d_i \cdot R_x)$.

Suppose $\sum_{j<i} R_j = \sum_{j=1}^{i-1} R_j - \sum_{rk<i} R_{rk}$ is the sum of the public keys of the meters remaining in the system whose subscripts are smaller than the ith meter in the original system, $\sum_{j>i} R_j = \sum_{j=i+1}^{n} R_j - \sum_{rk>i} R_{rk}$ is the sum of the public keys of the meters remaining in the system whose subscripts are larger than the ith meter in the original system

$$\begin{aligned}
&d_i \cdot \sum T \\
&= d_i \cdot \sum\left(\sum_{j=1}^{i-1} R_j - \sum_{j=i+1}^{n} R_j - \left(\sum_{rk<i} R_{rk} - \sum_{rk>i} R_{rk}\right)\right) \\
&= d_i \cdot \sum\left(\left(\sum_{j=1}^{i-1} R_j - \sum_{rk<i} R_{rk}\right) - \left(\sum_{j=i+1}^{n} R_j - \sum_{rk>i} R_{rk}\right)\right) \\
&= \sum\left(d_i \cdot \left(\sum_{j<i} R_j - \sum_{j>i} R_j\right)\right)
\end{aligned}$$

TABLE III
EQUALITY OF THE EQUATION OF DYNAMIC LEAVE

	d_1	d_4	d_x
d_1		$-d_1 \cdot d_4 \cdot P$	$-d_1 \cdot d_x \cdot P$
d_4	$+d_4 \cdot d_1 \cdot P$		$-d_4 \cdot d_x \cdot P$
d_x	$+d_x \cdot d_1 \cdot P$	$+d_x \cdot d_4 \cdot P$	

$$= \sum \left(d_i \cdot \left(\sum_{j<i} d_j - \sum_{j>i} d_j \right) \cdot P \right)$$

$$= \left(\sum \left(d_i \cdot \sum_{j<i} d_j \right) - \sum \left(d_i \cdot \sum_{j>i} d_j \right) \right) \cdot P.$$

As $(\sum_{i=1}^{n}(d_i \cdot \sum_{j<i} d_j) - \sum_{i=1}^{n}(d_i \cdot \sum_{j>i} d_j)) \cdot P = 0$, we can get $(\sum(d_i \cdot \sum_{j<i} d_j) - \sum(d_i \cdot \sum_{j>i} d_j)) \cdot P = 0 = \sum(d_i \cdot T)$. Then, we can get the following equation:

$$\begin{aligned} & d_x \cdot R'_A + \sum (d_i \cdot R'_{ci}) \\ &= d_x \cdot \left(\sum_{i=1}^{n} R_i - \sum R_{rk} \right) + d_i \cdot \sum T - \sum (d_i \cdot R_x) \\ &= d_x \cdot \sum R_i - \sum (d_i \cdot d_x \cdot P) + 0 \\ &= d_x \cdot \sum R_i - d_x \cdot \sum (d_i \cdot P) = 0. \end{aligned}$$

As $d_x \cdot R'_A + \sum(d_i \cdot R'_{ci}) = 0$, the proposed scheme will work correctly after $M_{r1}, M_{r2}, \ldots, M_{rk}$ are removed from the system. Table III illustrates the equality in a more intuitive way, suppose there are four meters: $M_1, \ldots, M_4$, and M_2, M_3 are removed from the system.

F. Dynamic Join

The proposed scheme supports dynamic join, if m smart meters $M_1, M_2, \ldots, M_k$ with public key $R_1, R_2, \ldots, R_k$ and private key $d_1, d_2, \ldots, d_k$want to join the system. The system still works correctly after the joining of these meters. Following steps must be conducted to ensure the system to work correctly.

1) Aggregator computes $\sum_{k=1}^{m} R_k$ and sends $\sum_{k=1}^{m} R_k$ to meter M_i, M_i updates $R'_{ci} = R_{ci} - \sum_{k=1}^{m} R_k$.
2) Aggregator computes $R'_A = R_A + \sum_{k=1}^{m} R_k$, and sends R'_A to the utility supplier, utility supplier updates $R_A = R'_A$.
3) For a new meter M_k, the aggregator calculates $R_{ck} = \sum_{i=1}^{n} R_i - R_x + (\sum_{j=1}^{k-1} R_j - \sum_{j=k+1}^{m} R_j)$.

Now, we are going to prove the system works correctly after the join of these meters. Suppose $\sum_{t=1}^{i-1} R_t - \sum_{t=i+1}^{n} R_t - \sum_{k=1}^{m} R_k = T_1$, we get the following equation:

$$\begin{aligned} R'_{ci} &= R_{ci} - \sum_{k=1}^{m} R_k \\ &= \sum_{t=1}^{i-1} R_t - \sum_{t=i+1}^{n} R_t - R_x - \sum_{k=1}^{m} R_k \\ &= \sum_{t=1}^{i-1} R_t - \sum_{t=i+1}^{n} R_t - \sum_{k=1}^{m} R_k - R_x \\ &= T_1 - R_x. \end{aligned}$$

$\sum_{i=1}^{n}(d_i \cdot R'_{ci})$ can be transformed to be

$$\begin{aligned} \sum_{i=1}^{n} (d_i \cdot R'_{ci}) &= \sum_{i=1}^{n} (d_i \cdot (T_1 - R_x)) \\ &= d_i \cdot \sum_{i=1}^{n} T_1 - \sum_{i=1}^{n} (d_i \cdot d_x \cdot P) \\ &= d_i \cdot \sum_{i=1}^{n} T_1 - d_x \sum_{i=1}^{n} (d_i \cdot P) \\ &= d_i \cdot \sum_{i=1}^{n} T_1 - d_x \cdot \sum_{i=1}^{n} R_i. \end{aligned}$$

Suppose $\sum_{i=1}^{n} R_i + (\sum_{j=1}^{k-1} R_j - \sum_{j=k+1}^{m} R_j) = T_2$, in a similar way, we get $R_{ck} = T_2 - R_x$

$$\begin{aligned} R_{ck} &= \sum_{i=1}^{n} R_i - R_x + \left(\sum_{j=1}^{k-1} R_j - \sum_{j=k+1}^{m} R_j \right) \\ &= \sum_{i=1}^{n} R_i - R_x + \left(\sum_{j=1}^{k-1} R_j - \sum_{j=k+1}^{m} R_j \right) \\ &= T_2 - R_x. \end{aligned}$$

$\sum_{k=1}^{m}(d_k \cdot R_{ck})$ can be transformed to be

$$\begin{aligned} \sum_{k=1}^{m} (d_k \cdot R_{ck}) &= \sum_{k=1}^{n} (d_k \cdot (T_2 - R_x)) \\ &= d_k \cdot \sum_{k=1}^{m} T_2 - d_x \cdot \sum_{k=1}^{m} R_k. \end{aligned}$$

As $d_x \cdot R'_A = d_x \cdot \sum_{i=1}^{n} R_i + d_x \cdot \sum_{k=1}^{m} R_k$, we get the following equation:

$$\begin{aligned} & \sum_{i=1}^{n} (d_i \cdot R'_{ci}) + \sum_{k=1}^{m} (d_k \cdot R_{ck}) \\ &= d_i \cdot \sum_{i=1}^{n} T_1 - d_x \cdot \sum_{i=1}^{n} R_i + d_k \cdot \sum_{k=1}^{m} T_2 - d_k \cdot \sum_{k=1}^{m} R_x \\ &= d_i \cdot \sum_{i=1}^{n} T_1 + d_k \cdot \sum_{k=1}^{m} T_2 - \left(d_x \cdot \sum_{i=1}^{n} R_i + d_k \cdot \sum_{k=1}^{m} R_x \right) \\ &= d_i \cdot \sum_{i=1}^{n} T_1 + d_k \cdot \sum_{k=1}^{m} T_2 - d_x \cdot R'_A. \end{aligned}$$

As $\sum_{i=1}^{n} (d_i \cdot \sum_{j=1}^{i-1} R_j - d_i \cdot \sum_{j=i+1}^{n} R_j) = 0$, we can get

$$d_i \cdot \sum_{i=1}^{n} T_1 = \sum_{i=1}^{n} \left(d_i \cdot \left(\sum_{t=1}^{i-1} R_t - \sum_{t=i+1}^{n} R_t - \sum_{k=1}^{m} R_k \right) \right)$$

$$= \sum_{i=1}^{n}\left(d_i \cdot \left(\sum_{t=1}^{i-1} R_t - \sum_{t=i+1}^{n} R_t\right)\right) - \sum_{i=1}^{n}\left(d_i \cdot \sum_{k=1}^{m} R_k\right)$$

$$= -\sum_{i=1}^{n}\left(d_i \cdot \sum_{k=1}^{m} R_k\right).$$

Similarly, as $\sum_{k=1}^{m}(d_k \cdot (\sum_{j=1}^{k-1} R_k - \sum_{j=i+1}^{m} R_k)) = 0$, we get $d_k \cdot \sum_{k=1}^{m} T_2 = \sum_{k=1}^{m}(d_k \cdot \sum_{i=1}^{n} R_i)$, then we get

$$d_i \cdot \sum_{i=1}^{n} T_1 + d_k \cdot \sum_{k=1}^{m} T_2$$

$$= -\sum_{k=1}^{m}\left(d_k \cdot \sum_{i=1}^{n} R_i\right) + \sum_{i=1}^{n}\left(d_i \cdot \sum_{k=1}^{m} R_k\right) = 0.$$

With these equations, we can prove the following equation:

$$d_x \cdot R'_A + \sum_{i=1}^{n}(d_i \cdot R'_{ci}) + \sum_{k=1}^{m}(d_k \cdot R_{ck})$$

$$= d_x \cdot R'_A + d_i \cdot \sum_{i=1}^{n} T_1 + d_k \cdot \sum_{k=1}^{m} T_2 - d_x \cdot R'_A$$

$$= d_i \cdot \sum_{i=1}^{n} T_1 + d_k \cdot \sum_{k=1}^{m} T_2 = 0.$$

Now we have proved $d_x \cdot R'_A + \sum_{i=1}^{n}(d_i \cdot R'_{ci}) + \sum_{k=1}^{m}(d_k \cdot R_{ck}) = 0$, which means the system still works correctly after the joining of meters $M_1, M_2, \ldots, M_k$.

V. Security Analysis

We will prove an adversary is unable to get a meter's real-time electricity consumption data if and only if the Bilinear Computational Diffie–Hellman (BCDH) problem is a computational hardness. Suppose G_1 is a cyclic additive group of prime order q, P is a generator of G_1. For any $a, b, c \in Z_n^+$, given an instance $\langle aP, bP, cP\rangle$, it is computationally intractable to compute $e(P,P)^{abc}$.

Theorem 1: An adversary is unable to get a meter's real-time consumption data if and only if the BCDH problem is unable to be solved in polynomial time.

Proof: ($\Rightarrow$) Suppose there is an adversary who is able to solve the BCDH problem, given $aP = R_i$, $bP = R_{ci}$, and $cP = H(t_i)$, an adversary is able to get $e(P,P)^{abc} = e(H(t_i), R_{si}) = g_i$, then he can get $c_{ik} \cdot g_i^{-1} = g_i \cdot g^{m_{ik}} \cdot g_i^{-1} = g^{m_{ik}}$, using Pollard's lambda method, the adversary is able to get a meter's real-time consumption data by: $\log_g(c_i \cdot g_i^{-1})$.

($\Leftarrow$) Suppose the utility supplier is able to get a meter's real-time electricity consumption data m_{ik}. Given $c_{ik} = e(H(t_i), R_{si}) \cdot g^{m_{ik}}$, an adversary is able to get $c_i \cdot (g^{m_{ik}})^{-1} = e(H(t_i), R_{si}) \cdot g^{m_{ik}} \cdot (g^{m_{ik}})^{-1} = e(H(t_i), R_{si})$. For the BCDH problem, suppose $aP = R_i$, $bP = R_{ci}$, and $cP = H(t_i)$, given aP, bP, and cP, an adversary is able to compute $e(P,P)^{abc} = e(H(t_i), R_{si})$ in polynomial time. This means the adversary is able to solve the BCDH problem.

Theorem 2: An external adversary is unable to get total electricity consumption data if and only if the BCDH problem is unable to be solved in polynomial time.

TABLE IV
Parameters of the Curve

Name	Value
q	77854349062394727098758421072735961481422250431953908824328205163735124284803
r	14134776518227074636666380005943348126619871175009787368251308127039782913
h	5508

TABLE V
Computation Time of Different Operations

Operation	G_{bp}	GT_{exp}	GT_{mul}	G_{mul}	G_{muls}	G_{add}	H_{G}	H_b
Time(ms)	4.444	0.674	0.004	8.169	2.254	0.030	1.563	0.008

Proof: ($\Rightarrow$) Suppose there is an adversary who is able to solve the BCDH problem, given $aP = R_x$, $bP = \sum_{i=1}^{n} R_i$, and $cP = H(t_i)$, an adversary is able to get $e(P,P)^{abc} = e(H(t_i), d_x \cdot \sum_{i=1}^{n} R_i) = g_x$, then he can get: $g_x \cdot c_{jk} = g^{\sum_{i=1}^{n} m_{ik}}$. Using the Pollard's lambda method, the adversary is able to get the total consumption data: $\log_g(g^{\sum_{i=1}^{n} m_{ik}})$.

($\Leftarrow$) Suppose the utility supplier is able to get the total consumption data $\sum_{i=1}^{n} m_{ik}$. Given $c_{jk} = \prod_{i=1}^{n} c_{i\mathrm{k}}$, an adversary is able to get $c_{jk} \cdot (g^{\sum_{i=1}^{n} m_{ik}})^{-1} = e(H(t_i), R_{sx})$. For the BCDH problem, suppose $aP = R_x$, $bP = \sum_{i=1}^{n} R_i$ and $cP = H(t_i)$, given aP, bP, and cP, an adversary is able to compute $e(P,P)^{abc} = e(H(t_i), R_{sx})$ in polynomial time. This means the adversary is able to solve the BCDH problem.

VI. Comparison

We used the Java Pairing-Based Cryptography Library [37], Type A pairings are constructed on the curve $y^2 = x^3 + x$ over the field F_q for some prime $q = 3 \bmod 4$. Both G_1, G_T are the group of points $E(F_q)$, $\#E(F_q) = (q+1)$, and $\#E(F_{q^2}) = (q+1)^2$, the embedding degree $k = 2$, G_T is a subgroup of F_{q^2}. We chose a 256-bit Z_n^+, the order is 224 bit, as the recommended elliptic curve length is 256 bit for 2016-2030 by NIST and for 2018 - 2028 by ECRYPT [38]. SHA-256 is used in this article. Table IV lists the parameters of the selected curve, and Table V lists the computation time of different operations. Note that we did not take into account the computation time of Pollard's lambda method in G_1 or G_T, however, this does not affect the results.

1) G_{bp} bilinear map pairing.
2) GT_{exp} element exponentiation in G_T.
3) GT_{mul} element multiplication in G_T.
4) G_{mul} element multiplication in G_1.
5) G_{muls} element multiplication in G_1 with a integer less than 60 bit.
6) G_{add} element addition in G_1.
7) H_{G} hash to an element of G_1.
8) H_b hash to a big integer.

The experiment was conducted on a computer with 64-bits windows 7 education edition; the CPU is an Intel(R) Core (TM) M460 2.53 GHz processor, 2 GB memory. The Java source code has been uploaded to a public repository in github.com [39].

TABLE VI
COMPUTATION COST OF THE REGISTRATION PHASE

Schemes	G_{bp}	G_{mul}	G_{add}	H_G	H_b
Boudia [16]		3n+5			
Fan [17]		4n	n		3n
Wang [19]	1	n+3		n+2	
Our scheme		2n+3	3n-1		

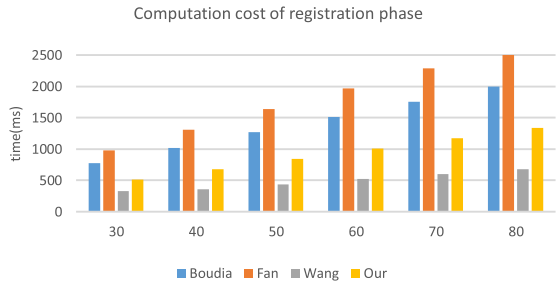


Fig. 2. Computation cost of registration phase.

TABLE VII
METER'S COMPUTATION COST AT AGGREGATION PHASE

Schemes	G_{bp}	GT_{exp}	GT_{mul}	G_{mul}	G_{add}	H_G	H_b
Boudia [16]				4	1		1
Fan [17]				4	2	2	
Wang [19]		2	1	2	1	1	
Our scheme	1	1	1	1		2	

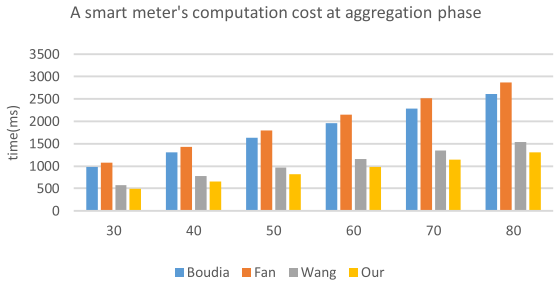


Fig. 3. Smart meter's computation cost at aggregation phase.

TABLE VIII
COMPUTATION COST OF THE VERIFICATION PROCESS

Schemes	G_{bp}	GT_{mul}	G_{mul}	G_{muls}	G_{add}	H_G	H_b
Boudia [16]			n+1		2n-1		n
Fan [17]	n+1	n-1		2n	n-1	n	
Wang [19]	n+2	n		3n	2n-2	2n	
Our scheme	n+1	n-1		2n	n-1	n	

A. Computation Complexity

First, we analyzed the computation cost of the registration phase. Suppose there are n meters and one aggregator, a meter reports one type of data at a time. For the scheme of Boudia [16], it requires $(3n+5)\,G_{\text{mul}}$ operations. For the scheme of Fan [17], it requires $4nG_{\text{mul}}$, nG_{add}, and $3nH_b$operations. For the scheme of Wang [19], it requires $1G_{bp}$, $(n+3)G_{\text{mul}}$ and $(n+2)H_G$ operations. For the proposed scheme, it requires $(2n+3)G_{\text{mul}}$ and $(3n-1)G_{add}$ operations, for the G_{add} operations, the first meter computes $R_{ci} = (\sum_{j=1}^{i-1} R_j - \sum_{j=i+1}^{n} R_j - R_x)$, it requires $(n-1)\,G_{add}$, from the second meter on, it only requires $2G_{add}$ to compute R_{ci}, as $R_{ci} = R_{c(i-1)} + R_i + R_{i-1}$. The results are listed at Table VI.

Fig. 2 shows the computation cost of registration phase, the vertical axis indicates the computation time in milliseconds, the horizontal axis represents n. The computation cost of the scheme of Wang [19] is less than the proposed scheme. In their scheme, meters encrypt their electricity consumption data using the electricity service provider's public key, while in the proposed scheme, every meter will be assigned an independent key to encrypt its electricity consumption data, this requires extra computation cost. For the scheme of Wang [19], if the electricity service provider accidentally gets a meter's encrypted electricity consumption data, it is able to decrypt the encrypted data using its private key. However, in the proposed scheme, even if the utility supplier gets a meter's encrypted electricity consumption data, it is unable to decrypt the encrypted data using its private key.

The computation cost of the schemes of Boudia [16] is higher than that of the proposed scheme. In their scheme, meters and the aggregator have to build shared keys, which needs $2nG_{\text{mul}}$ operation, this requires extra computation cost. Besides, the electricity consumption data are encrypted by the control center's public key, their scheme has the same problem as the scheme of Wang [19].

For the scheme of Fan [17], the signature scheme they used at the registration phase requires more computation cost, to generate the signatures, meters need to conduct nG_{mul} and $2nH_b$ operations, to verify these signatures, the aggregator needs to conduct nG_{add}, $2nG_{\text{mul}}$, and nH_b operations.

Second, we compared a meter's computation cost at aggregation phase, suppose a meter reports one type of data. For the scheme of Boudia [16], it requires $4G_{\text{mul}}$, $1G_{add}$, and $1H_b$ operations. For the scheme of Fan [17], it requires $4G_{\text{mul}}$, $2G_{add}$, and $2H_G$ operations. For the scheme of Wang [19], it requires $2G_{\text{exp}}$, $1GT_{\text{mul}}$, $2G_{\text{mul}}$, $1G_{add}$, and $1H_G$ operations. For the proposed scheme, it requires $1G_{bp}$, $1G_{\text{exp}}$, $1GT_{\text{mul}}$, $1G_{\text{mul}}$, and $2H_G$ operations. The results are listed in Table VII.

Fig. 3 shows a meter's computation cost at aggregation phase, the vertical axis indicates the computation time in milliseconds, the horizontal axis represents n. The computation costs of the other schemes are higher than that of the proposed scheme, because it costs more time to encrypt a meter's electricity consumption data. For the schemes of Boudia [16], a smart meter encrypts its electricity consumption data using the control center's public key, it costs 3 G_{mul} and 1 G_{add} operations. For the schemes of Fan [17], it costs 3 G_{mul}, 2 G_{add}, and 1 H_G operations. For the scheme of Wang [19], it costs 1 G_{mul}, 2 GT_{exp}, and 1 GT_{mul} operations. While for the proposed scheme, it requires 1 H_G, 1 GT_{exp}, 1 GT_{mul}, and 1 G_{bp} operations.

Third, we analyzed the message verification process, the results are shown in Table VIII. As the proposed scheme and the scheme of Fan [17] use the same verification method, their computation costs are the same.

Fig. 4 shows the computation cost of the message verification process, the vertical axis indicates the computation time in milliseconds, the horizontal axis represents n. Boudia [16] used a batch verification method to accelerate the message verification process, however, there is a potential risk in the

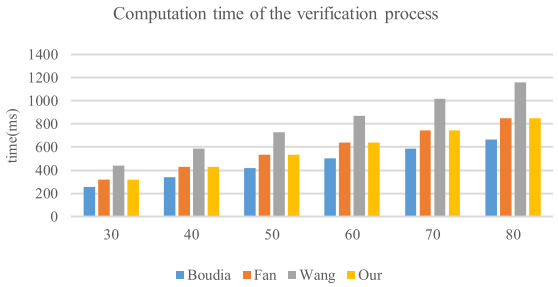


Fig. 4. Computation cost of signature schemes.

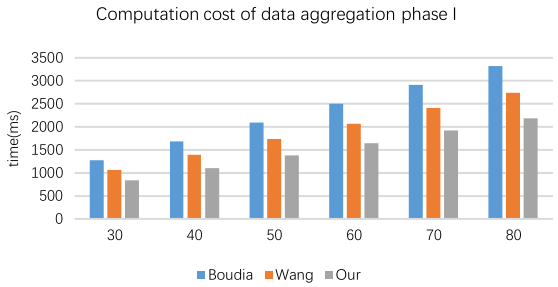


Fig. 5. Computation cost of data aggregation phase I.

batch verification process. The aggregator checks the equation $\sum_{i=1}^{N}(r_{ij2}\cdot G)=(\sum_{i=1}^{N}u_{ij})\cdot G+\sum_{i=1}^{N}(v_{ij}\cdot P_{ij})$ holds or not, if the sum $\sum_{i=1}^{N}u_{ij}$ is correct, the equation holds, it does not check the correctness of a single u_{ij}, if an adversary modifies u_{ij1} and u_{ij2}, let u_{ij1} and u_{ij2} to be $(u_{ij1}+k_1)$ and $(u_{ij2}-k_1)$, k_1 is a random number. The sum $\sum_{i=1}^{N}u_{ij}$ is the same, the equation still holds, the aggregator is unable to find out this situation, which means there is a potential security risk. The computation cost of the verification processes of Wang [19] is higher than that of the proposed scheme.

Fourth, we analyzed the computation cost of the aggregation phase, we compared our scheme with the schemes of Boudia [16] and Wang [19], as these three schemes can be applied to a three-layer system. Fig. 5 shows the results. For the scheme of Boudia [16], a meter encrypts its electricity consumption data using the control center's public key, this needs $3G_{\text{mul}}$ and $1G_{add}$ operations, to decrypt the data, the control center needs to conduct $1G_{\text{mul}}$, $1G_{add}$ operation. This is the main reason their scheme's computation cost is higher than that of the proposed scheme. For the scheme of Wang [19], a meter's computation cost is higher than that of the proposed scheme, besides, the verification process of their scheme needs more computation time, thus their scheme is not as efficient as the proposed scheme.

The scheme of Fan [17] and the proposed scheme can be applied to a two-layer system, we compared these two schemes separately. Fig. 6 shows the computation cost of the aggregation phase. The two schemes use the same signature scheme, however, the encryption scheme used in our scheme is more efficient, thus the proposed scheme is more efficient.

Last, at data aggregation phase, in the schemes of Boudia [16] and in the proposed scheme, a meter is able to report k types of data at once, we analyzed the computation costs of these two schemes separately. For the scheme of Boudia [16], a meter conducts $(2k+2)$ G_{mul}, kG_{add}, and $1H_b$ operations, while it

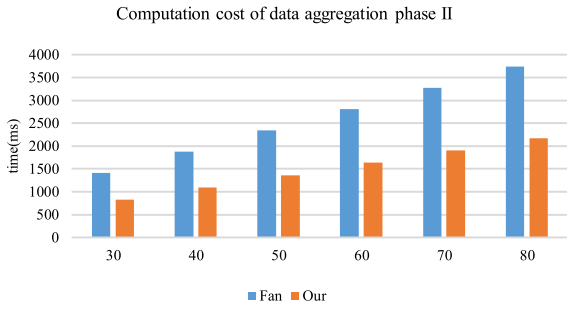


Fig. 6. Computation cost of data aggregation phase II.

TABLE IX
SMART METER'S COMPUTATION COST FOR REPORTING K TYPES OF DATA

Schemes	G_{bp}	GT_{exp}	GT_{mul}	G_{mul}	G_{add}	H_G	H_b
Boudia [16]				2k+2	k		1
Our scheme	1	k	k	1		2	

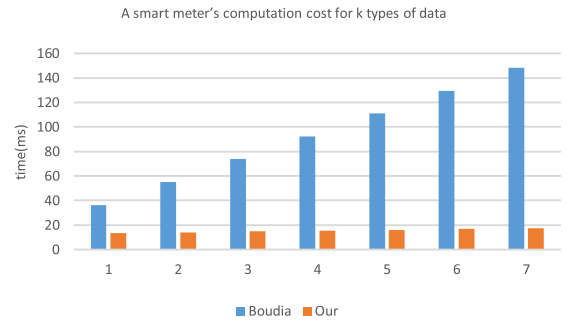


Fig. 7. Smart meter's computation cost for k types of data.

requires $1G_{bp}$, kG_{exp}, kGT_{mul}, $1G_{\text{mul}}$, and $1H_G$ operations for the proposed scheme. The results are shown in Table IX.

Fig. 7 shows a meter's computation cost for reporting k types of data, the vertical axis indicates the computation time in milliseconds, the horizontal axis represents k. The computation time of the scheme of Boudia [16] is higher than that of the proposed scheme, because the encryption method they used costs more computation time. In their scheme, to encrypt k types of data, it cost $(2k+1)$ G_{mul} and k G_{add} operations, while in the proposed scheme, it requires k GT_{exp}, k GT_{mul}, 1 H_G, and 1 G_{bp} operations. To decrypt these data, in their scheme, it requires k G_{mul} and k G_{add} operations, while in the proposed scheme, it requires k GT_{mul}, 1 H_G, and 1 G_{bp}operations. The computation cost of G_{mul} is higher than the other types of operations, thus the proposed scheme is more efficient.

Fig. 8 shows the computation time of aggregation phase for k types of data, when there are 20 smart meters in the system, the vertical axis indicates the computation time in milliseconds, the horizontal axis represents k. The computation time of the scheme of Boudia [16] is higher than that of the proposed scheme, as we have discussed, the encryption method they used costs more computation time.

B. Simulation Results

The simulation environments are the same as that mentioned at the beginning of this section, the scheme of Fan is based on

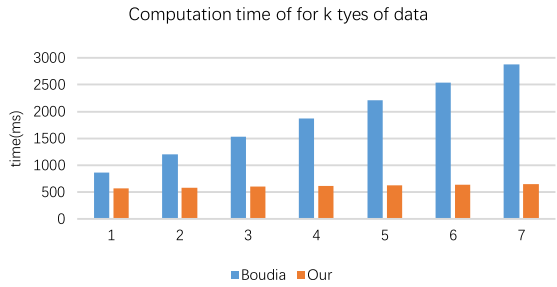


Fig. 8. Computation cost of aggregation phase for k types of data.

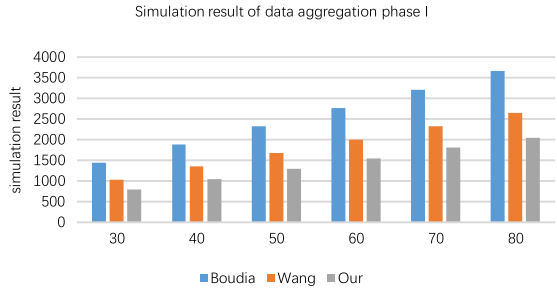


Fig. 9. Simulation result of data aggregation phase I.

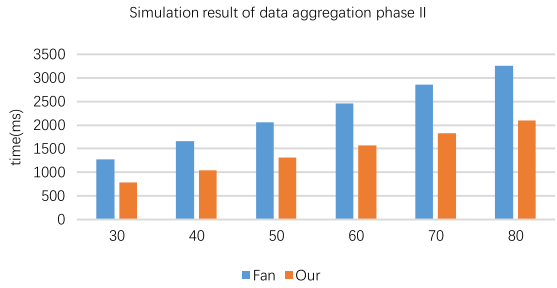


Fig. 10. Simulation result of data aggregation phase II.

type A1 pairings with parameters of equal length as that on type A pairings.

The schemes of Boudia [16], Wang [19], and the proposed scheme are simulated. Fig. 9 shows the simulation result, the advantage of the proposed scheme becomes larger compared to the analysis for Fig. 5. In Fig. 5, the computation time of Pollard's lambda method in G_1 or G_T are not considered, while the computation time of Pollard's lambda method in G_T is less than that in G_1, the advantage of the proposed scheme becomes larger in Fig. 9.

The simulation result of Fan [17] and the proposed scheme for a two-layer system are shown in Fig. 10, the results are similar to that of our analysis for Fig. 6. The proposed scheme is more efficient. When the number of meters grows, the gaps between the two schemes become larger.

Fig. 11 shows the simulation result of aggregation phase for k types of data, the number of meters is set to be 20, a meter's data is in [0, 1000], the simulation result is similar to that of our analysis for Fig. 8, the proposed scheme is more efficient. For the server to get k types of data, it needs to conduct k Pollard's lambda operations, thus the gap between the two schemes becomes larger when the number of types increases.

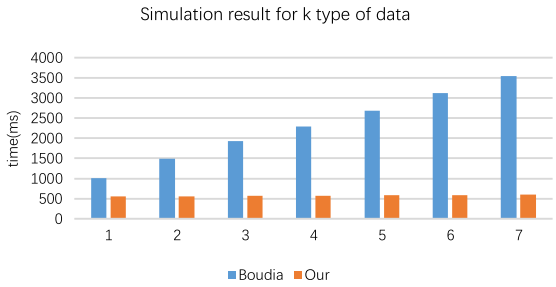


Fig. 11. Simulation results for k types of data.

Besides, when the number of types grows, the computation cost of the scheme of Boudia grows more quickly than that of the proposed scheme, because the computation cost at the meter side is higher than that of the proposed scheme, which is similar to the analysis of the computation cost at meter side in Fig. 7.

C. Communication Overhead

In this section, we analyzed the communication cost. Z_n^+ is 256 bit, an element of G_1 is 512 bit, an element of G_T is 512 bit. The result of SHA-256 is 256 bit. A timestamp is 32 bit, an identity is 32 bit. Suppose there are n meters and an aggregator in the system, a meter reports one type of data at a time.

We first analyzed the communication cost of the registration phase. For the proposed scheme, the messages between an aggregator and a meter are $\{id_i, R_i\}$ and $\{R_{ci}\}$. The messages between an aggregator and a utility supplier are $\{\mathrm{id}_j, R_j, R_A = \sum_{i=1}^{n} R_i\}$ and $\{R_x\}$. id_i is a 32-bit identity, $R_i, R_{ci}, R_j, R_A, R_x$ are 512-bit elements of G_1. The communication cost is $(512 \times 2 + 32)n + 512 \times 3 + 32 = 1056n + 1568$ bit.

For the scheme of Boudia [16], a meter sends $\{\mathrm{ID}_{ij}, P_{ij}\}$ to an aggregator, the aggregator sends $\{P_j\}$ back to this meter and $\{P_j\}$ to the control center. The control center sends $\{Y_i\}$ to a meter and $\{Y_{cc}\}$ to the aggregator. ID_{ij} is a 32-bit identity. P_j, Y_i, Y_{cc}, P_{ij} are 512-bit elements of G_1. The communication cost is $(512 \times 3 + 32)n + 512 \times 2 = 1568n + 1024$ bit.

For the scheme of Fan [17], a meter sends $\{Y_i, \alpha_i, \beta_i, \gamma_i, \mathrm{ID}_i\}$ to an aggregator, Y_i and α_i are 512-bit elements of G_1, β_i and γ_i are 224-bit moduli of the order. ID_i is a 32-bit identity. Besides, at system initialization phase, an offline trusted third party sends π_i to meter M_i and π_0 to the aggregator, π_0 and π_i are 1024-bit numbers. The communication cost is $(512 \times 2 + 224 \times 2 + 32)n + 1024(n + 1) = 2528n + 1024$ bit.

For the scheme of Wang [19], a meter sends $\{ID_i\}$ to the system, the system sends $\{d_{\mathrm{ID}i}, W\}$ back to this meter. ID_i is a 32-bit identity, $d_{\mathrm{ID}i}$ is a 512-bit element of G_1, W is a 512-bit element of G_T. Besides, the communication cost for a collector and an aggregator are both $(512 + 32) = 544$ bit. The communication cost is $(512 \times 2 + 32)n + 544 \times 2 = 1056n + 1088$ bit. The results are shown in Fig. 12.

At data aggregation phase, for the proposed scheme, a meter reports $\{c_i = c_{i1}, S_i, t_i, id_i\}$ to an aggregator. c_{i1} is a 512-bit element of G_T, S_i is a 512-bit element of G_1, t_i is a 32-bit timestamp, id_i is a 32-bit identity. The communication cost of a meter to an aggregator is $(512 \times 2 + 32 \times 2) = 1088$ bit. The communication cost of an aggregator to the utility supplier is 1088 bit, too.

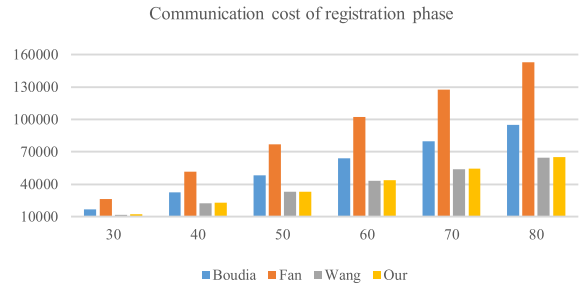


Fig. 12. Communication cost of registration phase.

TABLE X
COMMUNICATION COST OF THE SCHEMES

Schemes	Meter to aggregator	Aggregator to utility supplier
Boudia [16]	1824 bits	1824 bits
Fan (2014) [17]	1056 bits	---
Wang (2017) [19]	1600 bits	1600 bits
Our scheme	1088 bits	1088 bits

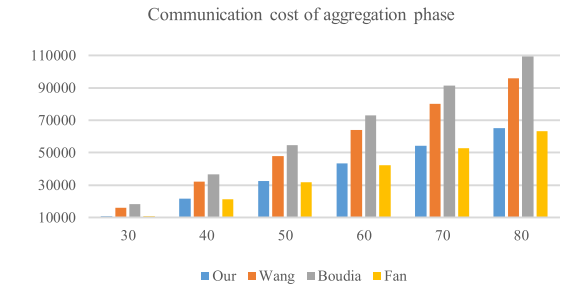


Fig. 13. Communication cost of aggregation phase.

For the scheme of Boudia [16], a meter reports $\{C_{ij}, \mathrm{ID}_{ij}, \mathrm{TS}, S_{ij}\}$ to an aggregator, $C_{ij} = (r_{ij1} \cdot G, r_{ij1} \cdot Y_1 + M_{ij1}), r_{ij1} \cdot G$ and $(r_{ij1} \cdot Y_1 + M_{ij})$ are 512-bit elements of G_1, $S_{ij} = (r_{ij2} \cdot G, z_{ij}), r_{ij2} \cdot G$ is a 512-bit element of G_1, z_{ij} is a 224-bit modulus of the order. TS is a 32-bit timestamp, ID_{ij} is a 32-bit identity. The communication cost from a meter to an aggregator is $(512 \times 3 + 224 + 32 \times 2) = 1824$ bit. The communication cost from an aggregator to the utility supplier is 1824 bit, too.

For the scheme of Fan [17], a meter reports $\{CT_i, \sigma_i, id_i\}$ to an aggregator, CT_i and σ_i are 512-bit elements of G_1, id_i is a 32-bit identity. The communication cost from a meter to an aggregator is $(512 \times 2 + 32) = 1056$ bit.

For the scheme of Wang [19], a meter reports $\{CT_i = (g^{r_i}, g_T^{m_i} * W_i), V_i, T_i, id_i\}$ to an aggregator, g^{r_i} and V_i are 512-bit elements of G_1, $g_T^{m_i} * W_i$ is a 512-bit element of G_T. T_i is a 32-bit timestamp, id_i is a 32-bit identity. The communication cost from a meter to an aggregator is $(512 \times 3 + 32 \times 2) = 1600$ bit. The communication cost from an aggregator to the utility supplier is 1600 bit, too.

The results are shown in Table X and Fig. 13. For scheme of Fan [17], a meter does not send a timestamp to an aggregator, thus their communication cost is 32 bit less than ours, however, this requires the meter and the aggregator to keep their time consistent, even if there is one-millisecond error, the aggregator will not get the exact electricity consumption data. Compared to the scheme of Wang [19], a meter reports one more element of G_1 in their scheme than the proposed scheme. In the scheme of Boudia [16], a signature is one element of G_1 and a modulus of the order, while in the proposed scheme, a signature is only one element of G_1, besides, the encrypted electricity data are two 512-bit elements of G_1 in their scheme, while in the proposed scheme, the encrypted result is one 512-bit element of G_T.

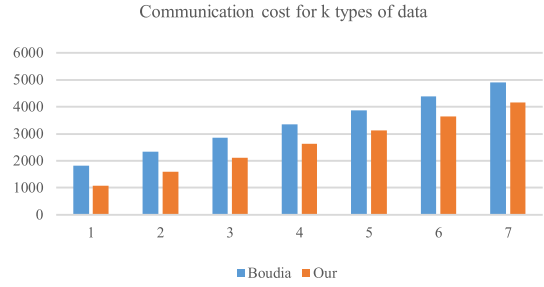


Fig. 14. Communication cost for k types of data.

TABLE XI
SECURITY FEATURES

Comparison	Boudia [16]	Fan [17]	Wang [19]	Our
Dynamic join	√	×	√	√
Dynamic leave	√	×	√	√
Meter replacement	√	×	√	√
Failure recovery	√	×	√	√
Flex structure	×	√	×	√
Collusion attack	×	√	×	√
Eavesdrop attack	×	√	×	√
Private key leakage problem	×	√	√	√
Private key privacy	√	√	×	√
Batch verification problem	×	√	√	√

For the scheme of Boudia [16] and the proposed scheme, a meter is able to report k types of data at a time, we analyzed the these two schemes separately. The result is shown in Fig. 14.

For the scheme of Boudia [16], a meter sends $\{C_{ij}, \mathrm{ID}_{ij}, \mathrm{TS}, S_{ij}\}$ to an aggregator, $C_{ij} = (r_{ij1} \cdot G, r_{ij1} \cdot Y_1 + M_{ij1}, r_{ij1} \cdot Y_2 + M_{ij2}, \ldots, r_{ij1} \cdot Y_l + M_{ijl},), \quad r_{ij1} \cdot G$ and $(r_{ij1} \cdot Y_1 + M_{ij1}, r_{ij1} \cdot Y_2 + M_{ij2}, \ldots, r_{ij1} \cdot Y_l + M_{ijl})$ are 512-bit elements of G_1. $S_{ij} = (r_{ij2} \cdot G, z_{ij}), r_{ij2} \cdot G$ is a 512-bit element of G_1, z_{ij} is a 224-bit modulus of the order. TS is a 32-bit timestamp, ID_{ij} is a 32-bit identity. The communication cost is 512(k + 1) + 512 + 224 + 32 + 32 = 512k + 1312 bit.

For the proposed scheme, a smart meter sends $\{c_i, S_i, t_i, id_i\}$ to an aggregator. $c_i = (c_{i1}, c_{i2}, .., c_{ik})$ are 512-bit elements of G_T. S_i is a 512-bit element of G_1, t_i is 32-bit timestamp, id_i is a 32-bit identity. The communication cost is 512k + 512 + 32 + 32 = 512k + 576 bit.

VII. SECURITY FEATURE COMPARISON

We compared all the schemes under different metrics, the results are shown in Table XI.

A. Dynamic Join and Dynamic Leave

The proposed scheme supports dynamic join and dynamic leave, which are two necessary features for the smart grid scenario, as it is common to add a new meter to the grid or to remove a broken meter from the grid. However, for the scheme of Fan [17], once the system is deployed, it is unable to add or to remove a smart meter from the system.

B. Meter Replacement

If a meter is broken, we need to replace it with a new one, for the proposed scheme, we can replace the broken one with a new one in two steps, first, remove the broken meter from the system, second, add a new meter to the system, the system will still work correctly. For the scheme of Fan [17], their scheme does not support meter replacement, once a meter is broken, the system fails to work correctly.

C. Failure Recovery

For the scheme of Fan [17], if a meter fails to work normally, the utility supplier is unable to get the total consumption data, and their scheme does not provide a failure recovery procedure, while in the proposed schemes, even if one or some of the meters fail to work normally, the system can be recovered, either by conducting dynamic leave procedure or dynamic leave and dynamic join procedures together.

D. Flex Structure

In the scheme of Wang [19], a meter's electricity consumption data is encrypted using the control center's public key, the control center is able to get a meter's consumption data if it gets the encrypted data accidentally, thus there has to be an aggregator to add all meters' encrypted electricity data before they are sent to the utility supplier, which means their scheme can only be applied to a three-layer system. For the proposed scheme, a meter is assigned an independent key to encrypt the encrypted electricity consumption data, the encrypted data can be sent to the utility supplier directly, as the utility supplier is unable to get a single smart meter's electricity consumption data with its key, the proposed scheme can be applied to a two-layer or three-layer system.

E. Collusion Attack and Eavesdrop Attack

In the scheme of Boudia [16], a meter's electricity consumption data are encrypted using the control center's public key, if the aggregator and the control center collude, the aggregator sends a meter's encrypted electricity consumption data to the control center, the control center is able to get the meter's electricity consumption data using its private key. Besides, if the control center accidentally gets a meter's encrypted electricity consumption data by eavesdropping, it is able to get the electricity consumption data. The scheme of Wang [19] has the same problem.

F. Private Key Leakage Problem

In the scheme of Boudia [16], a part of the signature is $z_{ij} = r_{ij2}^{-1}(H(D) + a_{ij}r)$, if the adversary accidentally learns the session empirical information r_{ij2}, it is able to get a meter's private key a_{ij}, which means there is a private key leakage problem.

G. Private Key Privacy

In the scheme of Wang [19], the private keys of the meters, the collector and the electricity service provider are generated by the system, the system knows the private keys of these entities. The information encrypted using their public keys is transparent to the system. While in the proposed scheme, the key pairs are generated by the entities themselves.

H. Batch Verification Problem

As we discussed in Section VI-A, the batch verification process of Boudia [16] is not secure.

After comparison, we can get the conclusion that the proposed scheme has more security features and is more flexible, while it is more efficient in computation cost and communication cost.

VIII. Conclusion

In this article, we introduced an elliptic curve-based scalable smart meter electricity consumption data aggregation scheme, the proposed scheme is scalable, which allows the dynamic join and dynamic leave of smart meters, when some meters fail to work, it is relatively easy to replace them with new ones. In addition, the proposed scheme enables a meter to report multiple types of data at a time, while the utility supplier is unable to learn a single smart meter's real-time electricity consumption data. We also analyzed the security features of the schemes, it turns out that the proposed scheme can resist various attacks. Our simulation results show that the proposed scheme is more effective.

Acknowledgment

The work presented in this article is part of the work made in the I3RES research project, an FP7 initiative (reference number 318184) that targets the seamless integration of Renewable Energy Sources and development of management tools for the Smart Grid. The proposal presented in this article is part of the work made in the European project (e-GOTHAM) (reference number 295378).

References

[1] "Smart Metering deployment in the European Union | JRC Smart Electricity Systems and Interoperability," 2014. [Online]. Available: http://ses.jrc.ec.europa.eu/smart-metering-deployment-european-union. Accessed on: Oct. 15, 2018.

[2] "Smart metering in Europe," 2018. [Online]. Available: http://www.berginsight.com/ReportPDF/ProductSheet/bi-sm13-ps.pdf. Accessed on: Oct. 15, 2018.

[3] K. D. Anderson, M. E. Bergés, A. Ocneanu, D. Benitez, and J. M. F. Moura, "Event detection for non intrusive load monitoring," in *Proc. 38th Annu. Conf. IEEE Ind. Electron. Soc.*, 2012, pp. 3312–3317.

[4] J. Kelly *et al.*, "NILMTK V0.2: A non-intrusive load monitoring toolkit for large scale data sets: Demo abstract," in *Proc. 1st ACM Conf. Embedded Syst. Energy-Efficient Buildings*, New York, NY, USA, 2014, pp. 182–183.

[5] N. Batra *et al.*, "NILMTK: An open source toolkit for non-intrusive load monitoring," in *Proc. 5th Int. Conf. Future Energy Syst.*, New York, NY, USA, 2014, pp. 265–276.

[6] F. Hao, P. Y. A. Ryan, and P. Zieliński, "Anonymous voting by two-round public discussion," *IET Inf. Secur.*, vol. 4, no. 2, pp. 62–67, Jun. 2010.

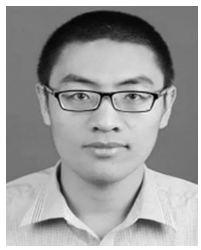
[7] F. Hao and P. Zieliński, "A 2-round anonymous veto protocol," in *Proc. Int. Workshop Secur. Protocols*, 2009, pp. 202–211.

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

[9] P. Paillier, "Public-key cryptosystems based on composite degree residuosity classes," in *Proc. Int. Conf. Theory Appl. Cryptographic Techn.*, 1999, pp. 223–238.

[10] S. Li, K. Xue, Q. Yang, and P. Hong, "PPMA: Privacy-preserving multi-subset data aggregation in smart grid," *IEEE Trans. Ind. Informat.*, vol. 14, no. 2, pp. 462–471, Feb. 2018.

[11] R. Lu, "Privacy-preserving subset data aggregation," in *Privacy-Enhancing Aggregation Techniques for Smart Grid Communications*, R. Lu, Ed. Cham, Switzerland: Springer, pp. 61–84, 2016.
[12] R. Lu, K. Heung, A. H. Lashkari, and A. A. Ghorbani, "A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT," *IEEE Access*, vol. 5, pp. 3302–3312, 2017.
[13] H. Shen, M. Zhang, and J. Shen, "Efficient privacy-preserving cube-data aggregation scheme for smart grids," *IEEE Trans. Inf. Forensics Secur.*, vol. 12, no. 6, pp. 1369–1381, Jun. 2017.
[14] L. Chen, R. Lu, Z. Cao, K. AlHarbi, and X. Lin, "MuDA: Multifunctional data aggregation in privacy-preserving smart grid communications," *Peer-to-Peer Netw. Appl.*, vol. 8, no. 5, pp. 777–792, Sep. 2015.
[15] S. Ge, P. Zeng, R. Lu, and K.-K. R. Choo, "FGDA: Fine-grained data analysis in privacy-preserving smart grid communications," *Peer-to-Peer Netw. Appl.*, vol. 11, no. 5, pp. 966–978, Sep. 2018.
[16] O. R. M. Boudia, S. M. Senouci, and M. Feham, "Elliptic curve-based secure multidimensional aggregation for smart grid communications," *IEEE Sens. J.*, vol. 17, no. 23, pp. 7750–7757, Dec. 2017.
[17] C. I. Fan, S. Y. Huang, and Y. L. Lai, "Privacy-enhanced data aggregation scheme against internal attackers in smart grid," *IEEE Trans. Ind. Informat.*, vol. 10, no. 1, pp. 666–675, Feb. 2014.
[18] D. He, N. Kumar, S. Zeadally, A. Vinel, and L. T. Yang, "Efficient and privacy-preserving data aggregation scheme for smart grid against internal adversaries," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2411–2419, Sep. 2017.
[19] Z. Wang, "An identity-based data aggregation protocol for the smart grid," *IEEE Trans. Ind. Informat.*, vol. 13, no. 5, pp. 2428–2435, Oct. 2017.
[20] 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, Jun. 2014.
[21] K. Kursawe, G. Danezis, and M. Kohlweiss, "Privacy-friendly aggregation for the smart-grid," in *Proc. Int. Symp. Privacy Enhancing Technol.*, 2011, pp. 175–191.
[22] L. Chen, R. Lu, and Z. Cao, "PDAFT: A privacy-preserving data aggregation scheme with fault tolerance for smart grid communications," *Peer-to-Peer Netw. Appl.*, vol. 8, no. 6, pp. 1122–1132, Nov. 2015.
[23] J. Ni, K. Zhang, K. Alharbi, X. Lin, N. Zhang, and X. S. Shen, "Differentially private smart metering with fault tolerance and range-based filtering," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2483–2493, Sep. 2017.
[24] H. Bao and R. Lu, "A new differentially private data aggregation with fault tolerance for smart grid communications," *IEEE Internet Things J.*, vol. 2, no. 3, pp. 248–258, Jun. 2015.
[25] Z. Shi, R. Sun, R. Lu, L. Chen, J. Chen, and X. S. Shen, "Diverse grouping-based aggregation protocol with error detection for smart grid communications," *IEEE Trans. Smart Grid*, vol. 6, no. 6, pp. 2856–2868, Nov. 2015.
[26] 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, Mar. 2013.
[27] G. Si, Z. Guan, J. Li, P. Liu, and H. Yao, "A comprehensive survey of privacy-preserving in smart grid," in *Proc. Int. Conf. Secur., Privacy, Anonymity Comput., Commun., Storage*, 2016, pp. 213–223.
[28] E. Liu and P. Cheng, "Achieving privacy protection using distributed load scheduling: A randomized approach," *IEEE Trans. Smart Grid*, vol. 8, no. 5, pp. 2460–2473, Sep. 2017.
[29] F. D. Garcia and B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," in *Proc. 6th Int. Conf. Secur. Trust Manage.*, 2010, pp. 226–238.
[30] N. Busom, R. Petrlic, F. Sebé, C. Sorge, and M. Valls, "Efficient smart metering based on homomorphic encryption," *Comput. Commun.*, vol. 82, pp. 95–101, May 2016.
[31] S. Ruj and A. Nayak, "A decentralized security framework for data aggregation and access control in smart grids," *IEEE Trans. Smart Grid*, vol. 4, no. 1, pp. 196–205, Mar. 2013.
[32] A. Abdallah and X. Shen, "Lightweight security and privacy preserving scheme for smart grid customer-side networks," *IEEE Trans. Smart Grid*, vol. 8, no. 3, pp. 1064–1074, May 2017.
[33] A. Abdallah and X. S. Shen, "A lightweight lattice-based homomorphic privacy-preserving data aggregation scheme for smart grid," *IEEE Trans. Smart Grid*, vol. 9, no. 1, pp. 396–405, Jan. 2018.
[34] J. M. Bohli, C. Sorge, and O. Ugus, "A privacy model for smart metering," in *Proc. IEEE Int. Conf. Commun. Workshops*, 2010, pp. 1–5.
[35] X. He, X. Zhang, and C. C. J. Kuo, "A distortion-based approach to privacy-preserving metering in smart grids," *IEEE Access*, vol. 1, pp. 67–78, 2013.
[36] P. Barbosa, A. Brito, and H. Almeida, "A technique to provide differential privacy for appliance usage in smart metering," *Inf. Sci.*, vol. 370–371, pp. 355–367, Nov. 2016.
[37] "JPBC library," Dec. 4, 2013. [Online]. Available: http://gas.dia.unisa.it/projects/jpbc/. Accessed on: Nov. 27, 2019.
[38] "Key length comparison," Jun. 10, 2018. [Online]. Available: https://www.keylength.com/en/compare/. Accessed on: Feb. 21, 2019.
[39] "Source code of this study," Mar. 8, 2019. [Online]. Available: https://github.com/SevenBruce/AggVote



Yuwen Chen received the M.S. degree in computer software and theory from Zhengzhou University, Zhengzhou, China, in 2015. He is currently working toward the Ph.D. degree in telematic engineering with the Technical University of Madrid, Madrid, Spain.

His research interests include IoT security and privacy, and smart grid privacy and security.

José-Fernán Martínez-Ortega received the Ph.D. degree in telematic engineering from the Technical University of Madrid, Madrid, Spain, in 2001.

He is an Associate Professor with the Department of Engineering and Telematic Architectures, Technical University of Madrid. He has authored several national and international publications included in the Science Citation Index in his interest areas. He has participated in several International and European Projects. He is responsible for different Spanish and European public-funded research projects and also research contracts with different IT companies. His main research interest areas and expertise include ubiquitous computing and Internet of Things, smart cities and wireless sensor and actuators networks, next-generation telematic network and services, software engineering and architectures, distributed applications and intermediation platforms (middleware), and high-performance and fault-tolerant systems.

Dr. Martínez-Ortega is a Technical Reviser and Chair of technical national and international events on telematics, as well as a member of different international and scientific committees.

Pedro Castillejo received the Ph.D. degree in telematic engineering from the Technical University of Madrid, Madrid, Spain, in 2015.

He is a member of the GRyS (Group of Next-Generation Networks and Services) researching group at UPM, where he is also working as a Researcher in different European projects, like LIFEWEAR, DEMANES, E-GOTHAM, I3RES, and SWARMs. His current research interests include wireless sensor networks, network security algorithms, network protocols, knowledge management, and tiny devices middleware. He has several conference presentations and paper published in indexed journals. He has also participated as an Invited Lecturer in different undergraduate, master, and doctoral courses.

Lourdes López received the degree in mathematical sciences from the Universidad Complutense de Madrid, Madrid, Spain, in 1985, and the Ph.D. degree in computers engineering from the Universidad Politécnica de Madrid, Madrid, in 1998.

Since 1991, she has been a Professor with the Department of Engineering and Telematics Architectures, UPM. From 2000 to 2009, she has been the Director of the Department of Engineering and Telematics Architectures, EUIT Telecomunicación - UPM. In 1992, she initiated her R&D activity in the Group of Information and Network Security, EUIT Telecommunication, UPM, joining the Group of Telematics Services for the Information Society in 2005. In 2010, she launched a new research group that focused on research on new Telematic Networks, GRyS (Group of Next-Generation Networks and Services). Since January 2012, she has been the Secretary of the Research Center for Software Technology and Multimedia Systems for Sustainability.