Hindawi Publishing Corporation
International Journal of Distributed Sensor Networks
Volume 2013, Article ID 230140, 8 pages
http://dx.doi.org/10.1155/2013/230140

Research Article

Perturbation-Based Schemes with Ultra-Lightweight Computation to Protect User Privacy in Smart Grid

Wei Ren,[1,2] Liangli Ma,[3] and Yi Ren[4]

[1] *School of Computer Science, China University of Geosciences, Wuhan 430074, China*
[2] *Shandong Provincial Key Laboratory of Computer Network, Jinan 250014, China*
[3] *School of Electronic Engineering, Naval University of Engineering, Wuhan 430033, China*
[4] *Department of Computer Science, National Chiao Tung University, Hsinchu 30010, Taiwan*

Correspondence should be addressed to Wei Ren; weirencs@cug.edu.cn

Received 22 August 2012; Revised 4 February 2013; Accepted 5 February 2013

Academic Editor: Sunho Lim

In smart grid, smart meters are deployed to collect power consumption data periodically, and the data are analyzed to improve the efficiency of power transmission and distribution. The collected consumption data may leak the usage patterns of domestic appliances, so that it may damage the behavior privacy of customers. Most related work to protect data privacy in smart grid relies on cryptographic primitives, for example, encryption, which induces a large amount of power consumption overhead. In this paper, we make the first attempt to propose solutions without any cryptographic computation to protect user privacy. The privacy in smart grid is formally defined in the paper. Three schemes are proposed: random perturbation scheme (RPS), random walk scheme (RWS), and distance-bounded random walk with perturbation scheme (DBS). Three algorithms are also proposed in each scheme, respectively. All schemes are ultra-lightweight in terms of computation without relying any cryptographic primitive. The privacy, soundness, and accuracy of proposed schemes are guaranteed and justified by strict analysis.

1. Introduction

Smart grid is a typical application of Internet of Things, M2M, or IP-based sensor networks. It has been envisioned as a key method to reduce the emission of carbon dioxide and retard climate changes, by improving the efficiency of power distribution and transmission.

Smart grid relies on smart meters to collect power consumption data at user ends instantly. Smart meters report the power consumption data periodically to smart grid control center (SGCC). SGCC thus can allocate necessary power distribution and schedule required power transmission. In addition, the SGCC can relocate the power requirements at user ends by delivering power price to users. Users thus can schedule the usage of their household appliances according to the forthcoming price.

As smart meters report the power consumption data periodically, the data may leak user privacy in daily life. For example, the data may be used for deducing user behavior patterns, such as when she gets up according to the data of using microwave oven or toaster in the morning, when she goes back home according to the data of using electric stove for cooking at afternoon, or when she takes bath or goes to bed at night according to the data of using water heater or lamps. Such privacy concerns have already been acknowledged and reported by NIST [1] and significantly affect the deployment of smart meters.

Although there exist several privacy protection or security improvement for smart grid currently [2–6], most of them rely on cryptographic primitives, for example, encrypting the uploading data at smart meters. Cryptographic operations are usually not lightweight, so that they will induce extra power consumption at smart meters. In addition, the data uploading may occur frequently and periodically, so the computation for data encryption occurs extensively. For example, data are uploaded to SGCC once in 10 minutes. The encryption for the data has to be 144 times a day. Thus, the energy consumption for encryption computation

would be large for a month even at single smart meter. Moreover, the extra power consumption will be accumulated to an unsatisfactory waste, because the number of smart meters in smart grid is huge. Furthermore, the decryption computation at SGCC has to be conducted if the uploading data are encrypted at smart meters. The energy consumption of decryption at SGCC will thus extremely increase. Last but not least, the smart meters usually have resource and power constraints, like traditional sensors. As the privacy protection must be conducted at smart meters, any computation for privacy protection should cost low energy to tackle these constraints. The frequent encryption operations are undesirable. Even though the encryption is lightweight in certain situations, the key management for encryption is also a difficult issue for deployment. Therefore, privacy protection by encryption unfortunately contradicts the intention of smart grid for saving energy; an ultra-lightweight method without any cryptographic computation for privacy protection is mandatory for a long run and a large scale.

In this paper, we propose perturbation-based schemes with ultra-lightweight computation without any cryptographic computation. Besides, we strictly and formally define and proof its privacy protection strength. We adapt a rigorous method to state, present, and analyze the privacy protection achievements. All our presentations strictly follow the formal expressions for better clarity and generality.

The contributions of the paper are listed as follows: (i) we propose ultra-lightweight privacy-protection schemes in terms of computation (and thus energy consumption) without any cryptographic computation; (ii) we strictly define the requirements on privacy, soundness, and accuracy in smart grid and proof the guarantee of those requirements.

The rest of the paper is organized as follows. In Section 2 we discuss the basic assumption and models used throughout the paper. Section 3 provides the detailed description of our proposed models and analysis. Section 4 gives an overview on relevant prior work. Finally, Section 5 concludes the paper.

2. Problem Formulation

2.1. Network Model. Two major entities exist in smart grid: smart meter (denoted by SM hereafter) and SGCC.

SM computes power consumption data and uploads them to SGCC periodically. The period for computing power consumption data at SM is called sensing period. The period for uploading power consumption data to SGCC is called uploading period. Without loss of generality, suppose the sensing period and uploading period are both t minutes. The sensing times and uploading times in a day will thus be $n = [24 * 60/t]$. The total sensing data for a day are denoted as a set $DATA_s = \{ds_1, \ldots, ds_n\}$. The total uploading data for a day are denoted as a set $DATA_u = \{du_1, \ldots, du_n\}$. If SM does not hide $DATA_s$, $DATA_s$ will be the same as $DATA_u$.

In smart grid, utility price may vary in different time slots. The price information is delivered by SGCC in advance. Users use such information to guide the power consumption. SM receives such information to calculate utility charge in a month for users. Suppose the prices for n uploading periods in a day are denoted as a set $PRICE = \{p_1, p_2, \ldots, p_n\}$. Thus, the total utility charge for a day is $\sum_{i=1}^{n} du_i * p_i$. The total utility charge for a month is the summation of charges for all days in this month. If the sensing data are changed into the uploading data for protecting privacy, the total utility charge for a day should be remained correct.

2.2. Attack Model and Trust Model. Only adversaries who attack user privacy are considered in this paper. Adversaries can eavesdrop the channels between SM and SGCC; those are denoted as $\mathscr{A}_c$. Adversaries at SGCC can access all uploading data by SM; those are denoted as $\mathscr{A}_s$. Both adversaries desire to deduce the user behaviors in a day by analyzing the uploading data from SM, namely, $DATA_u$. As $\mathscr{A}_c$ and $\mathscr{A}_s$ have the same view on $DATA_u$, we further do not distinguish those two adversaries. Both are denoted by the same notation $\mathscr{A}$.

SGCC is untrustworthy, as we assume adversaries at SGCC are interested in user privacy. SM should be trustworthy. It is a prerequisite for any further discussion, sensing data are at SM, and all possible solutions are conducted at SM. Besides, if SM is untrustworthy, users will not choose them. SM can be easily evaluated and authorized by a Trusted Third Party (TTP).

2.3. Security Definition and Design Goal. Informally speaking, the privacy is guaranteed if the adversaries (not only at SGCC but also at channels between SGCC and SM) cannot deduce the user activities in a day. More specifically, we formally state the privacy requirement definition as follows.

Definition 1. User activities. They are the activities that damage user privacy and are related to using one or multiple household appliances in a daily life. They are denoted as a set $ACT = \{a_1, a_2, \ldots, a_m\}$, where a_i $(i = 1, \ldots, m)$ is an activity related to one or multiple appliances.

Definition 2. Deduce. It means an activity in ACT can be inferred by data in $DATA_b$. If an activity $a_j \in ACT$ is inferred by data $d_i \in DATA_b$, it is denoted as a relation $(d_i, a_j) \in R = DATA_b \times ACT$, where R is a deduction relationship set and defined previously and empirically; $DATA_b$ is the set of "bad" data that can infer to at least one in ACT.

Definition 3. Perfect full privacy (denoted as $\mathsf{Privacy}^{p}_{\mathsf{full}}$). Simply speaking, any adversary $\mathscr{A}$ cannot deduce from anyone in $DATA_u$ to one in ACT after viewing $DATA_u$. More specifically, given anyone $du_i \in DATA_u$, it is impossible for any adversary $\mathscr{A}$ to find $a_j \in ACT$, such that $(du_i, a_j) \in R$. That is,

$$\Pr\left\{\forall du_i \in DATA_u, a_j \Longleftarrow ACT, \text{s.t.} \left(du_i, a_j\right) \in R : DATA_u\right\} = 0, \tag{1}$$

where $\Pr\{A : B\}$ denotes after viewing "B"; the probability of event "A" happens; "$\Leftarrow$" means "is selected from"; "," means two operations happen consequently; "s.t." is a shorthand for "such that."

Definition 4. Computational full privacy (denoted as $\mathsf{Privacy}^{c}_{\mathsf{full}}$). Given anyone $du_i \in DATA_u$, it is computationally infeasible for any Probabilistic Polynomial Turing Machine (PPTM) adversary $\mathscr{A}$ to find $a_j \in ACT$, such that $(du_i, a_j) \in R$. That is,

$$\Pr\left\{\forall du_i \in DATA_u, a_j \Longleftarrow ACT, \text{s.t.}\left(du_i, a_j\right) \in R : DATA_u\right\} < \mathsf{negl}(z), \tag{2}$$

where $\mathsf{negl}(z)$ is a negligible function with security parameter z.

Claim 1. Perfect (computational) full privacy can protect user privacy on all user activities in a day, as no activity can be deduced from data in $DATA_u$ by any (PPTM) adversary.

In previous claim the content in "()" is corresponded with each other. Similarly, the perfect (computational) partial privacy can be defined in the following.

Definition 5. Perfect (computational) partial privacy, denoted as $\mathsf{Privacy}^{p(c)}_{\mathsf{partial}}$. Given at least one $du_i \in DATA_u$, it is computationally infeasible for any (PPTM) adversary $\mathscr{A}$ to find $a_j \in ACT$; such that $(du_i, a_j) \in R$ after viewing $DATA_u$. Besides, given at least one $a_j \in ACT$, it is computationally infeasible for any (PPTM) adversary $\mathscr{A}$ to find $du_i \in DATA_u$, such that $(du_i, a_j) \in R$ after viewing $DATA_u$. That is,

$$\begin{aligned}&\Pr\left\{\exists du_i \in DATA_u, a_j \Longleftarrow ACT, \text{s.t.}\left(du_i, a_j\right) \in R : DATA_u\right\} = 0\,(<\mathsf{negl}(z)),\\ &\Pr\left\{\exists a_j \in ACT, du_i \Longleftarrow DATA_u, \text{s.t.}\left(du_i, a_j\right) \in R : DATA_u\right\} = 0\,(<\mathsf{negl}(z)).\end{aligned} \tag{3}$$

Claim 2. Perfect (computational) partial privacy can protect certain privacy-sensitive activities, as these activities cannot be deduced by $DATA_u$ by any (PPTM) adversary.

Claim 3. Full privacy has stronger strength than partial privacy in terms of the number of deducible data in $DATA_u$. Perfect privacy has stronger strength than computational privacy due to the adversary's ability. That is,

$$\mathsf{Privacy}^{c}_{\mathsf{partical}} < \mathsf{Privacy}^{p}_{\mathsf{partical}} < \mathsf{Privacy}^{c}_{\mathsf{full}} < \mathsf{Privacy}^{p}_{\mathsf{full}}, \tag{4}$$

where "$A < B$" means that the privacy protection strength of "A" is weaker than that of "B".

Roughly speaking, full privacy protects all activities; partial privacy protects partial activities. Perfect privacy defends against any adversary; computational privacy defends against any PPTM adversary. As perfect full privacy has the strongest privacy strength, we thus concentrate on the perfect full privacy protection in the following.

Definition 6. Full privacy attacking experiment on the scheme Π defending against any adversary $\mathscr{A}$-$\mathsf{ExpPrivacy}^{p,\mathscr{A},\Pi}_{\mathsf{full}}$ is defined as follows:

(1) the scheme Π is executed in the presence of any adversary $\mathscr{A}$;

(2) $\mathscr{A}$ fully accesses $DATA_u$, ACT, and R. Given any $du_i \in DATA_u$, if $\mathscr{A}$ can find $a_j \in ACT$, such that $(du_i, a_j) \in R$, $\mathscr{A}$ outputs 1, otherwise, outputs 0;

(3) if and only if $\mathscr{A}$ outputs 1, the experiment outputs 1.

Definition 7. The scheme Π that can guarantee the perfect full privacy in presence of any adversary $\mathscr{A}$ (denoted as $\mathsf{Privacy}^{p,\mathscr{A},\Pi}_{\mathsf{full}} = 1$) is defined as follows.

For any adversary $\mathscr{A}$ that the scheme Π defends against, the probability that the output of the full privacy attacking experiment equals one is 0. That is, if and only if

$$\Pr\left[\mathsf{ExpPrivacy}^{p,\mathscr{A},\Pi}_{\mathsf{full}} = 1\right] = 0, \tag{5}$$

$\mathsf{Privacy}^{p,\mathscr{A},\Pi}_{\mathsf{full}} = 1$.

Therefore, the design goal is to propose a scheme Π satisfying $\mathsf{Privacy}^{p,\mathscr{A},\Pi}_{\mathsf{full}}$ and importantly, with ultra-lightweight computation without any cryptographic computation.

3. Proposed Schemes

3.1. Problem Reduction. To protect the privacy of sensing data $DATA_s$, a naive method is encrypting them at SM and then uploading them to SGCC. As SGCC is untrustworthy, SGCC cannot decrypt them and has to consult a TTP. The TTP decrypts the data, and the result cannot be sent to SGCC. The TTP should compute accumulative values (or metadata) and send them to SGCC for further scheduling and charging. It obviously arises multiple overheads: a large volume of computation overhead at SM; extra communication overhead at SM and SGCC; extra entity TTP; key management overhead between SM and TTP.

As SM is trustworthy, SM is proposed to equip a trusted mixing layer between sensing layer and communication layer. That is, SM is modeled as three tuples: $\langle L_s, L_m, L_c\rangle$, where L_s is a sensing layer computing the power consumption periodically. The output of layer L_s is $DATA_s$; L_m is a mixing layer that transfers $DATA_s$ into $DATA_u$; L_c is a communication layer that uploads $DATA_u$ to SGCC. That is,

$$\begin{aligned}\text{SM} &::= \left\langle L_s, L_m, L_c\right\rangle,\\ L_s &\Longrightarrow L_m : DATA_s,\\ L_m &::= \ F : \ DATA_s \longrightarrow DATA_u,\\ L_m &\Longrightarrow L_u : DATA_u,\end{aligned} \tag{6}$$

where "::=" means "is defined as"; "⇒" means "data transferring between layers"; F is a data transforming function; "$\rightarrow$" that means the input of the function F is transformed into the output of the function F. Therefore, it becomes the

concentration to search an ultra-lightweight transformation function F with $\mathsf{Privacy}_{\mathsf{full}}^{p,\mathscr{A},F} = 1$ in the rest of the paper.

Definition 8. "Bad" data set ($DATA_b$). It consists of all power consumption data that can deduce to one or multiple activities in ACT. $DATA_b = \{db_1, db_2, \ldots, db_o\}$, where o is the total number of $db_i \in DATA_b$ $(i = 1, \ldots, o)$.

The characteristics of $DATA_b$, ACT, and deduction relationship set R are as the following.

(1) Without loss of generality, $DATA_b$ is a sorted set of positive numbers. That is, $db_1 < db_2 < \cdots < db_o$. db_1 is equal to or greater than the power consumption of the minimum power consumption appliance in a period. db_o is equal to or less than the power consumption of all appliances in a period.

(2) Any $db_i \in DATA_b$ $(i = 1, \ldots, o)$ may represent the usage of one appliance in a period. For example, db_1 (30 wh) is the power consumption of a lamp for a period. db_1 is related to an event (e.g., a_1) that means the lamp is on in the period.

(3) Any $db_i \in DATA_b$ $(i = 1, \ldots, o)$ may also represent the usage of multiple household appliances. For example, db_9 represents two household appliances used simultaneously. $db_9 = db_1 + db_2$, where db_1 is the power consumption of the lamp in a period; db_2 is the power consumption of the washing machine in the period. Thus, db_9 means using lamp and washing machine simultaneously in the period.

(4) Similarly, any $a_j \in ACT$ $(j = 1, \ldots, m)$ may represent the usage of one appliance or multiple household appliances simultaneously.

(5) Any $db_i \in DATA_b$ $(i = 1, \ldots, o)$ is related to at least one $a_j \in ACT$, $(j = 1, \ldots, m)$; any $a_j \in ACT$ $(j = 1, \ldots, m)$ is related to at least one in $db_i \in DATA_b$ $(i = 1, \ldots, o)$.

(6) Different $db_i \in DATA_b$ $(i = 1, \ldots, o)$ cannot be related to the same $a_j \in ACT$, $(j = 1, \ldots, m)$, as any $a_j \in ACT$ $(j = 1, \ldots, m)$ has single power consumption in a period.

(7) $db_i \in DATA_b$ $(i = 1, \ldots, o)$ may be related to multiple a_j, because such db_i may be the power consumption for multiple appliances, and those appliances may have the same power consumption in total. For example, $db_9 = db_1 + db_2 = db_3 + db_4$. db_9 is related to $a_5, a_6 \in ACT$, where a_5 means using lamp and washing machine simultaneously and a_6 means the usage of the other two appliances.

In summary, the deduction relationship set $R = DATA_b \times ACT$ can be further refined from a general relationship set to a relationship set with following properties:

$$\begin{aligned}
&\Pr\left\{\forall db_i \in DATA_b, \exists a_j \in ACT, \text{s.t. } \left(db_i, a_j\right) \in R\right\} = 1,\\
&\Pr\left\{\forall a_j \in ACT, \exists db_i \in DATA_b, \text{s.t. } \left(db_i, a_j\right) \in R\right\} = 1,\\
&\Pr\left\{\exists db_i \in DATA_b, a_{j1} \in ACT, a_{j2} \in ACT,\right.\\
&\qquad \left.\text{s.t. } \left(db_i, a_{j1}\right) \in R, \left(db_i, a_{j2}\right) \in R\right\} > 0,\\
&\Pr\left\{\exists d_{i1} \in DATA_b, d_{i2} \in DATA_b, a_j \in ACT,\right.\\
&\qquad \left.\text{s.t. } \left(d_{i1}, a_j\right) \in R, \left(d_{i2}, a_j\right) \in R\right\} = 0.
\end{aligned} \tag{7}$$

In other words, mapping $DATA_b \rightarrow ACT$ is not a function, and mapping $ACT \rightarrow DATA_b$ is a surjective and not a injective function.

Definition 9. After transformation F, the privacy of $DATA_u$ is guaranteed (denoted as $\mathsf{Privacy}_{DATA_u}^{F} = 1$). $\mathsf{Privacy}_{DATA_u}^{F} = 1$, if

$$\forall ds_i \in DATA_s, \qquad du_i = F\left(ds_i\right) \notin DATA_b, \tag{8}$$

where $F : DATA_s \rightarrow DATA_u$.

Definition 10. After transformation F, the soundness of $DATA_u$ is guaranteed ($\mathsf{Soundness}_{DATA_u}^{F} = 1$). The utility summation remains unchanged. That is,

$$\sum_{i=1}^{n} ds_i * p_i = \sum_{i=1}^{n} du_i * p_i. \tag{9}$$

Due to the concentration in the rest of the paper, the research problem is reduced to as follows: given $DATA_s$, find an ultra-lightweight transformation $F : DATA_s \rightarrow DATA_u$, such that the privacy and soundness of $DATA_u$ are both guaranteed. That is, given $DATA_s$, find $F : DATA_s \rightarrow DATA_u$, s.t. $\mathsf{Privacy}_{DATA_u}^{F} = 1$ and $\mathsf{Soundness}_{DATA_u}^{F} = 1$.

Next, we propose a family of schemes to solve the problem. We list all major notations used in the remainder of the paper in Table 1.

3.2. Random Perturbation Scheme (RPS). We firstly propose a basic scheme-random perturbation scheme (RPS) to illustrate our motivations. In RPS, any $ds_i \in DATA_s$ is perturbed into a new value in the middle of db_j and db_{j-1} or in the middle of db_j and db_{j+1}. The two cases are selected randomly. A Random Perturbation Algorithm called RPA is proposed for transformation F as follows.

3.2.1. Analysis of Algorithm 1

Proposition 11. *After the transformation of algorithm RPA, the soundness of $DATA_u$ is guaranteed.* ($\mathsf{Soundness}_{DATA_u}^{RPA} = 1$.)

Table 1: Notation.

$\mathscr{A}$	Adversary
$DATA_s = \{ds_1, \ldots, ds_n\}$	Sensing power consumption data set
$DATA_u = \{du_1, \ldots, du_n\}$	Uploading power consumption data set
F	Transforming function from $DATA_s$ to $DATA_u$
$DATA_b = \{db_1, \ldots, db_o\}$	"Bad" power consumption data set
$PRICE = \{p_1, p_2, \ldots, p_n\}$	Price set
$ACT = \{a_1, a_2, \ldots, a_m\}$	Activity set

Required: $DATA_s$, $DATA_b$
Ensure: $DATA_u$, $\mathsf{Privacy}^{\mathrm{RPA}}_{DATA_u}$ =1, $\mathsf{Soundness}^{\mathrm{RPA}}_{DATA_u} = 1$.
  $ds_i \Leftarrow GetData(DATA_s)$ //Get a data from $DATA_s$.
  for $i = 1$ to $n - 1$ **do**
    if $((ds_i == db_1).\mathrm{OR}.\ (db_1 < ds_i < db_2))$ **then**
      $du_i \Leftarrow (db_1 + db_2)/2$
      $\delta \Leftarrow ds_i - du_i$
      $BIAS \Leftarrow BIAS + \delta * p_i$
    end if
    if $((ds_i == db_o).\mathrm{OR}.\ (db_{o-1} < ds_i < db_o))$ **then**
      $du_i \Leftarrow (db_{o-1} + db_o)/2$
      $\delta \Leftarrow ds_i - du_i$
      $BIAS \Leftarrow BIAS + \delta * p_i$
    end if
    for $j = 2$ to $o - 1$ **do**
      if $(ds_i == db_j)$ **then**
        $du_i \Leftarrow iff(random()\%2, (db_j + db_{j+1})/2, (db_j + db_{j-1})/2)$
        $\delta \Leftarrow ds_i - du_i$
        $BIAS \Leftarrow BIAS + \delta * p_i$
      end if
        if $(db_j < ds_i < db_{j+1})$**then**
          $du_i \Leftarrow (db_j + db_{j+1})/2$
          $\delta \Leftarrow ds_i - du_i$
          $BIAS \Leftarrow BIAS + \delta * p_i$
        end if
    end for
  end for
$du_n \Leftarrow ds_n + BIAS/p_n$ //For soundness

Algorithm 1: Random Perturbation Algorithm—RPA.

Proof. The biases of $du_i\,(i = 1, \ldots, n-1)$ comparing to $ds_i(1, \ldots, n-1)$ are accumulated into a total value *BIAS*. *BIAS* is changed into extra power consumption and added to the last one du_n. Thus, $\sum_{i=1}^{n} ds_i * p_i = \sum_{i=1}^{n} du_i * p_i$. The total cost of power consumption in a day maintains the correct value, so $\mathsf{Soundness}^{\mathrm{RPA}}_{DATA_u} = 1$. □

Proposition 12. *The scheme RPS is ultra-lightweight.*

Proof. As algorithm RPA is ultra-lightweight, the number of loops is $(n-1) * (o-2)$. The computation in each loop is only simple operations such as modulo, minus, plus, division, and multiplication. The computation complexity of algorithm RPA is $O(n * o)$. □

Proposition 13. *The scheme RPS can guarantee the perfect full privacy. (*$\mathsf{Privacy}^{p,\mathscr{A},RPS}_{full} = 1$.*)*

Proof. It is clear that for all $ds_i \in DATA_s, du_i = F(ds_i) \notin DATA_b$. Thus, $\mathsf{Privacy}^{\mathrm{RPA}}_{DATA_u} = 1$. According to the definition of the perfect full privacy, $\mathsf{Privacy}^{p,\mathscr{A},\mathrm{RPS}}_{\mathsf{full}} = 1$. □

3.3. Random Walk Scheme (RWS). If the gap between db_j and db_{j+1} $(j = 1, \ldots, o-1)$ is small, the perturbation (namely, δ) in RPS will be small. It can be proofed as a claim in the following.

Claim 4. If the gap between db_j and db_{j+1} $(j = 1, \ldots, o-1)$ is small, the perturbation in RPS will be small.

Proof. Suppose $\max(|db_j - db_{j+1}|) = g(j = 1, \ldots, n-1)$. If $ds_i = db_j \in DATA_b$) in RPA, $\max(\delta) \leq g/2$. If $ds_i \neq db_j \in DATA_b$, $\max(\delta) < g/2$. Thus, the perturbation δ is small if g is small. □

Required: $DATA_s$, $DATA_b$
Ensure: $DATA_u$, $\mathsf{Privacy}^{\mathrm{RWA}}_{DATA_u} = 1$, $\mathsf{Soundness}^{\mathrm{RWA}}_{DATA_u} = 1$.
  $ds_i \Leftarrow GetData(DATA_s)$
  for $i = 1$ to $n - 1$ **do**
    $j \Leftarrow (random()\%o + 1)$
    $du_i \Leftarrow db_j$
    $\delta \Leftarrow ds_i - du_i$
    $BIAS \Leftarrow BIAS + \delta * p_i$
  end for
  $du_n \Leftarrow ds_n + BIAS/p_n$

Algorithm 2: Random Walk Algorithm (RWA).

If the perturbation is small, adversaries may guess the ds_i correctly, and adversaries can guess the activity is either of two activities. To address this issue, we propose a random walk scheme called RWS in which $ds_i \in DATA_s$ randomly jumps to a value in $DATA_b$. In this case, the privacy definition is extended to include unlinkability, in which the possibility of $db_j \in DATA_b$ for ds_i is equal. Thus, the revealed user activity occurs with equal possibility.

Definition 14. After transformation F, the privacy of $DATA_u$ is guaranteed (denoted as $\mathsf{Privacy}^{F}_{DATA_u} = 1$), if

$$\begin{aligned} &\forall ds_i \in DATA_s, \quad du_i = F\left(ds_i\right) \in DATA_u, \\ &\forall db_p, \quad db_q \in DATA_b, \quad db_p \neq db_q, \\ &\Pr\left\{du_i = db_p\right\} = \Pr\left\{du_i = db_q\right\}. \end{aligned} \tag{10}$$

The definition for privacy is thus extended to include the definition here and Definition 9.

In RWS, any $ds_i \in DATA_s (i = 1, \ldots, n-1)$ is perturbed to a value $db_j \in DATA_b$ $(j = 1, \ldots, o)$, which is randomly selected. This algorithm is thus, especially, ultra-lightweight in terms of computation. A random walk algorithm (RWA) is proposed for the transformation function F as follows.

3.3.1. Analysis of Algorithm 2

Proposition 15. *After the transformation of algorithm RWA, the soundness of $DATA_u$ is guaranteed. ($\mathsf{Soundness}^{RWA}_{DATA_u} = 1$.)*

Proof. The proof is similar to the proof of Proposition 11. As $\sum_{i=1}^{n} ds_i * p_i = \sum_{i=1}^{n} du_i * p_i$, the total cost of power consumption in a day maintains the correct value. Thus, the soundness of RWA is guaranteed. □

Proposition 16. *The scheme RWS is ultra-lightweight.*

Proof. The number of loops is $n - 1$, so algorithm RPA is ultra-lightweight. The computations in loops are only simple operations such as modulo, minus, plus, and multiplication. Moreover, algorithm RWA is more lightweight than algorithm RPA. Thus, scheme RWS is ultra-lightweight. □

Proposition 17. *The scheme RWS can guarantee the perfect full privacy. ($\mathsf{Privacy}^{p,\mathscr{A},RWS}_{full} = 1$.)*

Proof. According to the algorithm, for for all $ds_i \in DATA_s$, if $du_i = F(ds_i) \in DATA_u$, we have $\forall db_p, db_q \in DATA_b$, $db_p \neq db_q$, and $\Pr\{du_i = db_p\} = \Pr\{du_i = db_q\}$. Thus, $\mathsf{Privacy}^{\mathrm{RWA}}_{DATA_u} = 1$ According to the definition of the privacy in Definition 14, $\mathsf{Privacy}^{p,\mathscr{A},\mathrm{RWS}}_{\mathrm{full}} = 1$. □

3.4. Distance-Bounded Random Walk with Perturbation Scheme (DBS).

In smart grid, the uploading data will be used as a feedback for future scheduling of distribution and transmission. It thus requires the uploading data can accurately present the power consumption (namely, sensing data). However, thanks to the power distribution and transmission serve not for a single SM, but a large number of SMs (e.g., a campus, a community, or a county scale), only the accuracy for a scale of SMs is sufficient for scheduling.

In RPS and RWS, although the bias exists (that is, uploading data is not equal to sensing data) at single SM, the uploading data for a large number of SMs can still represent power consumption in a scale. More specifically, the deviation between the summation of uploading data and the summation of sensing data is randomly positive or negative in one SM, thus the overall summation remains almost unchanged in expectation in a large scale. It is explained as follows.

Definition 18. After the transformation F, the accuracy of $DATA_u$ is guaranteed in expectation for a scheduling area (denoted as $\mathsf{Accuracy}^{F}_{DATA_u} = 1$). The summation of $DATA_u$ equals the summation of $DATA_s$, in scheduling area and scheduling period. More specifically, suppose that each scheduling period consists of x sensing (uploading) period and each scheduling area consists of y SMs. The uploading data for them is $\mathrm{SUM}_u = \sum_{t=1}^{y} \mathrm{SM}_t$, $\mathrm{SM}_t = \sum_{i=1}^{x} du_i$. The sensing data for them is $\mathrm{SUM}_s = \sum_{t=1}^{y} \mathrm{SM}_t$, $\mathrm{SM}_t = \sum_{i=1}^{x} ds_i$.

Required: $DATA_s, DATA_b, BOUND$
Ensure: $DATA_u$, $\text{Privacy}_{DATA_u}^{\text{DBA}} = 1$, $\text{Soundness}_{DATA_u}^{\text{DBA}} = 1$
  $ds_i \Leftarrow GetData(DATA_s)$
  for $i = 1$ to $n - 1$ **do**
    WHILE (1)
    {
    $j \Leftarrow (random()\%(o - 2) + 2)$
    $\delta \Leftarrow iff(random()\%2, (ds_i - (db_j + (db_{j+1} - db_j)/2)), (ds_i - (db_j + (db_j - db_{j-1})/2)))$
    if $(abs(\delta) \leq BOUND)$ **then**
      $du_i \Leftarrow ds_i - \delta$
      $BIAS \Leftarrow BIAS + \delta * p_i$
      EXIT;
    else
      CONTINUE;
    end if
    }
  end for
  $du_n \Leftarrow ds_n + BIAS/p_n$

Algorithm 3: Distance-Bounded Algorithm (DBA).

The accuracy of $DATA_u$ is guaranteed, if and only if $\text{SUM}_s = \text{SUM}_u$.

Proposition 19. *After the transformation RPA or RWA, the accuracy of $DATA_u$ is guaranteed in expectation for a scheduling area. ($\text{Accuracy}_{DATA_u}^{RPA\|RWA} = 1$.)*

Proof. In each sensing (uploading) period, ds_i is changed into du_i at single SM. $\delta = ds_i - du_i$. Suppose that each scheduling period consists of x sensing (uploading) period and each scheduling area consists of y SMs. The uploading data for them is $\text{SUM}_u = \sum_{t=1}^{y} \text{SM}_t, \text{SM}_t = \sum_{i=1}^{x} du_i$; the sensing data for them is $\text{SUM}_s = \sum_{t=1}^{y} \text{SM}_t, \text{SM}_t = \sum_{i=1}^{x} ds_i$. The expectation of both is equal, as the expectation of δ is 0 in a scheduling area. That is, $\overline{\text{SUM}_s} = \overline{\text{SUM}_u}$, as $\overline{\delta} = 0$, where $\overline{H}$ means the expectation of H. □

To further guarantee the scheduling accuracy, we propose a distance-bounded scheme, in which the perturbation value (i.e., δ) is bounded. The accuracy is thus guaranteed within a threshold value. It takes the advantages of former two algorithms RPA and RWA. A distance-bounded algorithm (DBA) for the transformation F is proposed as follows.

3.4.1. Analysis of Algorithm 3

Proposition 20. *After the transformation of algorithm DBA, the soundness of $DATA_u$ is guaranteed. ($\text{Soundness}_{DATA_u}^{DBA} = 1$.)*

Proof. The proof is similar to the proof of Propositions 11 and 15. □

Proposition 21. *The scheme DBS is ultra-lightweight.*

Proof. The proof can be reduced to the proof of Propositions 12 and 16. □

Proposition 22. *The scheme DBS can guarantee the perfect full privacy. ($\text{Privacy}_{full}^{p,\alpha,DBS} = 1$.)*

Proof. The proof is similar to the proof of Propositions 13 and 17. □

Proposition 23. *After the transformation DBA, the accuracy of $DATA_u$ is guaranteed in expectation for a scheduling area. ($\text{Accuracy}_{DATA_u}^{DBA} = 1$.)*

Proof. The proof is reduced to the proof of Proposition 19. □

Proposition 24. *The summation of uploading data equals the summation of the sensing data with deviation bounded by $\alpha * \beta * BOUND$, where α is the number of SMs in a schedule area, β is the number of sensing (uploading) period in a schedule period. (That is, $|SUM_u - SUM_s| \leq \alpha * \beta * BOUND$.)*

Proof. The schedule accuracy is the deviation between the summation of uploading data and the summation of sensing data. As it is proofed in Proposition 19, it depends on the number of SMs in the schedule area and the number of sensing (uploading) period in the schedule period. The expectation value is proofed to be 0, as the expectation of δ is 0. Concerning the accuracy of one schedule period, the maximal bias between the summation of uploading data and the summation of sensing data is bounded by $\alpha * \beta * BOUND$. □

4. Related Work

The security architectures and overall security requirements in smart grid were discussed in the recent years [3, 7]. Currently, the privacy issue in smart grid starts to attract more attentions. The requirements of privacy were explored in some previous works [8–11]. They pointed out the importance

and urgency of privacy issues. Efthymiou and Kalogridis proposed a privacy protection scheme via anonymization of data [12]. Their work relied on Escrow and Public Key Infrastructure (PKI); thus the flexibility and scalability may be tampered. Tomosada and Sinohara proposed to use virtual energy demand to estimate the energy load and protecting consumer privacy [13], but the estimation may take much computation overhead, and accuracy may be damaged. Lu et al. [10] proposed an efficient and privacy-preserving aggregation scheme (EPPA). Their scheme relied on homomorphic Paillier cryptosystem and induces much computation overhead. Cheung et al. [14] proposed a credential-based privacy-preserving power request scheme for smart grid, which relied on an advanced cryptographic primitive-blind signature. He et al. [15] proposed to use homomorphic encryption for smart grid communications. Comparing with all aforementioned related work, our final scheme does not rely on any cryptographic primitive but fulfils provable privacy and restrains ultra-lightweight in computation.

5. Conclusions

In this paper, we proposed three schemes to protect user privacy in smart grid without any cryptographic primitive and with ultra-lightweight computation. They are random perturbation scheme (RPS), random walk scheme (RWS), and distance-bounded random walk with perturbation scheme (DBS). We also proposed three algorithms for three schemes, respectively. Our schemes do not rely on any cryptographic computations, are sound in terms of maintaining the correct utility charge, can guarantee the privacy that were strictly proofed, and can ensure the scheduling accuracy in power transmission and distribution. All proposed schemes and algorithms were extensively analyzed, which justified their applicability.

Acknowledgments

W. Ren's research was financially supported by National Natural Science Foundation of China (61170217), the Open Research Fund from Shandong provincial Key Laboratory of Computer Network (SDKLCN-2011-01), and Fundamental Research Funds for the Central Universities (CUG110109). Y. Ren's research was sponsored in part by the "Aim for the Top University Project" of the National Chiao Tung University and the Ministry of Education, Taiwan.

References

[1] The Smart Grid Interoperability Panel Cyber Security Working Group, "Nistir 7628 guidelines for smart grid cyber security," in *Privacy and the smart grid*, vol. 2, 2010, http://csrc.nist.gov/publications/nistir/ir7628/nistir-7628_vol2.pdf.

[2] H. Khurana, M. Hadley, N. Lu, and D. A. Frincke, "Smart-grid security issues," *IEEE Security and Privacy*, vol. 8, no. 1, pp. 81–85, 2010.

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

[4] A. R. Metke and R. L. Ekl, "Security technology for smart grid networks," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 99–107, 2010.

[5] G. N. Ericsson, "Cyber security and power system communicationessential parts of a smart grid infrastructure," *IEEE Transactions on Power Delivery*, vol. 25, no. 3, pp. 1501–1507, 2010.

[6] A. Vaccaro, M. Popov, D. Villacci, and V. Terzija, "An integrated framework for smart microgrids modeling, monitoring, control, communication, and verification," *Proceedings of the IEEE*, vol. 99, no. 1, pp. 119–132, 2010.

[7] T. M. Overman, R. W. Sackman, T. L. Davis, and B. S. Cohen, "High-assurance smart grid: a three-part model for smart grid control systems," *Proceedings of the IEEE*, vol. 99, no. 6, pp. 1046–1062, 2011.

[8] J. Liu, Y. Xiao, S. Li, W. Liang, and C. Chen, "Cyber security and privacy issues in smart grids," *IEEE Communications Surveys Tutorials*, vol. 99, pp. 1–17, 2012.

[9] F. Maandrmol, C. Sorge, O. Ugus, and G. Peandrez, "Do not snoop my habits: preserving privacy in the smart grid," *IEEE Communications Magazine*, vol. 50, no. 5, pp. 166–172, 2012.

[10] R. Lu, X. Liang, X. Li, X. Lin, and X. Shen, "Eppa: an efficient and privacypreserving aggregation scheme for secure smart grid communications," *IEEE Transactions on Parallel and Distributed Systems*, vol. 23, no. 9, pp. 1621–1631, 2012.

[11] S. Wang, L. Cui, J. Que et al., "A randomized response model for privacy preserving smart metering," *IEEE Transactions on Smart Grid*, vol. 3, no. 3, pp. 1317–1324, 2012.

[12] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *Proceedings of the 1st IEEE International Conference on Smart Grid Communications (SmartGridComm '10)*, pp. 238–243, October 2010.

[13] M. Tomosada and Y. Sinohara, "Virtual energy demand data: estimating energy load and protecting consumers' privacy," in *Proceedings of the IEEE PES Innovative Smart Grid Technologies (ISGT '11)*, pp. 1–8, January 2011.

[14] J. Cheung, T. Chim, S. Yiu, L. Hui, and V. Li, "Credential-based privacy-preserving power request scheme for smart grid network," in *Proceedings of the IEEE Global Telecommunications Conference (GLOBECOM '11)*, pp. 1–5, December 2011.

[15] X. He, M. Pun, and C. Kuo, "Secure and efficient cryptosystem for smart grid using homomorphic encryption," in *Proceedings of the IEEE PES Innovative Smart Grid Technologies (ISGT '12)*, pp. 1–8, January 2012.