Storage-Saving Bi-Dimensional Privacy-Preserving Data Aggregation in Smart Grids

Chun-I Fan[1], Yi-Fan Tseng[2], Yi-Hui Lin[1], Fangguo Zhang[3]

[1] Department of Computer Science and Engineering, National Sun Yat-sen University, Taiwan
[2] Department of Computer Science, National Chengchi University, Taiwan
[3] School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China
cifan@mail.cse.nsysu.edu.tw, yftseng@cs.nccu.edu.tw, blue_6132@hotmail.com, isszhfg@mail.sysu.edu.cn

Abstract

Recently, lots of works on power consumption data aggregation have been proposed for the privacy-preservation of users against the operation center in smart grids. This is the *user-based* data aggregation, which accumulates the power consumption data of a group of users for every time unit. On the other hand, the accumulation of a user's data in a group of time units will facilitate the queries on the user's accumulated power usage in these specified time units, which is *time-based* data aggregation. It enables the operation center to perform individual energy consumption statistics and management and offer customized services. If a data aggregation scheme provides both *user-based* and *time-based* data aggregation, it is said to be it *bi-dimensional*. This manuscript presents the first privacy-preserving bi-dimensional data aggregation scheme, where the storage cost only linearly increases with the number of time units and is independent of the number of users.

Keywords: Smart grid, Privacy-Preserving, Data aggregation

1 Introduction

Being regarded as the next-generation energy grid, the developments and researches of smart grids [1, 5, 8-12, 14-18, 21, 25, 27, 29-30] have thrived over the world. Nowadays, government of these countries: the U.S., China, Australia, South Korea, and European Community (EC) invested heavily in smart grids [11]. The researches can be roughly classified into three topics [2]: energy management [13, 20], information management [4, 28] and security [3, 19, 22].

In smart gird environments, the power usage is monitored and managed by an operation center in order to adjust the supply and demand curve of power usage and detect threats and failures in real time. In such a system, each user will report the information of her/his power usage every time unit, such as 15 minutes. The operation center will estimate users' energy consumption of the next time unit with the information, and then distribute energy to users. With the real-time monitoring, smart grid efficiently reduces the energy consumption compared with traditional architectures. Nevertheless, the frequent monitoring may expose the routines and schedules of users, which causes privacy leakage in smart grids. The data aggregation mechanisms are thus introduced to this environment. The power usage information will be first transmitted to an aggregator. After receiving the data, the aggregator will aggregate every user's data, and send the aggregated data to the operation center. The data received by the operation center have been "accumulated", and thus, they reveal nothing about the private information of each user. This is the *user-based* data aggregation, which provides single-dimensional data aggregation only.

Nevertheless, except the *user-based* data aggregation, we also require *time-based* data aggregation which allows the operation center to retrieve the accumulated energy consumption of a user for some specified time units. It can support the operation center to do individual energy consumption statistics and management for customized services. For instance, there are several power plants where one gives a much lower price of energy on Mondays but higher on Sundays. The operation center needs to know the energy consumption of each user on Mondays and Sundays in order to provide appropriate or customized discounts to the users.

To achieve both *user-based* and *time-based* data aggregation, called bi-dimensional data aggregation, a typical approach is to record the power usage of each user in each time unit in the aggregator for the response to any possible query from the operation center. Assume that N is the number of the residential users and M is the total number of time units. Thus, the storage cost of the aggregator will be $O(N \times M)$, which might be enormous (Figure 1).

*Corresponding Author: Yi-Fan Tseng; E-mail: yftseng@cs.nccu.edu.tw
DOI: 10.3966/160792642019052003027

	Residential users				
	$User_1$	$User_2$	$User_3$	⋯	$User_N$
T_1	d_{11}	d_{21}	d_{31}		d_{N1}
T_2	d_{12}	d_{22}	d_{32}	⋯	
⋮					
T_M	d_{1M}				d_{NM}

The storage cost of the aggregator $(N \times M)$, where N is the number of the residential users and M is the total number of time units.

Figure 1. A typical approach

This manuscript presents the first bi-dimensional data aggregation scheme for smart grids, which requires $O(M)$ storage cost only. Compared to other schemes, the proposed scheme provides lower storage cost and bi-dimensional data queries while achieving privacy preservation simultaneously.

The remainder of this manuscript is organized as follows. In Section 2, we define the system model, and present some techniques for the proposed scheme and security requirements in smart grids. In Section 3, we describe the proposed scheme in detail. In Section 4, the security is analyzed. The features and performance are discussed in Section 5. A concluding remark of this research is given in Section 6.

2 Preliminaries

2.1 System Model

In the proposed scheme, we mainly focus on how to send residential users' data to the aggregator privately, without being eavesdropped or intercepted by the operation center. There are three entities in the system model as follows (Figure 2).

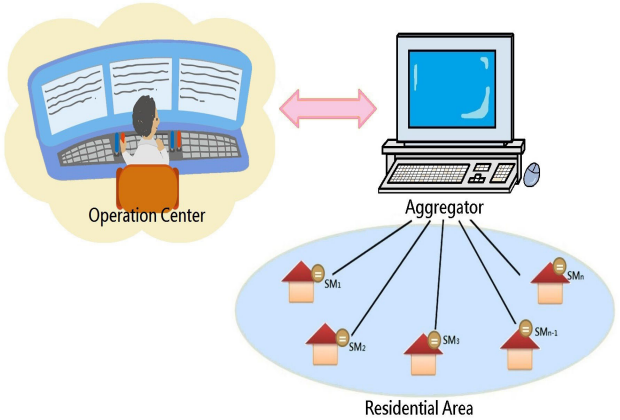

Figure 2. System model

Operation center. The operation center controls the transmission and distribution of electrical energy based on the aggregated data received from the aggregator. In order to achieve privacy preservation, it should be assumed that the operation center does not collude with the aggregator.

Aggregator. The aggregator is mainly responsible for the aggregation of the users' data.

Residential users. Residential users utilize smart meters to generate electricity usages and report their data to the aggregator.

The proposed scheme includes the following four phases.

System initialization. In this phase, all entities, including the operation center, the aggregator, and residential users, setup their public and private parameters.

Data generation. Residential users are equipped with smart meters to record electricity consumption data and compute encrypted data with their signatures, residential area tags, and time stamps. Then, they send the encrypted data to the aggregator.

User-based fata sggregation. After receiving encrypted consumption data from users, the aggregator accumulates these data, and sends addressed data to the operation center.

Time-based data aggregation. When the operation center would like to make a query with a set of indexes of time units $\{\bar{1},\cdots,\overline{M}\}$, which may not be consecutive, it sends the set to the aggregator. After receiving it, the aggregator aggregates the data and responds to the query.

2.2 Paillier Cryptosystem

Paillier cryptosystem [24] was proposed by Pascal Paillier in 1999, which provides additive homomorphism. It consists of the following algorithms.

Key generation. Given a security parameter k, two large prime numbers p, q, where $|p| = |q|$, calculate the RSA modules $n = pq$ and compute $\lambda = lcm(p-1, q-1)$. Define a function $L(u) = \frac{u-1}{n}$, choose a generator g, where the order of g is a nonzero multiple of the modulo n, and further calculate $\mu = (L(g^{\lambda} \bmod n^2))^{-1} \bmod n$. Then, the public key is $pk = (n, g)$, and the corresponding private key is $sk = (\lambda, \mu)$.

Encryption. Given a message $m \in Z_n$, choose a random number $r \in Z_n^*$. The ciphertext can be calculated as $C = E(m) = g^m \cdot r^n \bmod n^2$.

Decryption. Given the ciphertext $C \in Z_{n^2}^*$, the corresponding message can be recovered as $m = D(C) = L(c^{\lambda} \bmod n^2) \cdot \mu \bmod n$.

2.3 Super-Increasing Sequence

A sequence of positive real numbers $\{a_N\}$ is called k-superincreasing if every element of the sequence is greater than k times the sum of all previous elements in

the sequence [23, 26], i.e., $k(\sum_{i=1}^{j-1} a_i) < a_j$, $0 < j \le N$.

Minimally *k*-superincreasing integer sequences. A sequence of positive integers $\{a_N\}$ could thus be said to be minimally *k*-superincreasing if every element of the sequence is equal to *k* times the sum of all previous elements in the sequence, plus one, i.e.,

$$k(\sum_{i=1}^{j-1} a_i) + 1 = a_j, \ 0 < j \le N$$

The minimally *k*-superincreasing sequence of positive integers starting at $h \ge 1$, can be given by the formula

$$a_j = \begin{cases} h & if\ j = 1 \\ (kh+1)(k+1)^{j-1} & if\ j > 1 \end{cases}$$

2.4 Security Requirements

Security is a critical issue in smart grids. We consider that the operation center and the aggregator both are honest but curious. However, there exists an adversary *A* residing in a residential area to eavesdrop the users' reports. In addition, *A* may also intrude into the database of the operation center or the aggregator to steal the individual user reports. The adversary *A* could also take some active attacks to alter the data. Therefore, in order to prevent *A* from learning the users' information, we should meet the security requirements as follows in smart grids.

Privacy preservation. Adversary *A*, who intercepts the communications, cannot derive the contents of the data and significant information from the ciphertext and the public key in polynomial time. Furthermore, none of the participated parties, especially the operation center, can catch the detailed consumption data of any user in the region.

Authentication and sata integrity. All of the reported data should be authenticated, which can ensure that an encrypted report is really sent by a legal residential user and has not been forged or modified during the transmission.

3 The Proposed Scheme

The proposed privacy-preserving bi-dimensional data aggregation scheme for smart grids is presented in this section, where some notations are defined in Table 1. It contains the following phases.

3.1 System Initialization

Operation center. First, the operation center generates a public key (*n*, *g*) and the corresponding private key (λ, μ) of Paillier cryptosystem. Assume that the maximum number of households in a residential area is not greater than a constant *N* and every user's electricity consumption in a time unit is not greater than *d*. The operation center chooses a *k*-superincreasing sequence $\vec{a} = (a_1, a_2, \cdots, a_{N+2})$ such that $k(\sum_{i=1}^{j-1} a_i) < a_j$, $0 < j \le N+2$ where $k = \overline{qd}$. It then computes $(g_1, g_2, \cdots, g_{N+2})$, where $g_i = g^{a_i} \bmod n^2$ for $i = 1, 2, \cdots, N+2$. After that, it chooses and publishes a digital signature scheme $S = (KeyGen, Sign, Ver)$, the parameters as $pubs = \{(n, g)(g_1, g_2, \cdots, g_{N+2})\}$, and keeps ($\lambda$, μ, $\vec{a}$) secretly.

Table 1. The notations

Notation	Meaning
U_i	Residential user *i*
(PK_i, SK_i)	Public/Secret key pair of residential user *i*
$d_{i,j}$	U_i's power consumption data of the *j*-th time unit
M	The total number of time units
N	The total number of residential users
RA	Residential area tag
TS	Timestamp
q	The maximum number of time units in a query
d	The maximum power consumption in a time unit of a user

Aggregator. The aggregator chooses an asymmetric encryption scheme ε= (*KeyGen, Enc, Dec*).

– **User** U_i**:** Compute (PK_i, SK_i) ← *S.KeyGen*.

3.2 Data Generation

A user, say U_i, performs the following operations every time unit, e.g., 15 minutes.

(1) Choose a random number $r_{i,j} \in Z_n^*$.

(2) Let $d_{i,j}$ be U_i's power consumption of the *j*-th time unit. Compute

$$C_{i,j} = (g_i g_{N+2})^{d_{i,j}} r_{i,j}^n \bmod n^2.$$

(3) Compute $\sigma_{i,j}$ = *S.Sign*($C_{i,j}||RA||U_i||TS$) using SK_i, where *RA* represents the residential area and *TS* is the timestamp.

(4) Compute $CT_{i,j}$ = ε.*Enc*($C_{i,j}||RA||U_i||TS||\sigma_{i,j}$) and send it to the aggregator.

3.3 User-Based Data Aggregation (Data Aggregation for the *j*-th Time Unit)

After receiving the encrypted data from all users, the aggregator performs as follows.

(1) For *i* = *1* to *N*, compute $(C_{i,j}||RA||U_i||TS||\sigma_{i,j}) \leftarrow \varepsilon.Dec(CT_{i,j})$.

(2) For *i* = 1 to *N*, verify the signature $\sigma_{i,j}$ using *S.Ver*.

3. Compute $C_j = \prod_{i=1}^{N} C_{i,j} \bmod n^2 =$

$$g^{(\sum_{i=1}^{N} a_i d_{i,j}) + a_{N+2} \sum_{i=1}^{N} d_{i,j}} (\prod_{i=1}^{N} r_{i,j})^n \ mod\ n^2$$

and store C_j.

(4) Compute $C_j^{'} = C_j \prod_{i=1}^{N} g_i^{x_{i,j}} \ mod\ n^2$, where $x_{i,j}$, called a blinding factor, is randomly chosen from [1, k] for i = 1 to N.

5. Report addressed data $C_j^{'}$ to the operation center.

Remark. The purpose of the blinding factors $x_{i,j}$ is to prevent the operator from recovering the power usage of each user in *User-Based* Data Aggregation phase. Without the blinding factors, the operator can easily recover each user's power usage from the plaintext of C_j using $(a_1, a_2, \cdots, a_{N+2})$.

After receiving $C_j^{'}$, the operation center retrieves the aggregated power consumption data $m_{*,j}$ of the j-th time unit as follows.

(1) Use (λ, μ) to decrypt $C_j^{'}$ and obtain the plaintext

$$m_{*,j}^{'} = (\sum_{i=1}^{N} a_i (d_{i,j} + x_{i,j})) + a_{N+2} \sum_{i=1}^{N} d_{i,j} \ mod\ n .$$

(2) Compute

$$m_{*,j} = \frac{m_{*,j}^{'} - (m_{*,j}^{'} \ mod\ a_{N+2})}{a_{N+2}}$$

Which equals to $\sum_{i=1}^{N} d_{i,j}$.

Note that $\sum_{i=1}^{N} d_{i,j}$ has been bound with a_{N+2}, not a_{N+1}, so that $\sum_{i=1}^{N} a_i (d_{i,j} + x_{i,j})$ does not perturb $a_{N+2} \sum_{i=1}^{N} d_{i,j}$ even though it overflows into the message space bound with a_{N+1}.

Remark. The aggregator only requires to keep C_j after performing the above protocol, (i.e. *Data Aggregation for the j-th Time Unit*). For M time units, it just needs $O(M)$ storage to keep $\{C_j\}$, $1 \le j \le M$, which are enough to provide sufficient data for *Time-Based Data Aggregation*.

3.4 Time-based Data Aggregation

When the operation center sends a query with a set of indexes of time units $\{\bar{1}, \cdots, \overline{M}\}$ to the aggregator, the aggregator performs as follows.

(1) Retrieve $C_{\bar{1}}, C_{\bar{2}}, \cdots, C_{\overline{M}}$ from its storage.

(2) Compute $C_{query} = (C_{\bar{1}} C_{\bar{2}}, \cdots C_{\overline{M}}) \bmod n^2$

$$= g^{(\sum_{i=1}^{N} a_i m_{i,*}) + a_{N+2} \sum_{i=1}^{N} m_{i,*}} r^n \ mod\ n^2$$

Where $m_{i,*} = d_{i,\bar{1}} + d_{i,\bar{2}} + \cdots + d_{i,\overline{M}}$ for i = 1 to N and $r \in Z_n^*$

(3) Return C_{query} to the operation center.

After receiving C_{query}, the operation center executes the following steps.

(1) Decrypt $g^{-a_{N+2}(m_{*,\bar{1}} + m_{*,\bar{2}} + \cdots + m_{*,\overline{M}})} C_{query} \bmod n^2$ which equals to $g^{(\sum_{i=1}^{N} a_i m_{i,*})} r^n \ mod\ n^2$ and then obtain the plaintext $t_N = \sum_{i=1}^{N} a_i m_{i,*} \ mod\ n$. Note that $m_{*,\bar{i}}$ is the power consumption of the $\bar{i}$-th time unit, which can be obtained by the algorithm shown in Sec 3.3.

(2) For $i = N$ down to 2, compute and output $m_{i,*} = \frac{t_i - (t_i \ mod\ n)}{a_i}$, and then compute $t_{i-1} = t_i - a_i m_{i,*}$.

3. Compute and output $m_{1,*} = t_1 / a_1$.

Remark. Note that if we make an additional assumption that $(\sum_{i=1}^{N} a_1 m_{1,*}) + a_{N+2} \sum_{i=1}^{N} m_{i,*} < n$, then we can first compute the plaintext $\bar{t}$ of C_{query}, and set $t_N = \bar{t} \ mod\ a_{N+2} = \sum_{i=1}^{N} a_i m_{i,*}$. However, if the plaintext of C_{query} is greater than n, then it will fall out of the message space of Paillier cryptosystem, and then the plaintext will "overflow". Therefore the decryption result will not be correct and we cannot obtain the correct t_N. This is why we take advantage of the additive homomorphism of Paillier encryption to "subtract the unnecessary part from the plaintext" before decryption.

The comparison between the proposed scheme and the other existing schemes is summarized in Table 2.

Table 2. Feature comparison

	[1]	[5]	[6]*	[15]	[25]	Ours
Privacy-preserving agaist external attackers	Yes	Yes	Yes	Yes	Yes	Yes
Privacy-preserving against internal attackers	No	Yes	Yes	No	Yes	Yes
Assumption on aggregator	Semi-trusted	No Assumption	Semi-trusted	trusted	trusted	Semi-trusted
Authentication and data integrity	Yes	No	Yes	No	Yes	Yes
Bi-dimensional data aggregation	No	No	No	No	No	No

*Corrections shown in [7] are considered.

Correctness. Let the power consumption of each user in a time unit be not greater than a constant d, and the

operation center can only make a query with at most $\bar{q}$ time units, i.e., $\overline{M} \le \bar{q}$. For the correctness of the decryption, some restrictions are required.

(1) In the *Time-Based Data Aggregation* phase, it is necessary that

$$a_1 m_1 + a_2 m_2 + \cdots + a_{j-1} m_{j-1} < a_j .$$

Due to this restriction, we can derive an appropriate value of k for the superincreasing sequence $\vec{a} = (a_1, a_2, \cdots, a_{N+2})$ as follows.

$$\begin{aligned} & a_1 m_1 + a_2 m_2 + \cdots + a_{j-1} m_{j-1} \\ & \le a_1(\bar{q}d) + \cdots + a_{j-1}(\bar{q}d) \\ & = (\bar{q}d)((a_1 + \cdots + a_{j-1}) \\ & < a_j \end{aligned}$$

Thus we have $k = \bar{q}d$. In addition, it is also necessary that

$$m = a_1 m_1 + a_2 m_2 + \cdots + a_N m_N < n.$$

Due to this restriction, we can derive the relationship between $\bar{q}$ and N as follows. Assume $|n| = 1024$.

$$\begin{aligned} & a_1 m_1 + a_2 m_2 + \cdots + a_N m_N \\ & \le (\bar{q}d)(a_1 + \cdots + a_N) \\ & = (\bar{q}d)(a_1 + \cdots + a_{N-1}) + (\bar{q}d)a_N \\ & < a_N + (\bar{q}d)a_N \\ & = (\bar{q}d + 1)a_N \\ & = (\bar{q}d + 1)((\bar{q}dh + 1)(\bar{q}d + 1)^{N-1} \end{aligned}$$

Where $h = a_1$ is the staring value of the superincreasing sequence. Assume $a_1 = 1$, and thus, we have

$$\begin{aligned} & (\bar{q}d + 1)((\bar{q}dh + 1)(\bar{q}d + 1)^{N-1} \\ & = (\bar{q}d + 1)^{N+1} \\ & < n \\ & \approx 2^{1024} \end{aligned}$$

For the power consumption of a user within a time unit, e.g. 15 minutes, a reasonable evaluation of d would be 1. That is, the power consumption of a user within a time unit is usually not more than 1 degree. Then we have

$$(\bar{q} + 1)^{N+1} < 2^{1024} \Rightarrow N < 1024(\frac{ln2}{ln(\bar{q} + 1)}) - 1.$$

(2) In the *User-Based Data Aggregation* phase, it is required that

$$(\sum_{i=1}^{N} a_i(d_{i,j} + x_{i,j})) + a_{N+2} \sum_{i=1}^{N} d_{i,j} < n.$$

If $\sum_{i=1}^{N} d_{i,j} \le k$ is assumed, we have that:

$$\begin{aligned} & \sum_{i=1}^{N} a_i(d_{i,j} + x_{i,j})) + a_{N+2} \sum_{i=1}^{N} d_{i,j} \\ & \le (\sum_{i=1}^{N} a_i(d + k)) + k a_n + 2 \\ & = d \sum_{i=1}^{N} a_i + k \sum_{i=1}^{N} a_i + k a_n + 2 \\ & \le \frac{d}{k} a_{N+1} + a_{N+3} - k a_{N+1} \\ & = a_{N+3} + \frac{d - k^2}{k} a_{N+1} \end{aligned}$$

Let $a_1 = 1$, $|n| = 1024$, and $d = 1$. Thus,

$$\begin{aligned} & a_{N+3} + \frac{d - k^2}{k} a_{N+1} \\ & = (k+1)^{N+3} + \frac{d - k^2}{k}(k+1)^{N+1} \\ & = (\bar{q} + 1)^{N+2} \left[(\bar{q} + 1) - \frac{\bar{q} - 1}{\bar{q}} \right] \\ & = (\overline{q} + 1)^{N+2} (\bar{q} + \frac{1}{\overline{q}}) \\ & < (\overline{q} + 1)^{N+3} \\ & < 2^{1024}. \end{aligned}$$

It implies that

$$N < 1024 \frac{ln2}{ln(\overline{q} + 1)} - 3.$$

Combining the aforementioned restrictions in (1) and (2), the relationship between N and $\bar{q}$ is illustrated in Figure 3. Note that, by the assumptions "$\sum_{i=1}^{N} d_{i,j} \le k = \bar{q}d$" and "$d_{i,j} \le d = 1$", we have $N \le \bar{q}$ .Therefore the left side of the curve is omitted, where $N > \bar{q}$.

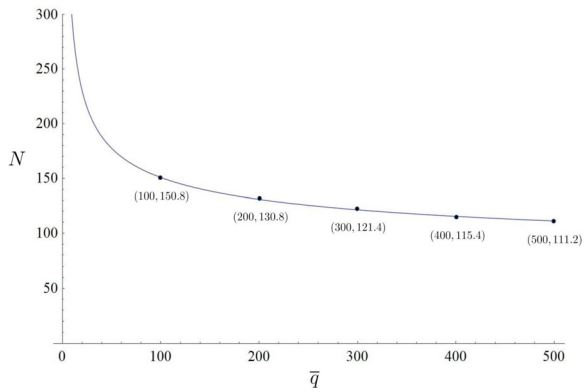

Figure 3. The relation between N and $\bar{q}$.

4 Security

The security of the proposed data aggregation scheme is discussed in this section. There are two types of attackers in smart grid:

External attackers. The external attackers may eavesdrop or intercept the data transmitted among users, the aggregator, and the operation center.

Internal attackers. The internal attackers are allowed to access the data in the storage of the aggregator and the operation center. In the proposed scheme, an internal attacker can be viewed as an insider of the operation center or the aggregator. In general, an internal attacker is much powerful than an external one.

4.1 Privacy Preserving

Paillier cryptosystem is a homomorphic encryption with IND-CPA security. With the existence of the randomization factor *r*, the same data will be transformed to different ciphertexts with different *r*, which makes the approach robust against the dictionary attack. The external attackers cannot obtain any useful information from the communication due to the usage of Paillier encryption and the asymmetric encryptionε. We then discuss the attacks from the internal attackers. Though the operation center owns the private key of Paillier encryption, it cannot know the detailed information of each user's power usage since the data sent to the aggregator are encrypted using the asymmetric encryption schemeε, and the data sent to the operation center are either aggregated (i.e. $\sum_{i=1}^{N} d_{i,j}$ and $d_{i,\bar{1}} + d_{i,\bar{2}} + \cdots + d_{i,\overline{M}}$ or blinded (i.e. $d_{i,j} + x_{i,j}$) by the aggregator. The proposed scheme can also prevent the aggregator from obtaining the users' power consumption data by encrypting the data using the public key of the operation center. It preserves the privacy against both external and internal attackers.

4.2 Authentication and Data Integrity

In the proposed scheme, each user generates a public/secret key pair using *S.Keytten*. All of the data sent by users have been signed by themselves. Owing to the unforgeability of the signature scheme *S*, it prevents users' power consumption data from being forged or modified by the attackers. The data reported by users can be authenticated in the smart grid. In the scheme, the aggregator will authenticate a user by verifying his signature and *TS*, where the unforgeability of the underlying signature scheme *S* can ensure user authentication, data integrity, and unforgeability against external and internal attackers.

5 Comparison

In this section we compare the proposed scheme with [1, 5, 6, 15, 25], where the comparison is summarized in Table 2. The proposed scheme and the scheme [6] with the corrections [7] are privacy-preserving against both external and internal attackers. The others can withstand external attackers. However, the scheme of [1] is not able to resist internal attackers. The author of [25] claimed that his scheme can withstand internal attackers, but he used only an administrative approach rather than cryptographic techniques to prevent internal attacks. Besides, the scheme [15] did not achieve data security. In [5], though the authors claimed that their scheme is able to measure both the spatial and temporal consumptions, their approach is straightforward. In their scheme, all the consumptions need to be stored individually, which is the typical solution illustrated in Figure 1. Therefore, the storage cost is also $O(N \times M)$, just as that of the other schemes. The advantage of [5] may be considered to be free from an on-line aggregator or a trusted third party. The need of trusted third parties is an additional assumption. Under such assumption, the third party may know the secret of users, which would cause some security issue, such as the key escrow problem. As a trade-off, however, the smart meters must share some secrets before encryption.

Consider the storage cost required for the aggregator. It needs $O(N \times M)$ if we apply the typical approach. Nevertheless, the proposed approach gains much lower storage cost, which is $O(M)$ only.

6 Conclusion

A novel privacy-preserving data aggregation scheme for smart grids has been presented in the manuscript. The security of Paillier encryption and the unforgeability of the underlying signature can guarantee the security of the proposed scheme.

Super-increasing sequences have first been applied to achieve bi-dimensional aggregation while gaining low storage cost. Although adopting super-increasing sequences may cause data expansion, we have exhaustedly utilized the unused message space in Paillier encryption. Compared with the typical approach, the storage cost has decreased tremendously, turning $O(N \times M)$ into $O(M)$. In the future, we will further improve the performance of the scheme. Furthermore, we will attempt to solicit a solution to release the limitation on N or M, which is caused by the data expansion owing to the involving of super-increasing sequences.

Acknowledgements

This work was partially supported by Taiwan Information Security Center at National Sun Yat-sen University (TWISC@NSYSU) and the Ministry of Science and Technology of Taiwan under grant MOST 107-2218-E-110-014. It also was financially supported by the Information Security Research Center at National Sun Yat-sen University in Taiwan and the Intelligent Electronic Commerce Research Center from The Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education (MOE) in Taiwan.

References

[1] A. R. Abdallah, X. S. Shen, Lightweight Lattice-Based Homomorphic Privacy-Preserving Aggregation Scheme for Home Area Networks, *2014 Sixth International Conference on Wireless Communications and Signal Processing (WCSP)*, Hefei, China, 2014, pp. 1-6.

[2] S. Bera, S. Misra, J. J. P. C. Rodrigues, Cloud Computing Applications for Smart Grid: A Survey, *IEEE Transactions on Parallel and Distributed Systems*, Vol. 26, No. 5, pp. 1477-1494, May 2015.

[3] C. Boyd, W. Mao, K. G Paterson, Key Agreement Using Statically Keyed Authenticators, *Applied Cryptography and Network Security. ACNS 2004*. Lecture Notes in Computer Science, Vol. 3089. Springer, Berlin, Heidelberg.

[4] S. Bu, F. R. Yu, P. X. Liu, Dynamic Pricing for Demand-Side Management in the Smart Grid, *2011 IEEE Online Conference on Green Communications*, New York, NY, 2011, pp. 47-51.

[5] Z. Erkin, G. Tsudik, Private Computation of Spatial and Temporal Power Consumption with Smart Meters, *International Conference on Applied Cryptography and Network Security*. Singapore, 2012, pp. 561-577.

[6] C. I. Fan, S. Y. Huang, Y. L. Lai, Privacy-Enhanced Data Aggregation Scheme against Internal Attackers in Smart Grid, *IEEE Transactions on Industrial Informatics*, Vol. 10, No. 1, pp. 666-675, February, 2014.

[7] C. I. Fan, S. Y. Huang, Y. F. Tseng, Corrections to Privacy-Enhanced Data Aggregation Scheme Against Internal Attackers in Smart Grid, *Technical Report*, doi: 10.13140/RG.2.1.4006.8649, February, 2015.

[8] M. M. Fouda, Z. M. Fadlullah, N. Kato, R. Lu, X. Shen, A Lightweight Message Authentication Scheme for Smart Grid Communications, *IEEE Transactions on Smart Grid*, Vol. 2, No. 4, pp. 675-685, December, 2011.

[9] S. Fu, J. Ma, H. Li, Q. Jiang, A Robust and Privacy-Preserving Aggregation Scheme for Secure Smart Grid Communications in Digital Communities, *Security and Communication Networks*, Vol. 9, No. 15, pp. 2779-2788, October, 2016.

[10] S. Galli, A. Scaglione, Z. Wang, For the Grid and Through the Grid: The Role of Power Line Communications in the Smart Grid, *Proceedings of the IEEE*, Vol. 99, No. 6, pp. 998-1027, June, 2011.

[11] V. C. Gungor, D. Sahin, T. Kocak, S. Ergüt, C. Buccella, C. Cecati, G. P. Hancke, Smart Grid Technologies: Communication Technologies and Standards, *IEEE Transactions on Industrial Informatics*, Vol. 7, No. 4, pp. 529-539, November. 2011.

[12] M. Hashmi, S. Hanninen, K. Maki, Survey of Smart Grid Concepts, Architectures, and Technological Demonstrations Worldwide, *2011 IEEE Pes Conference on innovative Smart Grid technologies Latin America (ISGT LA)*, Medellin, 2011, pp. 1-7.

[13] G. Koutitas, L. Tassiulas, A Delay Based Optimization Scheme for Peak Load Reduction in the Smart Grid, *2012 Third International Conference on Future Systems: Where Energy, Computing and Communication Meet (e-Energy)*, Madrid, 2012, pp. 1-4.

[14] C. Li, R. Lu, H. Li, L. Chen, J. Chen, PDA: A Privacy-Preserving Dualfunctional Aggregation Scheme for Smart Grid Communications, *Security and Communication Networks*, Vol. 8, No. 15, pp. 2494-2506, 2015.

[15] F. Li, B. Luo, P. Liu, Secure Information Aggregation for Smart Grids Using Homomorphic Encryption, *2010 First IEEE International Conference on Smart Grid Communications*, Gaithersburg, MD, 2010, pp. 327-332.

[16] Q. Li, G. Cao, Multicast Authentication in the Smart Grid with One-Time Signature, *IEEE Transactions on Smart Grid*, Vol. 2, No. 4, pp. 686-696, December, 2011.

[17] Y. Liu, C. Cheng, T. Gu, T. Jiang, X. Li, A Lightweight Authenticated Communication Scheme for Smart Grid, *IEEE Sensors Journal*, Vol. 16, No. 3, pp. 836-842, February, 2016.

[18] Y. Liu, W. Guo, C.-I. Fan, L. Chang, C. Cheng, A Practical Privacy-Preserving Data Aggregation (3PDA) Scheme for Smart Grid, *IEEE Transactions on Industrial Informatics*, 2018, doi: 10.1109/TII.2018. 2809672.

[19] J. Liu, Y. Xiao, S. Li, W. Liang, CL. Chen, Cyber Security and Privacy Issues in Smart Grids, *IEEE Communications Surveys & Tutorials*, Vol. 14, No. 4, pp. 981-997, January, 2012.

[20] T. Logenthiran, D. Srinivasan, T. Z. Shun, Demand Side Management in Smart Grid Using Heuristic Otimization, *IEEE Transactions on Smart Grid*, Vol. 3, No. 3, pp. 1244-1252, September, 2012.

[21] R. Lu, X. Liang, X. Li, X. Lin, X. Shen, EPPA: An Efficient and Privacy-Preserving Aggregation Scheme for Secure Smart Grid Communications, *IEEE Transactions on Parallel and Distributed Systems*, Vol. 23, No. 9, pp. 1621-1631, September, 2012.

[22] A. R. Metke, R. L. Ekl, Smart Grid Security Technology, *2010 Innovative Smart Grid Technologies (ISGT)*, Gothenburg, 2010, pp. 1-7.

[23] R. A. Mollin, *An Introduction to Cryptography*, CRC Press, 2006.
[24] P. Paillier, Public-Key Cryptosystems Based on Composite Degree Residuosity Classes, *Eurocrypt*, Prague, Czech Republic, 1999, pp. 223-238.
[25] R. Petrlic, A Privacy-Preserving Concept for Smart Grids, *Sicherheit in vernetzten Systemen*, 2010.
[26] B. Schneier, *Applied Cryptography: Protocols, Algorithms, and Source Code in C*, John Wiley & Sons, 1996.
[27] H. Son, T. Y. Kang, H. Kim, J. H. Roh, A Secure Framework for Protecting Customer Collaboration in Intelligent Power Grids, *IEEE Transactions on Smart Grid*, Vol. 2, No. 4, pp. 759-769, December, 2011.
[28] P. Vytelingum, T. D. Voice, S. D. Ramchurn, A. Rogers, N. R. Jennings, Agent-Based Micro-Storage Management for the Smart Grid, *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, Toronto, Canada, 2010, pp. 39-46.
[29] L. Yang, H. Xue, F. Li, Privacy-Preserving Data Sharing in Smart Grid Systems, *2014 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, Venice, 2014, pp. 878-883.
[30] Z. Yang, S. Yu, W. Lou, C. Liu, P^2: Privacy-Preserving Communication and Precise Reward Architecture for V2G Networks in Smart Grid, *IEEE Transactions on Smart Grid*, Vol. 2, No. 4, pp. 697-706, December, 2011.

Biographies

Chun-I Fan received the Ph.D. degree from the National Taiwan University, Taiwan, in 1998, and he is now a full professor in National Sun Yat-sen University, Taiwan. Prof. Fan is also the Chairman of the Chinese Cryptology and Information Security Association. His current research interests include applied cryptology, cryptographic protocols, and information and communication security.

Yi-Fan Tseng received the Ph.D. degree in computer science and engineering from National Sun Yat-sen University, Taiwan, in 2018. He joined the faculty of the Department of Computer Science, National Chengchi University, Taipei, Taiwan, in 2019. His research interests include information security, cryptographic protocols, and applied cryptography.

Yi-Hui Lin was born in Kaohsiung, Taiwan. She received the M.S. degree in computer science and engineering from National Sun Yat-sen University, Kaohsiung, Taiwan, in 2015. Her research interests include communication security, information security, and applied cryptography.

Fangguo Zhang received his Ph.D. from the School of Communication Engineering, Xidian University in 2001. He is currently a Professor at the School of Date and Computer Science of Sun Yat-sen University, China. He is the co-director of Guangdong Key Laboratory of Information Security Technology. His research mainly focuses on cryptography and its applications. Specific interests are elliptic curve cryptography, secure multiparty computation, anonymity and privacy, etc.