Consumer Privacy on Distributed Energy Markets

Niklas Büscher[1(✉)], Stefan Schiffner[2], and Mathias Fischer[3]

[1] Technische Universität Darmstadt, Darmstadt, Germany
buescher@seceng.informatik.tu-darmstadt.de
[2] ENISA, Athens, Greece
[3] Westfälische Wilhelms-Universität Münster, Münster, Germany

Abstract. Recently, several privacy-enhancing technologies for smart grids have been proposed. However, most of these solutions presume the cooperation of all smart grid participants. Hence, the privacy protection of consumers depends on the willingness of the suppliers to deploy privacy-enhancing technologies. Since electrical energy is essential for our modern life, it is impossible for consumers to opt out. We propose a novel consumer-only (do-it-yourself) privacy-enhancing approach under the assumption that users can obtain their energy from multiple suppliers on a distributed market. By splitting the demand over multiple suppliers, the information each of them can collect about a single consumer is reduced. In this context, we suggest two different buying strategies: a time and a sample diversification strategy. To measure their provided level of privacy protection, we introduce a new indistinguishability metric λ-Indistinguishability (λ-IND) that measures how relative consumption changes can be hidden in the total consumption. We evaluate the presented strategies with λ-IND and derive first privacy boundaries. The evaluation of our buying strategies on real-world energy data sets indicates their ability to hide load profiles of privacy sensitive appliances at low communication and computational overhead.

1 Introduction

Currently, users of the electrical grid are facing the risk of privacy breaches through the upcoming smart grid technology. The idea of the *smart* grid is to modernize the traditional electricity grid by establishing a communication infrastructure in parallel to the energy delivery network. This results in a constant flow of fine-grained consumption information from individual consumers to the energy suppliers. Furthermore, this data enables automatic billing, prediction and stabilizing tasks for suppliers. However, as research has shown, this data can also be used to infer detailed user profiles. Even further, Non-Intrusive Load Monitoring (NILM), the technique to disaggregate energy consumption, is still developing. Recent progress shows that given high resolution load profiles, content displayed on a larger LCD can be identified [12] as well as rendered web pages [5]. Thus, reporting the consumption information is bearing a risk for the

S. Schiffner et al. (Eds.): APF 2016, LNCS 9857, pp. 96–114, 2016.
DOI: 10.1007/978-3-319-44760-5_7

individuals privacy. This is especially the case in a scenario where 'opt out' is not an option, as is the participation in the electricity grid.

Previous presented solutions, which fulfill the suppliers' functional requirements and protect the privacy of users, depend on either the electricity suppliers voluntary commitment to complex cryptographic protocols or on the deployment of physical batteries. Cryptographic protocols are challenging in the correct implementation and require the willingness of the supplier to invest in the necessary hardware and software to run these protocols. Physical batteries require a huge investment in batteries for the consumer. From the individual's point of view, it would be preferable to be protected with less supplier dependency and without costly investments.

Based on these observations, we present a novel privacy enhancing approach that enables the clients to protect their consumption data without the need of involving suppliers. We discuss our solution in the context of smart grids, though it can be generalized for privacy protection on distributed markets. Our main contributions can be structured according to the following two research questions:

How can the consumer's privacy on distributed markets be protected without the technical involvement of suppliers? We answer this question by presenting a novel data perturbation based approach. The idea is to utilize the distributed market by randomly splitting the consumer's demand onto multiple suppliers. Thus, only a fraction of the total demand is observed by each supplier. This approach does not presume any further technical requirements while still guaranteeing accurate trades.

To which degree can privacy be protected and how can this protection be measured? On distributed markets, multiple parties usually trade a good directly and hence need to have knowledge of each other, which turns privacy definitions based on anonymity inapplicable. Furthermore, we show that that our buying strategies can hide only relative changes in the power consumption. As a consequence, the prerequisites of differential privacy or plain indistinguishability are too demanding. Therefore, we introduce a new privacy notion that uses strong formal guarantees to measure the protection of relative changes in the power consumption.

The paper is structured as follows. We discuss the related work in Sect. 2, before introducing our formal model and privacy metric in Sect. 3. Moreover, in Sect. 4 two novel buying strategies are presented and analyzed. Then, the strategies are evaluated on real world data sets in Sect. 5. Finally, we conclude our work in Sect. 6.

2 Related Work

In this section we discuss the state of the art in privacy protection mechanism for the smart grid. Furthermore, we discuss relevant statistical privacy metrics used to measure privacy in the smart grid.

Privacy Mechanisms. According Jawurek et al. [14] Privacy-enhancing technologies (PETs) for the smart grid can be classified into the following categories:

Data perturbation based protection mechanisms enable privacy friendly live monitoring by adding random noise to every raw reading, e.g., Bohli et al. [4] and Shuang et al. [26]. Hence, the actual reported readings are noisy. However, given a sufficiently large number of smart meters, the noise cancels out and thus, the supplier's aggregate becomes accurate. More sophisticated approaches for data perturbation are presented by Acs and Castelluccia [1] and Lin et al. [18] that combine data perturbation and additive blinding. All of these approaches either require a second protocol to allow accurate billing, an infeasible large number of smart meters, or an implementation of the encryption protocol at supplier side.

Batteries can reduce the entropy of the readings by flattening the actual electricity consumption [2,15,24]. Depending on the capacity and throughput of the battery, different privacy goals can be realized. However, it turns out that adequately sized batteries are expensive.

Furthermore, Trusted-Third-Parties (TTPs) have been utilized as PETs in smart grids, e.g., [4,10]. While TTPs can fulfill any privacy definition, they bear two risks: first, any trusted third party can also be compromised and represents a single point of failure; and second, deploying a TTP requires infrastructure and protocol changes at smart meters and suppliers.

One of the most promising solutions are aggregation protocols, which enable accurate live monitoring. Based on various cryptographic primitives, multiple variants have been presented. For example, Garcia et al. [11] and Kursawe et al. [17] presented protocols using either additive secret sharing or homomorphic cryptosystems. These protocols guarantee anonymity on a group level. However, they all make use of expensive computation or require bidirectional communication between groups of smart meters. Moreover, the proposed protocols have an inherent complexity and need to be implemented on the supplier side. Hence, they disqualify as consumer-only approaches.

Lastly, commitment schemes and zero knowledge proofs have been proposed to offload the bill calculation onto the consumers [7,20,23]. Here verifiable computation guarantees the correct calculation of the overall bill without revealing individual readings. This approach requires a protocol implementation on the supplier side and is incompatible to live monitoring, as only the smart meters sum is computed and verified.

Privacy Metrics. We focus on privacy metrics for smart grids that measure the protection level of approaches based on data perturbation.

Quantitative metrics based on statistical and information theoretic measures have been presented. Shuang et al. [26] use the F-Test measure to compare raw and noisy load profiles. Kalogridis et al. [15] measured this relationship using relative entropy and correlation metrics. Furthermore, the authors suggest to use the accuracy of clustering algorithms as a privacy measures. All these metrics are useful when comparing different privacy mechanisms. However, they have the drawback that a measurable threshold for a desired privacy level cannot be given.

To evaluate a battery based approach, Backes et al. [2] developed a metric based on differential privacy for streams, which ensures event-level privacy. The authors make use of the probabilistic variant of differential privacy, i.e., with a small probability δ the definition of differential privacy does not need to be met. Even so, this metric is based on the well defined grounds of differential privacy, it suffers practicability, as the authors note. This is because load signatures of appliances are typically characterized by more than one event, which are not necessarily covered by the presented definition.

Yet, a metric that shows the protection of multiple events is desirable. Bohli et al. present such a metric in [4], which is based on a cryptographic game of the type right-or-left. In this game, an adversary is challenged to identify the originating scenario from a transcript. A scenario consists of load profiles, i.e., load samples in a defined time span from multiple smart meters. We build on this idea in the reminder of this paper.

3 A Formal Smart Grid Model

In this section, we present the *distributed market model*. First, we define the major actors and actions. Then, we introduce the attacker model and the notion of λ-IND.

3.1 Energy Market Model

We define a distributed market as a virtual place where consumers purchase goods or services from multiple suppliers. In this paper, we focus on a single good market, i.e., the energy market. Nevertheless, for markets that offer multiple goods, the presented ideas can be applied multiple times in parallel. We assume that all communication is secured, i.e., communication channels are available all the time and guarantee confidentiality as well as integrity. Hence, a trade is not visible to any third party. The practical realization of such a market place requires supplier discovery and price formation services, which is beyond the scope of this work.

We deduce a formal model and its assumptions: the model consists of two participating parties, namely a set of consumers C and a set of suppliers S. Moreover, a discrete notion of time, denoted as t, is used. In each time period t, a consumer $c_i \in C$ is attributed with a demand $d_{i,t}$. Consumers can cover their demand by buying from one or multiple suppliers $s_j \in S$. As we are only interested in modelling consumption privacy instead of anonymity, it is sufficient to consider only one single consumer $c \in C$ in all following discussions.

The act of a consumer to buy a certain amount of energy in a given time period from a supplier is called trade. All trades of one consumer are denoted by a two dimensional matrix. Each entry $b_{j,t} \geq 0$ of this matrix describes the amount of the good bought by a consumer from supplier s_j at time t. Consequently, the demand at time t of the consumer is the sum of all trades with all suppliers $d_t = \sum_{s_j \in S} b_{j,t}$. In the following privacy analysis, we refer to the time series

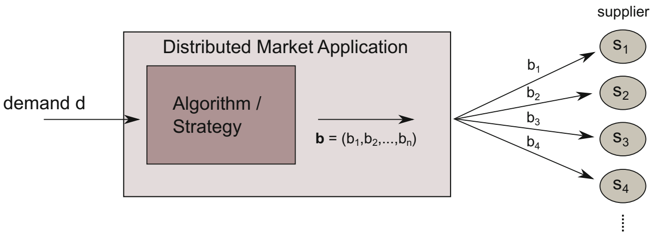


Fig. 1. Distributed market model. The consumer's demand d is split between multiple suppliers $s_1, s_2, \ldots, s_n$.

of a consumer $\mathbf{d} =< d_1, d_2, \ldots, d_n >$ as *original* load profile and for the time series that a supplier observes $\mathbf{b}_j =< b_{j,1}, b_{j,2}, \ldots, b_{j,n} >$ as *reported* load profile. Figure 1 illustrates the distribution model for a given demand d.

A consumer that cannot produce or store energy needs to cover all its demand via the market. Hence, its entire demand profile is at risk to be leaked. Contrary to consumers, so-called *prosumers* exist, who are capable of producing and storing energy to a certain extent, e.g., via a solar panel and an additional battery. Thus, by partially covering their demands through (unpredictable) third sources, they have more possibilities to protect their load profiles. We note, that given the possibility to report arbitrary and negative trades, two non-colluding suppliers are sufficient to trivially guarantee information theoretic security, by reporting $b_{1,t} = r_t$ to the first supplier with r_t being a random number and $b_{2,t} = d_t - r_t$ to the second supplier. Such a protocol guarantees correctness and privacy but is incompatible with time-of-use tariffs and practical live monitoring.

For the remainder of this paper we will focus on consumers only, as they are the more challenging case for privacy-protection. Therefore, to restrict our analysis adequately to the capabilities of consumers, we define all trades to be non-negative $b_{j,t} \geq 0$.

3.2 Attacker Model

Assuming a secure communication network, the only possible point to attack is at the end-users, namely compromising a supplier. Furthermore, we assume that the attacker is interested in reconstructing the original load profiles of consumers from reported consumption information. As this kind of attacker is completely passive, consumers are unable to differentiate between honest and compromised suppliers. Moreover, as a first step we assume that only one supplier is compromised, which is sufficient to show the impact of the considered attacker on distributed energy markets.

3.3 Privacy Metric - λ Indistinguishability

We define a Load Signature Hiding Game (LSHG) based on Bohli et al. [4] to measure the privacy of a Smart Metering Application (SMA). An adversary A

selects two possible load profiles, namely the vectors $\mathbf{d}^0 =< d_1^0, d_2^0, ..., d_n^0 >$ and $\mathbf{d}^1 =< d_1^1, d_2^1, ..., d_n^1 >$, and sends them to a challenger. After receiving the two scenarios the challenger randomly draws a bit $\beta \in \{0, 1\}$ and simulates $\mathbf{d}^\beta$. The simulation result is a transcript, which is then sent back to the adversary A. Following the described market scenario, the transcript consists of all trades with one randomly chosen supplier: $\mathbf{b} =< b_1, b_2, ..., b_n >$. The adversary outputs a bit δ and wins the game by correctly guessing which scenario was used to create the transcript, hence, iff $\delta = \beta$. The privacy of the SMA is measured by the difference between random guessing and correctly answering which of the two scenarios belongs to the transcript. As in [4] the two demand load profiles are required to have the same aggregate, since this information has to be known by the supplier for billing purposes. Otherwise, distinguishing load profiles is trivial.

To measure the privacy protection provided by the buying strategies introduced later in this paper, we present the idea of λ-IND. Even though deviating from common privacy metrics is bearing risks, we propose a new privacy metric and advocate the notions of indistinguishably, due to the following reasons:

- As discussed in Sect. 2, other common privacy metrics are either inapplicable, e.g., anonymity metrics, or provide insufficient protection in this scenario. For example, differential privacy under continual observation [9] only provides event-level protection that does not span over multiple events.
- Cryptographic games provide a strong formal tool and have successfully been applied as privacy metrics in different scenarios, e.g., for privacy preserving RFID tags [25].
- The strict indistinguishability notion for the smart grid by Bohli et al. [4] assumes a very strong adversary, who is allowed to choose arbitrary load signatures. This definition is too strong to show that only a part of the load profile is protected.

The goal of λ-IND is to show that high resolution attacks are infeasible. We are convinced that this is an important stepping stone between none and full protection, i.e., perfect indistinguishability in the LSHG. Our idea is as follows, instead of challenging the privacy mechanism with two totally different load profiles, the load samples from the same time period are restricted to be in relative distance to each other.

Formalizing this concept, we introduce the privacy parameter λ that expresses the maximal relative difference between two load samples taken from two load profiles in the privacy game, respectively. The new privacy metric λ-IND is based on the definition of the LSHG with the exception that two load samples in both scenarios are allowed to differ by at most a factor λ. Hence, given a load sample d_t^0 in the first scenario, the demand in the second scenario is restricted to $d_t^1 \in [d_t^0, \lambda \cdot d_t^0]$. Without loosing generality, $\lambda > 1$ is assumed for all further discussion. We refer to this restriction as the λ *requirement* and introduce the following definition:

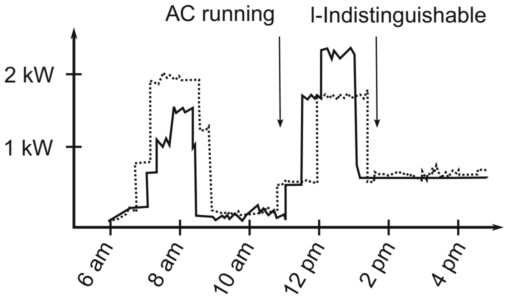


Fig. 2. Applicability of λ-IND for exemplary load profiles from two different households. At approximately 11 am both households switch on their AC, which leads to a similar power consumption. After a period of high energy usage in both households with different duration, no further activity in the household represented by the solid line is visible, whereas in the second household a TV is turned on. Thus, after 1 pm only a small λ is sufficient to show the running TV is hidden with λ-IND.

Definition 1. *Considering the LSHG(β) game fulfilling the λ requirement and the adversary A for a given distribution algorithm* ***Alg****, the (λ-Indistinguishability) advantage of A is defined as*

$$\mathbf{Adv}^{\lambda\text{-IND}}_{Alg} = |\Pr[\text{LSHG}(0)^{A}_{\text{Alg},\lambda} = 0] - \Pr[\text{LSHG}(1)^{A}_{\text{Alg},\lambda} = 0]|.$$

We illustrate λ-IND with an example. Given $\lambda = 1.2$ and load samples of 1000 Wh in the first profile, the maximum load sample an attacker can choose in the second profile is 1200 Wh. Thus, the chosen loads for the second profile have to be in between the corridor from 1000 Wh to 1200 Wh. Consequently, given a base load of 1000 Wh, an additional appliance with a load signature of maximal 200 Wh is undetectable. This concept is also illustrated in Fig. 2. Summarizing, a privacy mechanism guaranteeing λ-IND makes all load samples indistinguishable that are in relative distance to each other.

4 Buying Strategies

In this section, we introduce and evaluate multiple buying strategies in our formalized distributed market scenario. First, we introduce the notion of *fair* buying strategies, i.e., strategies where no supplier is favored. Second, we introduce the Temporal Diversification (TD) and Sample Diversification (SD) buying strategies and evaluate both strategies in the (unrestricted) LSHG model as well as under λ-IND.

A buying strategy is an algorithm that distributes an input demand d_t among multiple suppliers $s \in S$ in every time period. Thus, each supplier s_j observes a reported load profile of load samples $\mathbf{b}_j = < b_{1,j}, b_{2,j}, \ldots, b_{n,j} >$. By observing these load samples over a larger time period and assuming a steady input demand d, each approached supplier s_j observes an distribution $P_j(b)$ of load samples. We focus on *fair* buying strategies, i.e., buying strategies that do not favor any supplier over time. Thus, we propose the following formal definition for fair distribution algorithms:

Definition 2. *A distribution algorithm* ***Alg*** *with input load sample d_t and output vector consisting of $|S|$ load samples $b'_t =< b_{t,1}, b_{t,2}, ..., b_{t,|S|} >$ is called fair, iff for all $x \in [0, d_t]$ the following condition holds:*

$$\Pr[b_{t,1} = x] = \Pr[b_{t,2} = x] = ... = \Pr[b_{t,|S|} = x].$$

Note, even though unfair strategies might be interesting for the consumer, e.g., because of economic or ecological preferences, distribution algorithms that favor certain suppliers have the drawback that an attacker may obtain information on these preferences. This background information might undermine the consumers privacy. Hence, in the light of privacy protection, we recommend *fair* distribution algorithms. Among such fair algorithms are the TD and SD strategies that are introduced in the following two subsections.

4.1 Buying Strategy - Temporal Diversification (TD)

Consumers that cover their demand according to the TD strategy, have to meet their demand d_t per time period t through only one, yet changing supplier. Several variants of this strategy are possible w.r.t. the order (deterministic or stochastic) suppliers are approached.

An example for a deterministic variant is to use a round-robin scheme, i.e., suppliers are approached subsequently in an ordered sequence. Once the last supplier in the sequence is reached, the process starts with the first supplier again. In the second variant, suppliers are randomly chosen from the set of available suppliers. Several variations of such a random strategy are possible, e.g., the same supplier can be approached for k subsequent time periods. Hence, depending on the consumer's goals the granularity of the observed time frame can be controlled by parameter k.

Round-robin and random TD strategies can only offer limited privacy, as long as the number of suppliers is limited. This is because, consumers will inevitably return to the same supplier at some point. However, these strategies reduce the temporal resolution of a compromised attacker. For the indistinguishability analysis we apply LSHG and λ-IND on a randomized TD strategy and leave out the round-robin variant due to its static and predictable results. These are that each supplier is approached after at most $|S|$ time periods. An analysis of the TD strategy prepares the evaluation of the more complex SD strategy in the LSHG and λ-IND.

To analyze strategies with the help of cryptographic games, the attacker needs to construct two scenarios for the challenger. With respect to the TD strategy, it turns out that any two non-equal demand profiles are distinguishable, by setting one half of the first demand profile to an arbitrary $d^0 > 0$ and the other half to $d^1 \neq d^0$, $d^1 > 0$. The requirement for non-zero loads $d^0, d^1 > 0$ is necessary for the adversary to distinguish between zero consumption and not being approached at all. The second demand profile is constructed by swapping d^0's and d^1's. As a result of this construction, the sum of all load values is the same in both scenarios, as required. The adversaries advantage is then equivalent to the probability to observe a non-zero load sample:

$$\mathbf{Adv}_{\mathrm{TD}}^{\mathrm{LSHG}} = 1 - \left(\frac{|S|-1}{|S|}\right)^n .$$

We further observe that the attacker advantage in the LSHG is equal to the advantage in λ-IND, $\mathbf{Adv}_{\mathrm{TD}}^{\lambda\text{-IND}} = \mathbf{Adv}_{\mathrm{TD}}^{\mathrm{LSHG}}$. This is because, the λ-requirement does not prevent the adversary from choosing load samples that uniquely identify a load profile.

4.2 Buying Strategy - Sample Diversification (SD)

Consumers that deploy the Sample Diversification (SD) strategy cover their demand by using multiple suppliers simultaneously. A randomized algorithm splits the input demand into multiple smaller samples that are sent out to different suppliers. Hence, each supplier only observes a share of the total demand.

For example, given $|S| = 3$ suppliers and a demand of $d_t = 1000\,\mathrm{Wh}$. A consumer deploying a SD strategy could meet its demand by buying $b_{1,t} = 511\,\mathrm{Wh}$ from the first supplier, $b_{2,t} = 89\,\mathrm{Wh}$ and $b_{3,t} = 400\,\mathrm{Wh}$ from the second and third supplier. Several variations of this strategy are possible and can be differentiated by their distribution of load samples, e.g., exponential or uniform. Below we present an approach to derive the upper bound of the adversaries advantage for any SD variant.

Upper Bound for the Adversaries Advantage in the LSHG Game. We analyze the SD strategy in LSHG. For this the adversary needs to choose two load profiles that show the largest difference to maximize its advantage. However, a binary difference, namely zero and non-zero load is already sufficient, as we show. Thus, in the first load profile one half of the load samples is set to zero and the other half to a value greater than zero, e.g., one. The second scenario is constructed by swapping zeros and ones ensuring equal demands in both scenarios. Since any randomized reported consumption $b_{j,t}$ to supplier s_j is bounded by zero and the actual demand, i.e., $0 \leq b_{j,t} \leq d_t$, a smart meter has to report zero consumption in times of zero demand and non zero consumption to one or more suppliers in times of demand. In the following calculation, we denote the number n_s, as the number of suppliers being approached in every time period. Receiving a load sample greater than zero allows the challenger to deduce the simulated scenario, namely the one where the sample in the load sequence $\mathbf{b}_j$ is greater than zero. For simplicity reasons we assume an even number of load samples per profile. Since $\frac{n}{2}$ loads per scenario are greater than zero, the adversaries advantage is bound by the probability to observe such a non-zero load:

$$\mathbf{Adv}_{\mathrm{SD}}^{\mathrm{LSHG}} = 1 - \left(\frac{|S| - n_s}{|S|}\right)^{n/2}.$$

Upper Bound for the Adversaries Advantage under $\boldsymbol{\lambda}$-IND. To derive an upper bound on the adversaries advantage under λ-IND, we first have to describe an optimal adversary. According to the Neyman-Pearson Lemma [21], the best possible advantage when distinguishing distributions is achieved when using a *maximum likelihood-ratio distinguisher*. Given such an optimal distinguisher, its advantage is equal to the statistical distance, also known as total variation distance D_{TV}. The statistical distance between two discrete[1] probability functions P_0, P_1 for a given sample x is defined as

$$D_{TV}(P_0, P_1) = \frac{1}{2}\|P_0(x) - P_1(x)\|_1 = \frac{1}{2}\sum_{x \in \Omega} |P_0(x) - P_1(x)|\, dx.$$

The singular case can be extended to multiple samples by computing the 1-norm over all possible combinations [3]. As this can be computationally expensive, Pinsker's inequality [6,22] can be used to compute an upper bound on the distinguishing advantage more efficiently. Pinsker's inequality connects the statistical distance D_{TV} with the Kullback-Leibler divergence D_{KL} and is defined for multi-samples n as

$$D_{TV}(P_0^n, P_1^n) = \frac{1}{2}\|P_0^n(x) - P_1^n(x)\|_1 \leq \sqrt{2n \cdot D_{KL}(P_0\|P_1)}.$$

To minimize the adversaries advantage in λ-IND, an optimal strategy has to distribute demands d_0 and d_1, which differ by at most λ, in such a way that the distributions of observed load samples show minimal statistical distance. First, we consider the case where a load profile consists of only one demand ($n = 1$). The least statistical distance is achieved when the transport between the two distributions observed by the adversary in the LSHG is minimized. As the two distribution P_0 and P_1 have to differ, because the originate different input demands, the best possible way to construct distribution P_1 from a given P_0 is realized by transporting probability from the two extremes 0 and $\max(d_0, d_1)$. This minimizes the amount of transported probability and thus, the statistical distance. Given a number of available suppliers $|S|$ and the privacy parameter λ, the statistical distance is then bound to (cf. Appendix A):

$$\mathbf{Adv}_{\mathrm{SD,1}}^{\lambda\text{-IND}} = \frac{\lambda - 1}{|S| \cdot \lambda}.$$

Following the same strategy for load profiles consisting of multiple samples ($n > 1$), a maximum likelihood-ratio distinguisher can only decide according the transported probabilities and has thus an advantage of at most

$$\mathbf{Adv}_{\mathrm{SD,n}}^{\lambda\text{-IND}} = 1 - \left(1 - \frac{\lambda - 1}{|S| \cdot \lambda}\right)^{n}.$$

[1] For simplification purposes, in this work we make use of discrete instead of continuous probability distributions. This is reasonable when considering a finite metering resolution (e.g., $10^{(-7)}$ kWh).

4.3 Heuristics for the SD Strategy

A distribution strategy as presented above cannot directly be deployed in practical settings. This is because, in a real world deployments of a smart meter all values from zero to a households maximum consumption will be observed at some point. Thus, the static assignment with minimal transport from and to a single value has to be replaced by a continuous approach. Furthermore, the values for upcoming d_t are unknown to the distribution algorithm, and therefore a *proportional* distribution scheme is desirable. Thus, the fraction of the demand observed by an individual supplier is independent of the input demand. Moreover, a heuristic should function with little computational cost to avoid expensive smart meter hardware. Finally, a practical heuristic should reduce the communication costs and should only report noticeable consumption. Thus, tiny load samples could be grouped and sent out to only one supplier. However, this variation impacts the privacy and is evaluated further in Sect. 5.

We present an efficient heuristic that considers the afore-mentioned thoughts. It is uses the idea that the probability transport is kept minimal and that the distribution should become uneven towards the extremes. Moreover, as a variant, all samples below a threshold τ can aggregated and grouped together to avoid the communication of arbitrarily small samples. The core idea of the heuristic, as presented in Algorithm 1, is to iteratively draw the reported samples according a uniform distribution over the remaining demand:

Algorithm 1. Communication Optimized Distribution Algorithm

```
input $d, |S|, \tau$
$b_2 \leftarrow \cdots \leftarrow b_{|S|} \leftarrow 0$
$b_1 \leftarrow$ rand()                              ▷ First load is drawn uniformly from $[0, d]$
$l \leftarrow 1 - b_1$                                ▷ Remaining load
for $i = 2, \ldots, |S| - 1 \land l > 0$ do
    $b_i \leftarrow$ rand() $\cdot\, l$
    $l \leftarrow l - b_i$
    if $l < \tau$ then                    ▷ Threshold variant: Compare with threshold
        $b_i \leftarrow b_i + l$                      ▷ Aggregate the rest
        $l \leftarrow 0$
    end if
end for
$b_{|S|} \leftarrow l$
$\mathbf{b} \leftarrow$ shuffle($< b_1, b_2, \ldots, b_{|S|} >$)    ▷ Shuffle for fair distribution
$\mathbf{b} \leftarrow d \cdot \mathbf{b}$
output $\mathbf{b}$
```

The algorithm takes a demand d, a number of suppliers $|S|$ and (optional) a threshold τ as input and outputs a vector of load samples, whose sum is the given input demand. In a first step the interval $[0, 1]$ is split into two parts according to a value b_1 drawn uniformly from the same interval. The left part of the interval becomes the first reported load. The remaining load $l_1 = d - b_1$ is further split by a random value b_2 drawn from the uniform distribution on the interval $[0, l_1]$. In the further steps the remaining load is updated, $l_2 = l_1 - b_2$. This iterative procedure continues for all suppliers or until the remaining load reaches the threshold (if given). In both cases, the last reported load $b_{|S|}$ is set to $l_{|S|-1}$ to distribute the remaining load. As a result, earlier drawn b_j are more likely to be larger than those which have been drawn at the end of the recursive procedure. To achieve a *fair* distribution for all suppliers, in its final steps the algorithm performs a random permutation (shuffle) on $b_1, \ldots, b_{|S|}$ and multiplies each fraction b_j with the total demand.

5 Evaluation

We discuss the applicability of our results in an evaluation on real world data sets in this section. First, we identify a reasonable value for the privacy parameter λ. Then, we study the influence of different parameter choices, e.g., the number of suppliers, on the adversaries advantage against the SD strategy.

5.1 Privacy Sensitive Appliances

To show that λ-IND has practical relevance, we identify appliances that in our opinion show the highest privacy risk. In a second step, we evaluate their energy consumption in comparison with the total consumption. The latter give us an insight on a reasonable choice for λ.

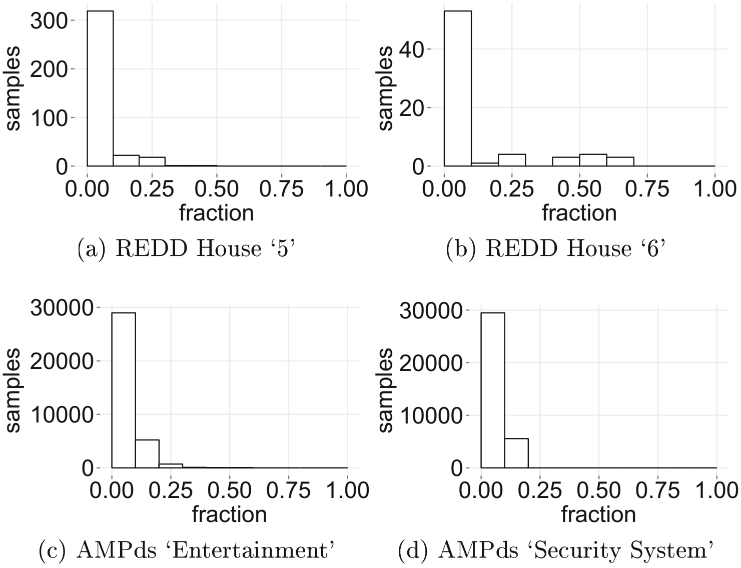
One group of privacy sensitive appliances are digital screens. Recently Greveler et al. [12] showed that the TV program can be identified in the aggregate power consumption. Moreover, Clark et al. [5] showed an attack, where rendered websites could be identified through power analysis. Since LC-Displays also display private information, we are convinced that digital devices need special protection. Similar concerns have been raised by Backes et al. [2]. Another example of noteworthy appliances are alarm systems. A remote detection of their functionality can compromises the households inhabitants safety [13].

We evaluate the energy consumption of the mentioned appliances on two larger public data sets that are used in NILM research:

The *Reference Energy Disaggregation Data Set (REDD)* was published by J. Zico Kolter and Matthew J. Johnson [16]. It contains fine granulated energy data collected from six houses around Boston, Massachusetts. Kolter et al. measured not only the total consumption but also monitored multiple labeled sub-circuits within the households. The dataset consists of low (1 Hz) and high frequency (15 kHz) measurements.

The *Almanac of Minutely Power data set (AMPds)* was released by Stephen Makonin et al. [19]. The AMPds provides one year of data from a single household from the Vancouver region in British Columbia. Similar to the *REDD* data set, the *AMPds* provides readings of 21 sub-metered circuits with a frequency of one reading per minute.

For our evaluation we used the statistics programming language *R*. First, all incomplete and implausible entries are removed from the data sets, e.g., entries where sub-metered circuits are not measured or the power consumption of appliances exceeds the total consumption. Second, all load samples are aggregated in 15 min intervals. Third, all time periods with zero consumption of sensitive devices are removed. Finally, a histogram is created over the fraction of energy used by the sensitive devices.



(a) REDD House ‘5’

(b) REDD House ‘6’

(c) AMPds ‘Entertainment’

(d) AMPds ‘Security System’

Fig. 3. Fraction of energy spent on electronic devices for two houses in the REDD and the energy spent on entertainment and security system in the AMPds.

Figure 3a and b show the results for the *REDD* for two distinct households, which have a sub-metered circuits labeled electronics. The histograms illustrate the number of time periods in which the fraction energy consumption of entertainment appliances is within the range printed on the x-axis. Figure 3c illustrates the fraction of energy used entertainment appliances in the *AMPds* and Fig. 3d illustrates the same for the alarm system. Taking these numbers into account, in more than 80 % of all time periods the measured fraction is below or equal 10 %. Furthermore, with the exception of ‘house 6’, in more than 95 % of all time periods, the sensitive appliances consume less than 20 % of the total energy. The alarm systems always require less than 20 % of the total energy consumption. Unfortunately, no breakdown of the sub-metered circuits is given. Thus, the actual consumption of a individual sensitive appliances could be even less. The results support the idea that λ-IND with small λ, e.g., $\lambda = 1.2$, is of practical use to measure the protection of privacy sensitive appliances.

5.2 λ-IND Evaluation of the SD Strategy

In Sect. 4, we have introduced the SD strategy and have proposed a theoretical distribution strategy as well as heuristics. In this section, we evaluate both with different parameters under λ-IND. Thus, the relationship between the λ, the number of suppliers, and the adversaries advantage is studied.

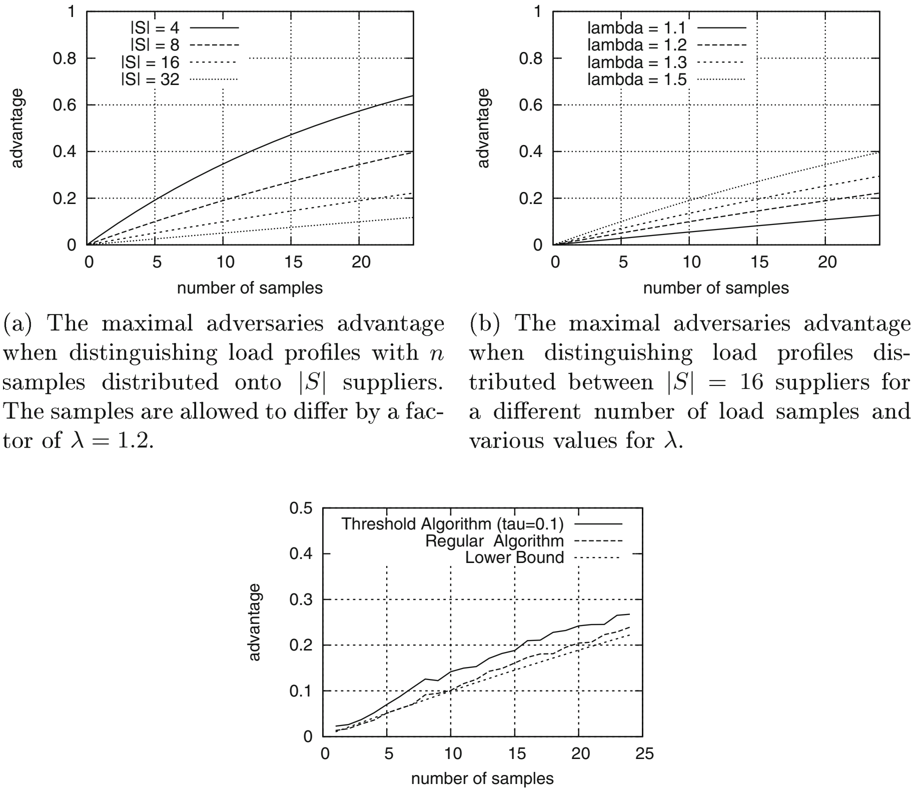
In Sect. 4 a formula for computing an upper bound on the adversaries advantage for given distribution is presented. The described heuristics, however, require a further investigation, as the resulting distribution are not described in closed-form. Therefore, to evaluate these we follow a numerical Monte Carlo approach. First, we distribute a constant demand onto $|S|$ suppliers by applying the heuristics. Repeating this experiment $k = 10^7$ times, a probability distribution of load samples is observed. Given this distribution, an optimal likelihood-ratio distinguisher is used to calculate the adversaries advantage under λ-IND. The heuristic and the evaluation itself are written and executed in R.

The upper bound on the advantage of the adversary as computed in Sect. 4 depending on the number of samples for a different number of suppliers is illustrated in Fig. 4a. The parameter λ is fixed to 1.2 and we observe that, as expected, an increasing number of suppliers decreases the adversaries advantage. Figure 4b shows the distinguishing advantage in dependence on the number of samples for different choices of λ using a fixed number of suppliers $|S| = 16$. When increasing λ, the maximal advantage of the attacker also increases. Thus, the consumer faces the trade-off between the protected time span and the level of protection, i.e., the maximal fraction of energy that can be protected. However, we observe that the advantage is never negligible. Moreover, as others have already discussed [8], the question which advantage is acceptable is of social concern and not of technical interest.

The results of the numerical evaluation of the heuristic described in Algorithm 1 are presented in Fig. 4. The advantage of the heuristics with/out threshold are compared with the earlier computed boundary. A value of $\lambda = 1.2$ is chosen and the number of suppliers is set to $|S| = 16$. We note that both heuristics perform close to the computed bound, with the threshold variant providing slightly less privacy. However, we observed that the threshold algorithm communicates on average with suppliers 3.29 per time period, which is far less than the available 16 suppliers. Thus, aggregating small samples reduces the required communication effort with minimal privacy trade-off.

5.3 Computation and Communication Complexity

The computation costs for distribution algorithms that implement the TD and SD strategy are very low in comparison to the proposed cryptographic aggregation protocols. The costs depend on a few, at most linear in the number of suppliers, symmetric cipher operations per time period. This is because the TD strategy only requires the generation of one secure random number per time period. The non-optimized heuristic for the SD strategy requires at most two random numbers per approached supplier.



(a) The maximal adversaries advantage when distinguishing load profiles with n samples distributed onto $|S|$ suppliers. The samples are allowed to differ by a factor of $\lambda = 1.2$.

(b) The maximal adversaries advantage when distinguishing load profiles distributed between $|S| = 16$ suppliers for a different number of load samples and various values for λ.

Fig. 4. Distinguishing advantage against both variants of the distribution algorithm.

Studying the communication patterns of both strategies, we observe that unidirectional communication is sufficient. Yet, the communication complexity varies for the TD and SD strategy. The TD strategy requires the same number of messages as an unprotected SMA, namely one message per load sample. In contrast, the SD strategy requires messages linear in the number of used suppliers $O(|S|)$. When using the presented threshold algorithm, on average the number of required messages reduces significantly.

In summary, being dependant on only symmetric ciphers and unidirectional communication, the computational and communication costs are very low when compared with other proposed solutions.

6 Conclusion

In this paper, we have introduced privacy-preserving, randomized buying strategies for an application in smart grids. Contrary to most approaches in the state of the art, these strategies do not presume the cooperation of suppliers nor expensive hardware at consumer side.

Our approach employs a distributed market to buy energy from multiple sources in order to protect the privacy of consumers. Our results indicate that it is not possible to conceal the complete energy consumption of a consumer, but at least it is feasible to conceal sensitive appliances, e.g., an alarm system. Based upon a formal model, we propose the indistinguishability notion of λ-IND that is capable of measuring the protection of such privacy sensitive appliances, which is supported by an evaluation on real-world data sets. Moreover, we have been able to show boundaries in the LSHG and under λ-IND in dependence on the number of readings to be protected and the number of available suppliers. Furthermore, we have developed an heuristic that approximates the SD strategy with low computational and communication overhead.

However, the provided level of privacy protection is fairly low compared to other approaches suggested so far. Even under the comparable weak definition of λ-IND, an adversary achieves non-negligible advantage when observing a larger number of samples. Privacy solutions in which consumers and utilities cooperate, e.g., aggregation protocols, provide stronger privacy protection.

Further work will be a detailed analysis of attackers with access to the information of multiple suppliers, e.g., colluding suppliers. Furthermore, hybrid strategies as well as algorithms that utilize unfair distribution strategies might be interesting candidates for a privacy analysis. Additionally, attacks against diversification strategies through pricing strategies could be evaluated.

Acknowledgments. This work has been co-funded by the German Federal Ministry of Education and Research (BMBF) within CRISP, by the DFG as part of project A.1 within the RTG 2050 "Privacy and Trust for Mobile Users" and by the Hessian LOEWE excellence initiative within CASED. At the time this research was conducted, Stefan Schiffner and Mathias Fischer were part of CASED at TU Darmstadt. Stefan Schiffner is currently employed at the European Union Agency for Network and Information Security (ENISA). The content of this article does not reflect the official opinion of ENISA. Responsibility for the information and views expressed in therein lies entirely with the authors.

A Constructing Minimal Distinguishable Distributions

To derive an optimal distribution strategy under λ-IND, multiple steps are necessary. First, we discuss the idea of probability transports. Then, given an input distribution and a new desired mean, we construct a new distribution with the specified mean, which has the least statistical distance to the input distribution. Finally, we compute the distinguishing advantage against this construction.

Probability Transport. A probability transport is the change of occurrence probabilities of two values in a (discrete) distribution. Transporting probability $y > 0$ from x_s to x_d implies that the likelihood to observe x_s decreases, while the likelihood to observe x_d increases by y. Given two distributions P_0 and P_1 that are separated by one transport, the change of mean $\Delta\mu = \mu^1 - \mu^0$ can be computed by $\Delta\mu = (x_d - x_s) \cdot y$, where y describes the transported probability, x_s the source, and x_d the destination value.

Optimal Construction. Given the definition of a transport and an input distribution P_0 with mean μ_0, we show how to construct the least distinguishable distribution P_1 that has a mean of $\mu_1 = \lambda \cdot \mu_0$. The best construction of P_1 is by transporting probability from the smallest possible x_s, where $P_0(x_s) > 0$ holds, to the largest possible $x_d = d^1 = \lambda \cdot d^0$. By this construction the mean increases with the least increase in the statistical distance, which only depends on the transported probability y. The accurate value y that is necessary for the transport to achieve a mean μ^1 is

$$y = \frac{\Delta\mu}{x_d - x_s} = \frac{\mu^1 - \mu^0}{d^1 - x_s}.$$

Note that multiple transports might be required if $P_0(x_s)$ does not provide sufficient probability.

Distinguishing Advantage. Given this construction, we show how the first distribution P_0 should be chosen, such that construction produces a pair of distributions that is the least distinguishable pair of distributions for the means μ_0 and μ_1. A transport from $x_s = 0$ to $x_d = d^1$ provides the best and thus least increase in the adversaries advantage while increasing the mean. Thus, we deduce that distribution P_0 needs sufficient probabilities $P_0(0) \geq y$ for a transport from 0. If this is the case then only one transport from 0 to d^1 is necessary to construct P_1 from P_0. A transport from some $x_s > 0$ implies that a larger amount has to be transported and therefore would result in a larger statistical distance.

Given two distributions constructed according the derived properties, we are able to link the advantage with the privacy parameter λ and the number of available suppliers $|S|$. The latter determines the required mean, when assuming a fair distribution algorithm. With only one transport, we can deduce the following distinguishing advantage:

$$\begin{aligned}\mathbf{Adv}^{\lambda\text{-IND}}_{SD,1} = y = \frac{\Delta\mu}{x_d - x_s} &= \frac{\mu^1 - \mu^0}{d^1 - 0} = \frac{d^1/|S| - d^0/|S|}{d^1} \\ &= \frac{\lambda \cdot d^0 - d^0}{|S| \cdot \lambda \cdot d^0} = \frac{(\lambda - 1) \cdot d^0}{|S| \cdot \lambda \cdot d^0} \\ &= \frac{\lambda - 1}{|S| \cdot \lambda}.\end{aligned}$$

References

1. Ács, G., Castelluccia, C.: I have a DREAM! (DiffeRentially privatE smArt Metering). In: Filler, T., Pevný, T., Craver, S., Ker, A. (eds.) IH 2011. LNCS, vol. 6958, pp. 118–132. Springer, Heidelberg (2011)
2. Backes, M., Meiser, S.: Differentially private smart metering with battery recharging. In: Garcia-Alfaro, J., Lioudakis, G., Cuppens-Boulahia, N., Foley, S., Fitzgerald, W.M. (eds.) DPM 2013 and SETOP 2013. LNCS, vol. 8247, pp. 194–212. Springer, Heidelberg (2014)

3. Baignères, T., Sepehrdad, P., Vaudenay, S.: Distinguishing distributions using chernoff information. In: Heng, S.-H., Kurosawa, K. (eds.) ProvSec 2010. LNCS, vol. 6402, pp. 144–165. Springer, Heidelberg (2010)
4. Bohli, J.-M., Sorge, C., Ugus, O.: A privacy model for smart metering. In: 2010 IEEE International Conference on Communications Workshops, pp. 1–5. IEEE, May 2010
5. Clark, S.S., Mustafa, H., Ransford, B., Sorber, J., Fu, K., Xu, W.: Current events: identifying webpages by tapping the electrical outlet. In: Jajodia, S., Mayes, K., Crampton, J. (eds.) ESORICS 2013. LNCS, vol. 8134, pp. 700–717. Springer, Heidelberg (2013)
6. Csisz, I., et al.: Information-type measures of difference of probability distributions and indirect observations. Studia Sci. Math. Hungar. **2**, 299–318 (1967)
7. Danezis, G., Kohlweiss, M., Rial, A.: Differentially private billing with rebates. In: Filler, T., Pevný, T., Craver, S., Ker, A. (eds.) IH 2011. LNCS, vol. 6958, pp. 148–162. Springer, Heidelberg (2011)
8. Dwork, C.: Differential privacy. In: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I. (eds.) ICALP 2006. LNCS, vol. 4052, pp. 1–12. Springer, Heidelberg (2006)
9. Dwork, C., Naor, M., Pitassi, T., Rothblum, G.N.: Differential privacy under continual observation. In: Proceedings of the 42nd ACM Symposium on Theory of Computing (STOC), pp. 715–724 (2010)
10. Efthymiou, C., Kalogridis, G.: Smart grid privacy via anonymization of smart metering data. In: International Conference on Smart Grid Communications (SmartGridComm), pp. 238–243. IEEE (2010)
11. Garcia, F.D., Jacobs, B.: Privacy-friendly energy-metering via homomorphic encryption. In: Cuellar, J., Lopez, J., Barthe, G., Pretschner, A. (eds.) STM 2010. LNCS, vol. 6710, pp. 226–238. Springer, Heidelberg (2011)
12. Greveler, U., Justus, B., Loehr, D.: Multimedia content identification through smart meter power usage profiles. Computers, Privacy and Data Protection CPDP, Brussels, Belgium (2012)
13. Hart, G.W.: Residential energy monitoring and computerized surveillance via utility power flows. IEEE Technol. Soc. Mag. **8**(2), 12–16 (1989)
14. Jawurek, M., Kerschbaum, F., Danezis, G.: Privacy technologies for smart grids - a survey of options. Technical report, Microsoft Research - Tech Report - 2012 - 119 (2012)
15. Kalogridis, G., Efthymiou, C., Denic, S.Z., Lewis, T.A., Cepeda, R.: Privacy for smart meters: towards undetectable appliance load signatures. In: IEEE International Conference on Smart Grid Communications (SmartGridComm), pp. 232–237 (2010)
16. Kolter, J.Z., Johnson, M.J.: REDD: a public data set for energy disaggregation research. In: SustKDD Workshop on Data Mining Applications in Sustainability, San Diego, CA, pp. 1–6 (2011)
17. Kursawe, K., Danezis, G., Kohlweiss, M.: Privacy-friendly aggregation for the smart-grid. In: Fischer-Hübner, S., Hopper, N. (eds.) PETS 2011. LNCS, vol. 6794, pp. 175–191. Springer, Heidelberg (2011)
18. Lin, H.-Y., Tzeng, W.-G., Shen, S.-T., Lin, B.-S.P.: A practical smart metering system supporting privacy preserving billing and load monitoring. In: Bao, F., Samarati, P., Zhou, J. (eds.) ACNS 2012. LNCS, vol. 7341, pp. 544–560. Springer, Heidelberg (2012)
19. Makonin, S., Popowich, F., Bartram, L., Gill, B., Bajic, I.V.: AMPds: a public dataset for load disaggregation and eco-feedback research. In: IEEE Electrical Power and Energy Conference, pp. 1–6 (2013)

20. Molina-Markham, A., Shenoy, P., Fu, K., Cecchet, E., Irwin, D.: Private memoirs of a smart meter. In: Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Building, pp. 61–66. ACM (2010)
21. Neyman, J., Pearson, E.S.: On the problem of the most efficient tests of statistical hypotheses. In: Kotz, S., Johnson, N. (eds.) Breakthroughs in Statistics. Springer Series in Statistics, pp. 73–108 (1992)
22. Pinsker, M.S.: Information and information stability of random variables and processes (1960)
23. Rial, A., Danezis, G.: Privacy-preserving smart metering. In: Proceedings of the 10th Annual ACM Workshop on Privacy in the Electronic Society, pp. 49–60. ACM (2011)
24. Varodayan, D., Khisti, A.: Smart meter privacy using a rechargeable battery: minimizing the rate of information leakage. In: IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 1932–1935 (2011)
25. Vaudenay, S.: On privacy models for RFID. In: Kurosawa, K. (ed.) ASIACRYPT 2007. LNCS, vol. 4833, pp. 68–87. Springer, Heidelberg (2007)
26. Wang, S., Cui, L., Que, J., Choi, D.-H., Jiang, X., Cheng, S., Xie, L.: A randomized response model for privacy preserving smart metering. IEEE Trans. Smart Grid **3**(3), 1317–1324 (2012)