© 2016 IJSRSET | Volume 2 | Issue 2 | Print ISSN : 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

Privacy Preserved Secured Data Aggregation Technique for Smart Grid Communication System

Saranya. A*, Anantha Prabha. P

Department of Computer Science, Autonomous/Sri Krishna College of Technology, Coimbatore, TamilNadu, India

ABSTRACT

Smart grid application is the most concerned application in the real world environment which will generate different power readings in different time periods. These readings are to be gathered and sent to the centralized server for further processing. The single node failure in the smart grid system might lead to entire system failure where the aggregation can't be performed well. During data aggregation user privacy, fault tolerance and data integrity are to be considered as the important factors by the smart grid system. The technique DPAFT (Differentially Private data aggregation with Fault Tolerance) provides efficient differential privacy with fault tolerance for smart metering. And also DPAFT uses Honest but curious model to maintain data integrity against internal/external attack and End to end signature to accidental attack. Performance evaluation illustrate that DPAFT is efficient in terms of differentially privacy storage cost, computational complexity and robustness of fault tolerance.

Keywords: Data integrity, Privacy Preserving, Data aggregation, Differential privacy

I. INTRODUCTION

The electrical brownout in North America in 2003 affected more than 100 power plants, and endangered tens of millions of people's lives. Later, post surveys exposed that the failure was mainly due to load imbalance in the electric power grid and the lack of active and real-time diagnosis. While rapid scientific and technological advances are driving radical inventions in many fields, today's power grid is unpredictably still grounded on a design the same as 100 years ago. Opportunely, in the last decade, huge numbers of efforts on the development of next-generation power grid, known as smart grid, have been made in many realms around the world.

Compared with traditional power grid, smart grid has presented new concepts and offered promising solutions for intelligent electricity generation, transmission, scattering and exploitation. By deploying various sensors along with the two-way flows of electricity and communication, a enormous amount of real-time information is collected and reported to the control center (CC) for timely monitoring and evaluating the health of power grid, as illustrated in Fig Specifically, all the intelligent electric utilizations in the residential user's home are connected to a key element, *smart meter*, which periodically records the power consumption of appliances and booms the metering data to a local area *gateway*.[12] Then the gateway collects and forward data to the control center for further analysis and processing, e.g. making real-time power pricing decisions and detecting power fraud/leakage. There are usually two ways of collecting data in smart grid, one is at a low frequency and the other is at a high frequency . The low-frequency data contains summary for some periodic power usage, brief enough to elude privacy leakage. The high-frequency data, e.g. those collected every 15 minutes, include specific power usage designs for fine grained optimization and real-time management. As they are related to users' private lives, the high frequency data have to be threatened from utilities. To solve this problem, we can simply use end-to-end data encryption from managers to the control center, but encryption may increase both communication and computation overhead of the control center, specifically when the user set is large. Thus we consider letting the gateway aggregate users' data before reportage them to

the control center, i.e. combine individual users' cipher texts together to get the cipher text of data summation. To protect user confidentiality, the local gateway should not be able to access the content of users' data, i.e. the gateway should not be able to decrypt users' cipher texts. In order to allow the gateway to perform aggregation on cipher texts, homomorphic encryption techniques can be applied. Existing data aggregation schemes apply different homomorphic encryptions to accomplish the same purpose; however, they only consider protection of user privacy against the gateway (aggregator), while the control center is static free to learn individual users' data. This is because private keys the control center keeps can not only be used to decrypt aggregated data, but also be used to reveal any user's electricity usage. This may also struggle residential users' privacy concerns, especially when the control center is vulnerable to some strong adversaries, i.e. the adversary that can cooperation a few servers at the control center and obtain their private keys. Consider a interested control center or a strong adversary that aims to spy on user privacy, these privacy-preserving data aggregation schemes are not robust enough to keep user activities unexposed.[15] Other aggregation schemes such as use another technique: each user surrounds a random number into their cipher text; and the aggregator also embraces a random number, where the sum of all of these casual numbers is 0. During the decryption phase, the random numbers will cancel out such that the aggregator can improve the sum of all users' data without learning individual ones. One major drawback of these existing works is that these schemes are not tolerant to user failures. Even if a single user fails to explosion data at a time point, the servers would not be able to learn anything as the sum of random numbers in informed data is no longer 0. This can be a big anxiety since failures may not be avoidable in reality. On the other hand, while servers at the control center are dependable and robust, they may also sometimes suffer from malfunction or shutdown initiatively to evade certain attacks. Even worse, some of the servers may be compromised by a strong adversary, thus the attendants should also be fault-tolerant or the whole system will fall into paralyzation.

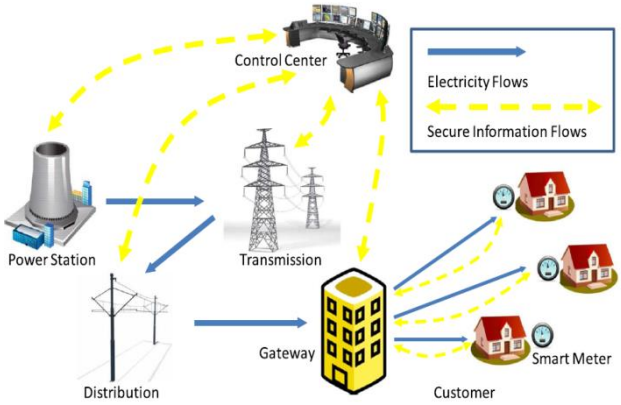


Figure 1. Smart Grid System Architecture

II. METHODS AND MATERIAL

A. Related Work

A privacy-preserving data aggregation scheme with fault tolerance for smart grid communication sle Chen et al [3] proposed a privacy-preserving data aggregation scheme with fault tolerance, named PDAFT, for secure smart grid communications. PDAFT uses the homomorphic Paillier Encryption technique to encrypt sensitive user data such that the control centre can obtain the aggregated data without knowing individual ones, who aims to threaten user privacy can learn nothing even though, already compromised a few servers at the control centre. PDAFT not supports the fault-tolerant feature, i.e., PDAFT cannot work well even when some user failures and server malfunctions occur. PDAFT not only resists various security threats and preserves user privacy, but also has significantly less communication overhead compared with those previously reported competitive approaches. The Paillier encryption is a popular public key encryption scheme, which can achieve the homomorphic properties and is widely desirable in many privacy-preserving applications. Power efficient data gathering and aggregation in wireless sensor networks huseyin et al [6] proposed method two new algorithms under name PEDAP (Power Efficient Data gathering and Aggregation Protocol), which are near optimal minimum spanning tree based routing schemes, where one of them is the power-aware version of the other. With the use of data fusion and aggregation techniques, while minimizing the total energy per round, if power consumption per node can be balanced as well, a near optimal data gathering. The PEDAP protocols assume the locations of all nodes are known by base station. The routing information is

computed using Prim's minimum spanning tree algorithm where base station is the root. Initially, put a node in the tree which is the base station in our case. After that, in each iteration select the minimum weighted edge from a vertex in the tree to a vertex not in the tree, and add that edge to the tree. Repeat this procedure until all nodes are added to the tree. by computing a minimum spanning tree over this graph with the cost functions given as above and by routing packets according to that spanning tree, to achieve a minimum energy consuming system. Switching between the two proposed algorithms requires only a small change in the base station and no changes in sensor nodes. PEDAP takes it further and tries to balance the load among the nodes. Minimizing the total energy of the system while distributing the load evenly to the nodes has a great impact on system lifetime. Efficient algorithms for maximum lifetime data gathering and aggregation in wireless sensor networks Kalpakis et al [2] proposed a near-optimal polynomial-time algorithm for solving the MLDA(Maximum Lifetime Data gathering with Aggregation)problem. The proposed algorithm, while performing significantly better than existing protocol in terms of system lifetime, is computationally expensive for large sensor networks. It describes clustering-based heuristics approaches for maximum lifetime data gathering and aggregation in large-scale sensor networks. Data aggregation performs in-network fusion of data packets, coming from different sensors enroute to the base station, in an attempt to minimize the number and size of data transmissions and thus save sensor energies. Finally, provide (i) for smaller sensor networks the MLDA algorithm achieves system lifetimes that are times better when compared to an existing data gathering protocol, (ii) for larger networks, our clustering-based heuristics achieve as much as a factor of increase in the system lifetime when compared to the same protocol. An efficient and privacy-preserving aggregation scheme for secure smart grid communications**.** Rongxing Lu et al [5] proposed an efficient and privacy-preserving aggregation scheme, named EPPA, for smart grid communications. EPPA uses a super increasing sequence to structure multidimensional data and encrypt the structured data by the homomorphic Paillier cryptosystem technique. For data communications from user to smart grid operation centre, data aggregation is performed directly on cipher text at local gateways without decryption, and the aggregation result of the original data can be obtained at the operation centre. EPPA also adopts the batch verification technique to reduce authentication cost. EPPA resists various security threats and preserve user privacy, and has significantly less computation and communication overhead than existing competing approaches. A lightweight privacy-preserving data aggregation scheme for smart grid.Xiaodong Linet et al [7] proposed an efficient lightweight privacy-preserving Data aggregation scheme, called LPDA, for smart grid. The proposed LPDA is characterized by employing one-time masking technique to protect user's privacy while achieving lightweight data aggregation. Detailed security analysis has shown that the proposed LPDA scheme is robust against many security and privacy threats in smart grid. In this scheme, aggregation is performed in a distributed manner in accordance to the aggregation tree, where each node collects data from its children, aggregates them with its own data, and sends the intermediate result to the parent node. Homomorphic encryption is employed to protect the privacy of the electricity use data, so that inputs and intermediate results are not revealed to smart meters on the aggregation path, while the aggregation is still correctly performed. . In smart grid, through a variety of sensors and smart electric meters installed on the power grid, the power grid can be monitored effectively to be more reliable and power companies can also control energy consumption through real-time pricing, especially, higher prices at peak times due to higher demand. Also, consumers can benefit from it, for example, reducing their electricity bills by lowering their power consumption at peak times.

B. Problem Formalization

In this section, we validate our research problems in smart grid communications, The main problem observed in the existing work is, cannot support fault tolerance and privacy preserving. Once the user fails to report, the whole data aggregation protocol is not workable [9]. Therefore, fault tolerance is a big concern for smart grid communications, because smart meters, as low cost devices and running in unprotected environments, are prone to failure. User's private data may often suffer from differential attack. This problem is resolved in the proposed work by using the DPAFT including system model, security requirements, and design goal.

C. System Model

Since residential users permanently care about their privacy when reporting their detailed electricity usage data to the control center in smart grid communication, in this work, we mainly focus on how to let the control center compute multiple arithmetic functions of users' data, such as average, variance, one-way ANOVA, etc., in a privacy-preserving way. Specifically, in our system model, we consider a representative smart grid communication architecture for residential users, which contains of a trusted authority (TA), a control center (CC) in charge of communication and control of the system, a local gateway (GW), and a large number of residential users U = {*U*1,*U*2, · · · ,*Un*} in a residential area (RA).The local GW is a controlling workstation that connects the CC and residential users, i.e., support the CC in collecting residential users nearly immediate electricity usage data. The communication between residential users and the GW is via Wi-Fi technology as suggested in the Standards. While the communication between the GW and CC is via wired links with high bandwidth and low delay. The concern of GW in our system is mainly twofold, one is collecting and relaying users' data, and the other is performance some aggregations based on the CC's requirements.

Servers: *(*S = {*S*1, *S*2, . . . , *Sk*}*) at CC* The CC is comprised of a set of servers S = {*S*1, *S*2, . . . , *Sk*}, $k \geq 3$, which run synergistically to collect, process and evaluate the nearly real-time data for providing reliable services for electrical grid, e.g., real-time monitoring the RA's usage for leakage detecting, fraud detecting, and forecasting . Unlike the TA, servers S = {*S*1, *S*2, . . . , *Sk*} are powerful entities in our system, but some of them could be compromised or paralyzed by a strong adversary. As S are powerful entities, it will take huge costs for an adversary to compromise even a single server. As a result, it is realistic for us to assume that the adversary can only compromise a limited number, i.e. no more than $d = \lfloor k/2 \rfloor - 1$, of servers. In other words, the adversary can only compromise minority of the servers S = {*S*1, *S*2, . . . , *Sk*}.

Gateway (GW) The GW is a powerful entity, which connects the CC and residential users, i.e., helping the CC to collect the residential users' nearly real-time usage data. The communication between residential users and the GW is relatively inexpensive Wi Fi technology, while the communication between the CC and the GW is through either wired links or any other links with high bandwidth and low delay.

Residential Users U = {*U*1,*U*2, · · · ,*Un*} Each residential user *Ui* ∈ U is equipped with a smart meter and various smart applications to form a Home Area Network(HAN), which can electronically record the real-time electricity usage data, and report to the CC via the GW in a certain period, i.e., every 15 minutes. As smart meter is not as powerful as the GW, some meters could be faulty occasionally, i.e., they could stop reporting for a while and will be reset in a late time. However, malfunction of smart meter can be viewed as a rare event in reality.

D. Security Requirements

While broadcasting fine-grained electricity usage data, users are also worried about leaking privacy of their activities. In our security model, we consider avoiding the adversary from revealing the individual user's electricity usage data, and at the same time allowing the CC to compute multiple functions upon users' data. Specifically, we consider the CC and GW are both trustable, and the users U = {*U*1,*U*2, · · · ,*Un*} are all honest. However, there exists a malicious adversary *A* eavesdropping the communication flows between users and the GW, and those between the GW and CC. In addition, the adversary *A* could intrude into the databases of GW and CC to steal the stored data. More seriously, based on the collected data obtained, the adversary *A* may launch some differential attacks to acquire the individual user's data. Therefore, we consider achieving the following security requirements to avoid the adversary *A* from disclosing individual user's sensitive data in smart grid communications. [10]Note that we mainly focus on protecting user data privacy in this paper, confirming integrity of user data is out of the scope of our work. In fact, by adding some authentication methods at the GW, the integrity of users' data could be ensured. A could also intrude in the database of the GW and the smart grid operation center to take the individual user reports.

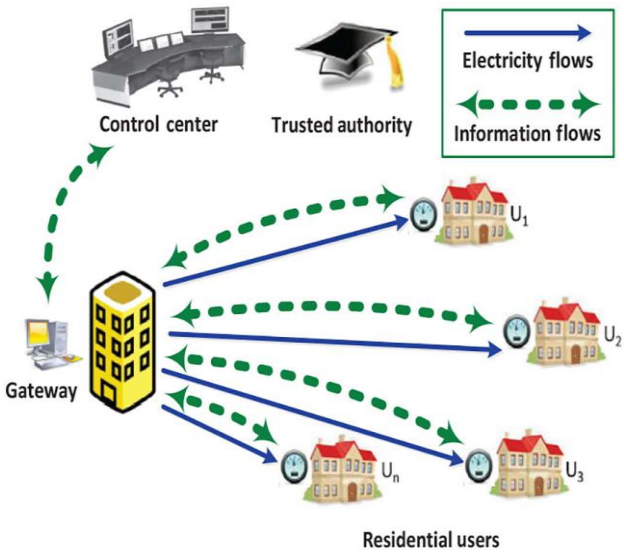


Figure 2. Overall System Design

Data Confidentiality: As the adversary *A* may reside in the RA to overhear the communication flows, data Transferred from residential users to the GW and that from the GW to CC should both be encrypted. In order to aggregate users' data, the GW could first decrypt the data informed from users, aggregate them together, and then encrypt the aggregation again and transmit it to the CC. However, this way may cost much time on decrypting all users' data.[11] More seriously, if the adversary *A* intrudes into the database of GW, then all users' private data is disclosed. As a result, the GW should be able to aggregate users' data in a privacy preserving way, without decryption. Then the confidentiality of users' private data can be ensured[1].

Differential Privacy: Although some controlling adversary may also intrude into the database of CC, the Individual user's data won't be released directly since the CC only stores aggregated data. Since we allow the CC to compute multifunctional aggregations, the adversary *A may* utilize multiple aggregations of similar sets to launch the differential attack and obtain the individual user's private data. Therefore, if differential privacy of users' data is also required in the secure smart grid communications.

Data integrity: Validating an encrypted report that is really sent by a legal residential user and has not been altered during the transmission, i.e., if the adversary A give up and/or modifies a report, the malicious operations should be detected [13]. Where an adversary is usually the contributors of the protocol including the GW or the CC, which could access or misuse the information of residential users to concession their privacy, or the curious residential users, who actively seek or infer other users' private usage data.

E. Design Goal

Under the above-mentioned system model and security requirements, our design goal is to develop an efficient privacy-preserving aggregation system with fault tolerance for smart grid communications.[4] Specifically, the following three objectives should be achieved.

The security requirements should be satisfied in the proposed aggregation scheme.

As stated above, if the smart grid does not reflect security issues, the residential users' privacy could be disclosed, and the smart grid cannot step into its display. Therefore, the proposed scheme should achieve the above security requirements consequently.

The communication-effectiveness should be achieved in the proposed aggregation scheme.

Although the communication between the GW and the CC is featured with high-bandwidth and low-latency, to support a large number of residential users' reports to the CC at almost the same time, the proposed aggregation scheme should also consider the communication-effectiveness, so that the near real-time user reports can be professionally transmitted to the CC[5].

The fault tolerance should be guaranteed in the proposed aggregation scheme.

Since some smart meters could be destroyed and d servers could possibly be compromised by the adversary *A*, the proposed aggregation scheme should also be fault tolerant, i.e., $k - d$ uncompromised servers can still improve the aggregated data from non-malfunctioning smart meters[8].

F. Preliminaries

In this section, we briefly recall the ideas of Boneh-Goh-Nissim cryptosystem which serve as the basis of the proposed scheme.

Boneh -Goh-Nissim Cryptosystem

The Boneh-Goh-Nissim cryptosystem is a public key encryption scheme that proposed by Boneh, Goh and Nissim . It has been widely used in many privacy-preserving applications since it can achieve some nice homomorphic properties. Definitely, the Boneh-Goh-Nissim encryption is comprised of three algorithms: key generation, encryption and decryption as follows.

Key Generation: Given the security parameter $\tau \in Z+$, run $G(\tau)$ to obtain the tuple $(p, q, G, G1, e)$ as defined above. Randomly chose two generators $g, x \in G$ and set $h = xq$. Then h is a random generator of the subsection of G of order p. The public key is $PK = (N, G, G1, e, g, h)$. The private key is $SK = p$.

Encryption: Given a message $m \in \{0, 1, ..., W\}$, where $W << q$ is the destined of the message space, choose a random number $r \in ZN$. Then the cipher text can be calculated as $C = gm \cdot hr \in G$.

Decryption: Given the private key $SK = p$ and the cipher text $C \in G$, first compute $Cp = (gm \cdot hr)p = (gp)m$. Let $gp = gp$, then $Cp = gmp$. To recover m, it suffices to compute the discrete logarithm of gmp .Note that when m is a message, say $m \leq W$ for some small bound W, the decryption takes expected time $O(\sqrt{W})$ using Pollard's lambda method. The Boneh-Goh-Nissim cryptosystem also has some nice homomorphic properties. Firstly, it is additively homomorphic. For any cipher texts $C1, C2 \in G$ of messages $m1, m2 \in \{0, 1, ..., W\}$, to obtain the cipher texts of $m1 + m2$, one can simply compute the product $C = C1C2hr$ for a random $r \in ZN$. In addition, one can also obtain the product of two messages by exploiting the bilinear map, i.e. by computing $C = e(C1, C2)e(g, h)r \in G1$, and decrypt it similarly in the group $G1$. Note that the homomorphic multiplication can be taken only *once* upon two cipher texts in G, and then the result will be in G1, but it still supports additive homomorphism.

III. RESULTS AND DISCUSSION

1. Proposed Method

A. System Initialization

The single TA can bootstrap the whole system in the beginning. Concretely, for the system initialization, given the security parameters τ , the TA first runs the algorithm $\zeta(\tau)$ to generate the tuple (p, q, G). Then, the TA builds up the Boneh–Goh– Nissim cryptosystem and acquires the tuple (N, G, g, h), where $N = pq$, $g \in$ is a random generator of G, and $h = gq\beta$ (for some private β) is a random generator of the subgroup of G of order p.

Finally, the TA publishes (N, G, g, h) as the public keys of our system. Although some private keys may also be assigned to users for the sake of authentication, since it is not our main focus, we do not consider it in this work. In addition, via performing the following steps, the key materials of the residential users.

Step 1) For each user $Ui \in U$, the TA first chooses a random number $si \in ZN$ and assigns si as Uis private key.

Step 2) The TA computes $s0 \in ZN$ such that $s0 \cdot (s1 + s2 + \cdot \cdot \cdot + sn) = 1 \bmod p$.

Step 3) The TA assigns $s0$ as the CC's private key.

Step 4) The TA computes Yi, such that $Yi = hs0si$, for $i = 1, \ldots, n$, and assigns Yi as the CC's private key as well.

B. Data aggregation Request Relay

Data aggregation is any procedure in which evidence is gathered and for purposes such as statistical analysis. Aggregation reduces the amount of network traffic which helps to reduce energy consumption on sensor nodes.

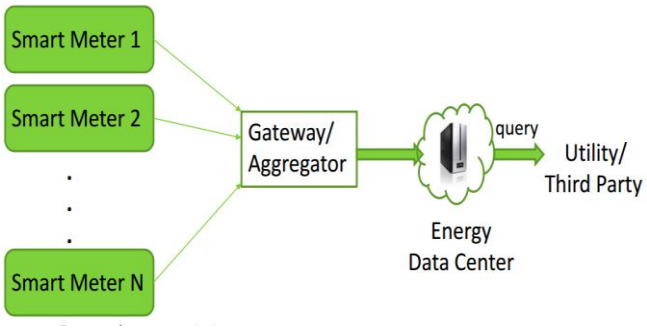


Figure 3. Data Aggregation

Fig 3. represents the collecting and analysing nearly real time electricity usage data. Gateway is used for collecting and analysing nearly real time electricity usage data at from all residential users in the residential

area. The control centre monitor the health of the whole smart grid system and further provides various high quality services. At the time point the control centre launches the request for collecting the usage data in the residential area. It collected every 15 minutes, contain personalized power usage patterns, which are highly relevant to user's privacy, thus must be protected from unauthorized entities. To encrypt user's data can aggregate all users' data without decryption. The control centre may not only utilized to decrypt the aggregated data, but also revealed any user's electricity usage. The CC launches a request for collecting the usage data in the RA (Residential Area)as follows.

Step 1: The CC first selects a random $r \in ZN$, and computes $A1 = gr$ and $A2 = hSor$.

Step 2: Then, the CC sends *A1* and *A2* to the GW.

After receiving *A1* and *A2* from the CC, the GW(Gateway) performs the following steps to relay the data aggregation request.

Step 1: The GW first selects a random $t \in ZN$, and computes $A3 = At\ 1 = grt$ and $A4 = At\ 2 = h\ So\ rt$.

Step 2: Then, the GW sends A3 and A4 to each Ui ∈ U, respectively.

Where r: secret key, h: random generator of subgroup of G having order P. *A1,A2*: Aggregated data.

Thus, the users data are all the intelligent electric appliances in the residential user's home are connected to a key element, smart meter, which periodically records the power consumption of applications and reports the metering data to a local area gateway and data to be aggregated.

C. User Report Generation

The security requirements should be guaranteed in the proposed scheme. The smart grid does not consider the security, the residential users' secrecy could be disclosed, and the real-time electricity use reports could be altered. Then, the smart grid cannot step into its flourish. Therefore, the proposed scheme should achieve the confidentiality, authentication, and data integrity requirements simultaneously.

Each user Ui ∈ U, after collecting its usage data mi ∈ {0,1, . . .,W} at time point tγ, performs the following steps to report the encrypted and noisy usage data.

Step 1: Ui first calculates the value $C\tilde{}i = A^{mi+G1(n,\lambda)-G2(n,\lambda)} A^{si} = g^{rt(mi+G1(n,\lambda)-G2(n,\lambda))} h^{rts0si}$, where G1(n, λ) and G2(n, λ) represent two random values independently sampled from the same gamma distribution, i.e., G1(n, λ) and G2(n, λ) are random variables having gamma distribution with the $g(x, n, \lambda) = 1/\lambda^{1/n} / \Gamma(1/n)\, x^{1/n-1}\, e^{-x/\lambda}$, where $x \geq 0$, and n is the number of all the smart meters.

Step 2: Ui then reports C˜i to the GW.

D. Privacy Preserving Report Aggregation

This scheme provides privacy-preserving aggregation beside, which may compromise a few servers at the control centre, and cares fault tolerance of smart grid users and servers. As a result, user data can be confidentially and reliably reported to smart grid control centre for real-time monitoring. First, an external attacker *cannot* disclose users' private usage data even though can eavesdrop the communication flows. Second, although can deploy some undetectable malwares to the GW or the CC, it still cannot disclose users' private usage data. Third, through eavesdropping and analysing all the inputs, intermediate communication flows and outputs that are not of one's own, any participant running in the honest-but-curious ad model cannot infer useful knowledge about residential users' privacy. Finally, cannot launch differential attack to obtain the individual user's privacy successfully[8]. The system can still aggregate the data of functioning meters effectively and efficiently even in the presence of malfunctioning ones.

STEP1: The GW first aggregates the encrypted and noisy measurements of all the users as

$$C_{\tilde{\gamma 1}} = \left(\prod_{i=1}^{n} \hat{C}_i\right)^{t^{-1}} \qquad (1.1)$$

$$\left(g^{r\sum_{i=1}^{n}(m_i+G_1(n,\lambda)-G_2(n,\lambda))} \times h^{rs0\sum_{i=1}^{n} s_i}\right)$$

Then, the gateway first aggregates the received cipher text Ci. For i=1,2,…n

STEP 2: The GW then complements the M noises of the malfunctioning smarter meters as

$$C_{\gamma 1}^{\sim} = C_{\gamma 1}^{\sim} \prod_{i=1}^{M} A_1 \; g^{r\sum_1^n (G_1(n,\lambda)-G_2(n,\lambda))} = g^{r\sum_{i=1 ui\in u/\hat{u}}^n (m_i+G_1(n,\lambda)-G_2(n,\lambda))}$$

$$= g^{r\sum_{Ui\in U/\hat{U}=1}^M mi)+Lap(\lambda))_h rso^{\sum_{ui\in U/\hat{u}}^n s_i}}$$

STEP 3: The GW finally sends and$C_{\gamma 1}^{\sim}$ ˆU to the CC. G**:**Cyclic group of order N=pq where p,q are prime numbersH**:**random generator of subgroup of G having order P**.**

C:Cipher text

M:short message

S0,Si:Assign private key for cc

Cγ:Encrypted data

From equ (1.1) and (1.2) thus, all the n smart meter work correctly after receiving total encrypted measurements Ci, for i=1,2,….n at the point. The gateway report the aggregated and encrypted data at the time point.

2. Security Analysis

In this section, we will discuss the security issues involved in the proposed aggregation scheme DPAFT, in particular, to protect the users' electricity usage privacy against a strong adversary *A*.

• The users' electricity usage privacy is protected from eavesdropping

As stated in our security model, an adversary *A* may reside in the residential area to eavesdrop the communication flows from users to the GW. Suppose *A* has eavesdropped a cipher text of user *Ui* at time point *tγ* , i.e. *gmi,γ · hkuiγ* . Since the electricity usage *mi,γ* within 15 minutes is probably a small value, the adversary *A* may try to launch a brute-force attack by exhaustedly testing each possible value of *mi,γ* , but before that, *A* needs to know *hkui γ* first, which is impossible if *kui* is unknown to *A* . Thus the privacy of users' electricity usage is guaranteed.

• The uncompromised users' electricity usage will not be revealed.

In our system, we consider that the adversary *A* can compromise some of the users, in that case, the privacy of compromised users is fully exposed. However, since there are a large number of users, the adversary *A* is discouraged to use this inefficient method to disclose more users' privacy. Instead the adversary *A* may try to threaten the uncompromised users' privacy by utilizing the secret information he obtained from the compromised ones, i.e. their private keys. Nevertheless, this attack won't success either since the privacy key of each user is randomly chosen by the TA, knowing one user's private key reveals nothing about another one's. Moreover, even if the adversary *A* compromises *n* – 1 users and obtains their private keys, he still cannot reveal the last user's private key and electricity usage, since the sum of all users' private keys is unknown to *A*.Thus the privacy of uncompromised users is also preserved.

• The users' private usage and sum usage data will not be disclosed at the GW.

At each time point *tγ* , the GW collects all users' cipher texts and directly aggregates them into one single cipher text. Thus if the adversary *A* deployed some undetectable malwares into the GW, he could only get the cipher texts of all users and the aggregated one. Since the GW does not decrypt any user's electricity usage data, the adversary *A* still cannot get any user's private usage data. Moreover, the aggregated cipher text is *Cγ = g_ni*=1 *mi,γ ·h_i*=1 *kuiγ* mod *N*2, which has the same form of one user's cipher text *gmi,γ · hkuiγ* . Since the sum of all users' private key is not known to the adversary *A* , similarly, the adversary *A* can neither disclose the sum usage of all users. Thus the users' electricity usage data privacy can be ensured even though the GW is deployed some undetectable malwares by the adversary.

3. Performance And Cost Evaluation

In this section, we evaluate the performance of the proposed DPAFT in terms of storage cost, computation complexity, utility of differential privacy, robustness of fault tolerance, and efficiency of user addition and removal. Because few of the existing schemes supports fault tolerance and differential privacy simultaneously, in this section, we compare our proposed scheme with the state-of-the-art schemes which support privacy-preserving aggregation with fault tolerance or differential privacy.

A. Storage Cost

In the scheme of it is necessary for the GW to configure the huge amount of memory buffers to store the future cipher texts for all the residential users. Even though such overhead is acceptable when the scale of the scheme is not large as pointed out by the authors in when the scale of the scheme turns large, e.g., with millions of smart meters, it brings considerable storage overhead to the GW. By contrast, in our scheme, the GW is just responsible for data aggregation and packages relay, thus there is no special storage requirements, which makes it more rational and practical

B. Computation Complexity

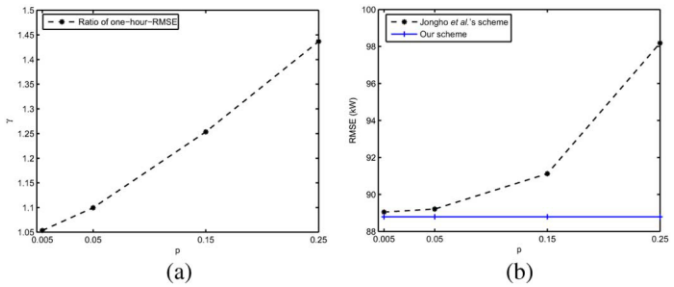
In the scheme of each user should select k other users as partners to encrypt the measurements. Specifically, in the initial setup phase of their scheme, the shared secret keys between every two users of the partner pairs should be generated and assigned secretly. Then, in dada report phase, the two parts of cipher texts, i.e., the current cipher text and the future cipher text, should be calculated and reported. Each user generates the current cipher text by adding the random number and the noise information to the real measurement. The random number is computed by using the shared secret key assigned in initial setup phase and the information of reporting time point. Then, the future cipher text should also be generated and reported simultaneously for achieving fault tolerance. In our proposed scheme, each user independently reports the measurement, thus there is no need to compute and assign the shared secret keys among the users. The additional computation of future cipher text is not necessary either. Our DPAFT only needs two exponentiation operations and one multiplication operation totally for reporting the measurement of each user. The computation complexity is less than or at least not heavier that the scheme . Because our scheme is secure in *honest-but-curious* adversary model as illustrated in aforementioned while the scheme is not constructed in this security model, other comparison of computation complexity for the GW and the CC cannot be carried out. Nevertheless all the computation cost of the GW and the CC include only a handful of exponentiation, add, and multiplication operations, which leads to the low computation complexity of our scheme. Since the techniques to support data aggregation of our scheme are also mainly inspired by the Boneh– Goh–Nissim additively homomorphic cryptosystem and the scheme is the most efficient one based on Boneh– Goh–Nissim cryptosystem, we also compare our scheme with in terms of computation complexity. In the most time consuming phase of *secure report reading*, both of our scheme and [12] take two exponentiation operations and one multiplication operation. Additionally, in order to support fault tolerance, for each round of data aggregation, our scheme costs two exponentiation operations and one multiplication operation in *data aggregation request* phase, and two exponentiation operations in *data aggregation request relay* phase, respectively. Since all these operations are performed by the entities of the CC or the GW, and both of which are with powerful computational capability, thus, this increases negligible computation overhead. Therefore, with additional negligible computational cost, our scheme achieves fault tolerance while the scheme of cannot support.

C. Utility of Differential Privacy

Similar to the scheme we implement an electricity consumption simulator having the ability of generating realistic 1-min consumption traces synthetically. It is extended from the basic simulator. We produce traces for 2000 households based on this simulator. Specifically, the distribution of the residents of each household consults the U.K. statistics on household sizes in 2011 . We select the day to be a weekday in February. For the appliances in a household, we choose them randomly among 33 available ones. For differential privacy, ε is set to 1, and the global sensitivity is set to 33 kW, which is the sum of power demands of all the appliances and lights.

D. Robustness of Fault Tolerance

When some meters fail to report, the smart meter system still needs to be able to aggregate the measurements of the remaining functioning ones to perform real-time data monitoring and analyzing successfully. In the scheme of the GW stores *B* pieces of *future cipher texts* for each smart meter to support fault tolerance. Without loss of generality, assume that the data report interval of the smart meter system is *T*. And suppose at T me point *Ta*, due to some fault, certain smart meter *Ui* cannot report the measurement successfully to the GW, and the fault recovery time point is *Tb*. Thus, the fault duration period *T*per is *Tb* − *Ta*. If *T*per > *B* · *T*, the system of cannot tolerate the fault any longer after the time point of *Ta* + *B* · *T*, because the prestored *future cipher texts* are used up, until the fault smart meter *Ui* is to be recovered again at *Tb*. The robustness of fault tolerance turns to be much worse when the number of the malfunctioning smart meters increases. In order to support more robust fault tolerance, the system parameter of the buffer size *B* of should be increased further. However, this causes heavy storage cost, computation complexity, and communication overhead as illustrated in the aforementioned performance analysis. By contrast, our scheme is more robust of fault tolerance and can support data aggregation with any rational number of malfunctioning smart meters with arbitrary long fault period, because the mechanism of our fault tolerance is not related to the malfunctioning smart meters directly and is independent of any external factors, e.g., *future ciphertexts*.



IV. CONCLUSION

In this paper, does not have capabilities for integrity check. Hence, it is vulnerable to accidental errors, as well as compromised/dishonest meters and other fake data injection attacks. In this paper, we first introduce an end-to-end signature scheme using homomorphic signatures. A checksum of the aggregation is generated and updated along with their-network aggregation process. With minimum overhead, it enables the collector to check the integrity of the aggregation result. However, such mechanism becomes insufficient in the presence of cyber-attacks, i.e., when forged data is injected by compromised meters or communication channels. We further present a hop-by-hop signature scheme and an incremental verification mechanism to defend against such attacks. In this solution, output from each smart meter is signed, and signatures are kept at parent nodes. The collector device initiates an incremental verification of signatures when suspicious aggregation results are received. The ex post facto verification process is computationally inexpensive, while certifying faithfulness and undeniability properties.

V. REFERENCES

[1] S. Ozdemir and Y. Xiao, "Secure data aggregation in wireless sensor networks: A comprehensive overview," Comput. Netw., vol. 53, no. 12, pp. 2022–2037, Aug. 2009.

[2] L. Wasserman, All of statistics : a concise course in statistical inference. New York: Springer.

[3] C. T. Chou, A. Ignatovic, and W. Hu, "Efficient computation of robust average of compressive sensing data in wireless sensor networks in the presence of sensor faults," Parallel and Distributed Systems, IEEE Transactions on, vol. 24, no. 8, pp. 1525–1534, Aug 2013.

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

[5] M. Hashmi, S. Hanninen, and K. Maki, "Survey of smart grid concepts,architectures, and technological demonstrations worldwide," in IEEE PES Conference on Innovative Smart Grid Technologies, 2011.

[6] X. Li et al., "Securing smart grid: Cyber attacks, countermeasures, andchallenges," IEEE Commun. Mag., vol. 50, no. 8, pp. 38–45, Aug. 2012

[7] 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., to be published.

[8] F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart gridsusing homomorphic encryption," in Proc. 1st IEEE Int. Conf. Smart GridCommun. (SmartGridComm), 2010, pp. 327–332.

[9] A. Metke and R. Ekl, "Smart grid security technology," in Innovative Smart Grid Technologies, Jan. 2010, pp. 1–7.
[10] Efthymiou C, Kalogridis G (2010) Smart grid privacy via anonymization of smart metering data. In: SmartGridComm, pp 238–243.
[11] P. Paillier, "Public-Key Cryptosystems Based on Composite Degree Residuosity Classes," Proc. 17th Int'l Conf. Theory and Application of Cryptographic Techniques (EUROCRYPT), pp. 223-238, 1999.
[12] P. McDaniel and S. McLaughlin, "Security and privacy challenges in the smart grid," Security Privacy, IEEE, vol. 7, no. 3, pp. 75 –77, 2009.
[13] J. Lu, D. Xie, and Q. Ai, "Research on smart grid in China," in Transmission Distribution Conference Exposition: Asia and Pacific, 2009, Oct. 2009, pp. 1 – 4.
[14] K. B. Frikken and J. A. Dougherty, IV, "An efficient integrity-preserving scheme for hierarchical sensor aggregation," in Proceedings of the first ACM conference on Wireless network security, 2008, pp. 68–76.
[15] T. E. Gamal, "A public key cryptosystem and a signature scheme based on discrete logarithms," in Proceedings of CRYPTO 84 on Advances in cryptology. Springer-Verlag New York, Inc., 1985, pp. 10–18.