On the Effectiveness of Re-Identification Attacks and Local Differential Privacy-Based Solutions for Smart Meter Data

Zeynep Sila Kaya, M. Emre Gursoy
Department of Computer Engineering, Koç University, Turkiye
{*zkaya18, emregursoy*}*@ku.edu.tr*

Keywords: Smart meter, energy consumption, privacy, differential privacy, re-identification attacks

Abstract: Smart meters are increasing the ability to collect, store and share households' energy consumption data. On the other hand, the availability of such data raises novel privacy concerns. Although the data can be de-identified or pseudonymized, a critical question remains: How unique are households' energy consumptions, and is it possible to re-identify households based on partial or imperfect knowledge of their consumption? In this paper, we aim to answer this question, and make two main contributions. First, we develop an adversary model in which an adversary who observes a pseudonymized dataset and knows a limited number of consumption readings from a target household aims to infer which record in the dataset corresponds to the target. We characterize the adversary's knowledge by two parameters: number of known readings and precision of readings. Using experiments conducted on three real-world datasets, we demonstrate that the adversary can indeed achieve high inference rates. Second, we propose a local differential privacy (LDP) based solution for protecting the privacy of energy consumption data. We evaluate the impact of our LDP solution on three datasets using two utility metrics, three LDP protocols, and various parameter settings. Results show that our solution can attain high accuracy and low estimation error under strong privacy guarantees.

1 Introduction

Conventional electricity meters are nowadays being replaced by *smart meters* which can record and transmit energy consumption data to consumers and utility providers. This adoption of smart meters is increasing the amount of energy consumption data that is available to be collected, stored, analyzed, and shared. On the other hand, such availability of energy consumption data raises novel privacy risks and concerns. Although attempts are made to de-identify or pseudonymize consumption data, a critical question remains: *How unique are households' energy consumption data, and therefore, is it possible to re-identify households based on partial and/or imperfect knowledge of their energy consumption?*

In this paper, we aim to answer this question by constructing a realistic adversary model and empirically measuring the adversary's success rate on real-world energy consumption datasets. We consider an adversary who observes a pseudonymized energy consumption dataset, and furthermore, knows a limited number of *(consumption, month)* pairs of a target household. The goal of the adversary is to infer which record in the pseudonymized dataset corresponds to the target household. To increase flexibility, we allow the adversary to know ℓ *(consumption, month)* pairs from the target household (we vary ℓ between 1 and 5), and we allow the adversary's knowledge to be precise (such as "exactly 915 kWh consumption") or imprecise (such as "consumption is between 900-1000 kWh"). The number of pairs is controlled by parameter ℓ and the degree of precision is controlled by parameter s.

We use three real-world datasets from London and Australia and two metrics to quantify the adversary's success rate: Uniqueness Ratio (UR) and Average Anonymity Degree (AAD). UR measures what percentage of households in the dataset are under risk of being uniquely identified. AAD measures the average degree of anonymity (similar to the notion of k-anonymity) with respect to the adversary's knowledge characterized by ℓ and s. When adversary's knowledge is precise, results show that UR $\geq 80\%$ as soon as $\ell = 2$. In addition, UR $\simeq 100\%$ when $\ell \geq 3$. Similarly, AAD values are low (e.g., households are only 3 or 4-anonymous), which shows that re-identification and de-anonymization are indeed serious risks, and psuedonymization may not be sufficiently effective for privacy protection. In order to achieve reasonably

strong privacy protection, precision needs to be heavily reduced, e.g., $s \geq 2$. However, such reduction in precision (e.g., through generalization) would come at a heavy cost of utility.

Motivated by this problem, we then explore a local differential privacy (LDP) based approach for privacy protection. We propose a solution in which: (i) each household bucketizes their consumption reading, (ii) buckets are perturbed to achieve LDP using popular protocols such as GRR, RAPPOR or OUE, (iii) perturbed buckets are sent to the data collector, and (iv) the data collector performs *estimation* to recover consumption statistics pertaining to the general population. In this solution, the data collector only observes perturbed outputs of LDP protocols, not the households' true consumption readings. The lack of a truthful energy consumption dataset prevents our aforementioned attack in the first place.

We experimentally measure the impact of our proposed LDP solution on three real-world datasets using various ε privacy parameters, various bucket range sizes R, three LDP protocols (GRR, RAPPOR, OUE), and two utility metrics (TCE and CHE) to quantify LDP estimation error. Results show that the three protocols agree in terms of optimal R values. Furthermore, our solution is able to attain high accuracy and low estimation error under reasonably strong privacy guarantees (e.g., $\varepsilon = 1$).

The rest of the paper is organized as follows. We review related work in Section 2. We present our adversary model and experimental results with our adversary model in Sections 3 and 4, respectively. We present our solution for applying LDP to energy consumption data and experimentally evaluate it in Section 5. Finally, Section 6 concludes the paper.

2 Related Work

Considering more and more data is becoming available nowadays, re-identification and de-anonymization attacks continue to be prominent privacy threats. The possibility of re-identification and de-anonymization attacks have been shown in various domains. The seminal work of (Sweeney, 2000) showed that 87% of the US population can be identified using a combination of zip code, gender, and date of birth. It is also possible to recover specific individuals' health records by linking voter registration records with health insurance data (Ohm, 2009). The work of (Benitez and Malin, 2010) provides a set of approaches for estimating the likelihood of de-identifying information in the context of data sharing policies associated with the HIPAA privacy rule. They defined two metrics and estimated the risk of a specific re-identification attack. In (Emam et al., 2013), a new model for estimating re-identification risk was developed and applied to Canada's post-marketing adverse drug event database. Locations and location trace data are also susceptible to de-anonymization and re-identification attacks. In (de Montjoye et al., 2013), it was shown that human location traces are highly unique, and as few as four low-resolution location points are enough to uniquely identify 95% of individuals within a population of half million. In (Yin et al., 2015), re-identification risks of mobile phone users in China were examined, and a quantitative relationship between re-identification risk and data utility for aggregate mobility analysis was studied.

Similar to these data types, *smart meter* and *energy consumption* data can also contain various privacy risks (McDaniel and McLaughlin, 2009), such as inferring which appliances are running (Eibl and Engel, 2014), predicting occupancy and holidays (Tang et al., 2015; Kleiminger et al., 2015; Eibl et al., 2019), and determining household characteristics as well as energy consumption profiles, even relating to socio-economic status (Beckel et al., 2013; Beckel et al., 2014; Anderson et al., 2017; Czétány et al., 2021). Among different privacy risks, closer to our work are attacks that focus on de-anonymization and re-identification. In (Buchmann et al., 2012), it was shown that households can be re-identified using simple statistical measures, and even simple means are sufficient to re-identify 68% of the records. Higher de-pseudonymization rates were achieved in later studies such as (Tudor et al., 2015) and (Cleemput et al., 2018). In a very recent work by (Radovanovic et al., 2022), the authors demonstrated that even if consumption data are anonymized, it is possible to identify a household with high accuracy by utilizing weekly consumption.

The prevalence of privacy risks associated with smart meters and energy consumption data has led to growing interest in engineering new privacy solutions based on obfuscation and/or perturbation (Pal et al., 2018; Khwaja et al., 2020). Towards this aim, in this paper, we consider the application of Local Differential Privacy (LDP). LDP has recently gained significant attention from academia and industry (Cormode et al., 2018; Gursoy et al., 2022), but its applications to smart meters and smart grid have been relatively few. In (Ou et al., 2020), a singular spectrum analysis-based LDP method has been proposed to prevent inference of household appliances. (Gai et al., 2022) developed a data aggregation scheme with LDP using randomized response. In (Parker et al., 2021), a new

Table 1: Sample energy consumption data (in kWh) from 4 households and across 4 months

id	01/2021	02/2021	03/2021	04/2021
1	1108	915	1013	972
2	802	712	788	793
3	278	241	267	312
4	551	462	495	479

variant of differential privacy called *spectral* DP was proposed, motivated by applications with unbounded time-series data such as smart meter data. A literature review covering the applications of different forms of differential privacy (not just LDP) was conducted in (Marks et al., 2021).

3 Adversary Model

3.1 Problem Setting and Notation

Consider a monthly energy consumption dataset as shown in Table 1. Each row corresponds to a different household and each column corresponds to a different month. The dataset shows how much energy (in kWh) was consumed by each household in each month, e.g., measured by a smart meter. For example, household 1 consumed 972 kWh in April 2021. The dataset is completely *pseudonymized*, i.e., random IDs are assigned to each row (determined arbitrarily) which aims to hide the identity of the corresponding households. Note that each row can be viewed as one *time series* corresponding to the monthly energy consumption of one household.

We denote the full dataset by D. Without loss of generality, we write $D = \{T_1, T_2, ..., T_n\}$ to denote that D contains n households, where T_i denotes the energy consumption time series of the i'th household. We write $T_i[j]$ to refer to the consumption amount of T_i in the j'th month. All time series in D have equal length, i.e., $|T_1| = |T_2| = ... = |T_n|$.

3.2 Adversary Formulation

We consider an adversary $\mathcal{A}$ who observes the pseudonymized dataset D, and furthermore, knows a limited number of *(consumption, month)* pairs from a certain household h. We denote this knowledge of the adversary by $\mathcal{K} = \{(c_1, m_1), (c_2, m_2), ...\}$ where c_i is to the consumption amount (in kWh) and m_i is the corresponding month. For example, $\mathcal{K} = \{(802, 01/2021), (712, 02/2021)\}$ means that the adversary knows the household h has consumed 802 kWh in January 2021 and 712 kWh in February 2021. The adversary is assumed to formulate such knowledge $\mathcal{K}$ through third-party resources or accessing meter readings in the physical world (e.g., observing h's electricity bill).

Given D and $\mathcal{K}$, the goal of $\mathcal{A}$ is to infer which time series in D corresponds to the household h. If the adversary is successful in uniquely identifying h from the dataset, then the adversary learns the remaining energy consumption readings of h. Continuing from the example in the previous paragraph, in Table 1, there is only one household which satisfies the knowledge $\mathcal{K} = \{(802, 01/2021), (712, 02/2021)\}$. Then, the adversary learns that id = 2 corresponds to household h; furthermore, the consumption readings of h for March 2021 is 788 kWh and for April 2021 it is 793 kWh. Note that such an inference can be quite sensitive in practice. For example, if a household consumes much lower energy in certain months compared to the rest of the months (e.g., in July-August the consumption is much lower) then this may indicate that the residents go on holiday or the household is unoccupied during these months.

The success of the adversary relies heavily on the amount of knowledge $\mathcal{K}$ available. We introduce two parameters for characterizing $\mathcal{K}$: length ℓ and precision s.

Length ℓ denotes how many (consumption, month) pairs are known by the adversary, i.e., $\ell = |\mathcal{K}|$. When ℓ is small, e.g., $\ell = 1$, it is more likely that there are multiple households in D which fit the adversary's knowledge. For example, in a dataset with thousands of households, there can be multiple households which have the same consumption reading for January 2021. On the other hand, as the value of ℓ increases, we expect the uniqueness of households to also increase.

Precision s: We account for cases in which the adversary's knowledge $\mathcal{K}$ may be imprecise. For example, instead of knowing that h's consumption in January 2021 is 802, the adversary may know that consumption is within the range [800, 900) without knowing the exact amount. The reason for the imprecision in $\mathcal{K}$ could be due to an inexact observation, as well as a potential privacy protection method (i.e., through generalization or masking).

In our work, we account for the imprecision using parameter s. The smallest possible value of s is $s = 0$, which means that $\mathcal{A}$ knows the exact integer consumption amount, e.g., consumption in January 2021 is 802. When $s = 1$, one least significant digit of the consumption amount is not known by the adversary. For example, instead of knowing the consumption amount is 802, the adversary knows that it is 80*, i.e., between [800, 810). When $s = 2$, two least significant digits are not known, e.g., the adversary's

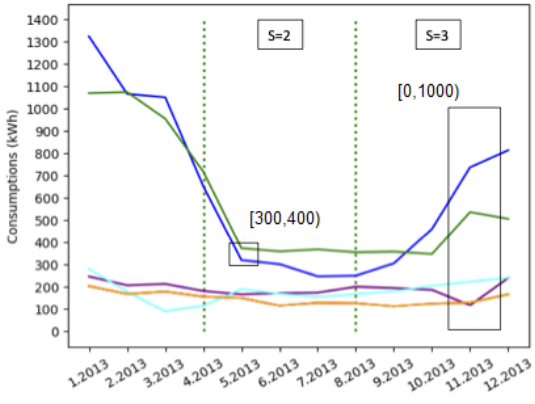

Figure 1: Monthly consumption of five households

knowledge is equal to 8** which corresponds to the range [800, 900). In general, s is equal to the number of least significant digits that are not known by the adversary (or are masked).

We visualize the impact of s using sample consumption data in Figure 1. When s is smaller, uniqueness of households increases. For example, when $s = 2$ and $\mathcal{K}$ consists of the knowledge [300, 400), there are two households which fit this criteria. When $s = 3$ and $\mathcal{K}$ consists of the knowledge [0, 1000), there are five households which fit this criteria. In general, as s increases, we expect households' uniqueness to decrease.

3.3 Measurement Metrics

Given an adversary who has D and $\mathcal{K}$, we would like to measure: (i) What percentage (ratio) of households can the adversary uniquely identify? (ii) How anonymous are the households? To answer these two questions, we define the following two metrics.

Uniqueness Ratio (UR): We say that $\mathcal{K}$ uniquely identifies a household in D if and only if there exists exactly one $T_i \in D$ such that:

$$\forall (c_j, m_j) \in \mathcal{K}: \quad T_i[m_j] = c_j$$

In other words, given $\mathcal{K}$, the adversary will be able to uniquely identify T_i in D since there does not exist any other time series which fits the conditions in $\mathcal{K}$.

Given D, ℓ and s, we generate all possible combinations of knowledge that an adversary may have which satisfy the ℓ and s parameters. Let $\mathbf{K}$ denote this set of all possible combinations. Note that $\mathbf{K}$ is potentially a very large set, e.g., for a single household and $\ell = 1$, there are $|T|$ different entries in $\mathbf{K}$, one for each month. Then, the Uniqueness Ratio of dataset D is defined as:

$$UR = \frac{\begin{array}{c}\#\text{ of } \mathcal{K} \in \mathbf{K} \text{ that uniquely identifies} \\ \text{a household in } D\end{array}}{|\mathbf{K}|}$$

By definition, the metric takes values between [0, 1]. Values closer to 0 imply better privacy for households (lower uniqueness).

Average Anonymity Degree (AAD): We say that T_i is k-anonymous with respect to $\mathcal{K}$ if and only if there exist $k-1$ other time series $S \subset D$ such that:

$$\forall T \in S, \ \forall (c_j, m_j) \in \mathcal{K}: \quad T[m_j] = T_i[m_j] = c_j$$

In other words, T_i appears indistinguishable from $|S| = k-1$ other time series to an adversary who has knowledge $\mathcal{K}$. Note that this is an adaptation of the well-known notion of k-anonymity (Sweeney, 2002; Samarati, 2001; Fung et al., 2010). Instead of requiring T_i to be indistinguishable from $k-1$ other time series in terms of *all* monthly consumption readings, our formulation requires T_i to be indistinguishable in terms of only the adversary's knowledge $\mathcal{K}$.

Let us define a function Φ which takes as input T_i, D and $\mathcal{K}$ and outputs the anonymity degree of T_i in D with respect to $\mathcal{K}$. That is: $\Phi(T_i, D, \mathcal{K}) = k$. Then, we generate all possible combinations of knowledge $\mathbf{K}$ similar to UR. Finally, Average Anonymity Degree (AAD) is formally defined as:

$$AAD = \frac{\sum_{\mathcal{K} \in \mathbf{K}} \sum_{i=1}^{n} \Phi(T_i, D, \mathcal{K})}{|D| \times |\mathbf{K}|}$$

Intuitively, AAD measures the anonymity degree of all time series in D in terms of all possible adversarial knowledge $\mathcal{K} \in \mathbf{K}$. Then, their average is computed to arrive at AAD. In the best case of privacy, all time series are indistinguishable from one another in terms of all $\mathcal{K}$. In this case, AAD becomes equal to $|D|$. In the worst case, where all time series are only 1-anonymous, AAD becomes equal to 1.

4 Experiment Results and Discussion

4.1 Experiment Setup and Datasets

We experimented with three real-world energy consumption datasets to measure the effectiveness of the aforementioned adversary model using the metrics given in Section 3.3. In our experiments, ℓ is varied between 1 and 5, and s is varied between 0 and 3. These parameters were chosen as such because of the following reasons: For s, the consumption readings in our datasets rarely went over 10.000; therefore all values of $s \geq 3$ would have had the same practical impact as $s = 3$. Hence, maximum s was determined as 3. For ℓ, we experimentally observed that $\ell = 5$ is sufficient

to cause either perfect or near-perfect uniqueness in many cases; thus, increasing it further would not have changed our results and findings.

Our three datasets come from two sources: The first dataset is from the London Datastore, whereas the second and third datasets are from Ausgrid, an electricity distribution company in Australia. The datasets are explained in more detail below.

London: The London Datastore is a platform established by the Greater London Authority to share data relating to the city of London. We extracted the Smart Meter Energy Consumption Data in London Households from (Datastore, 2023), which contains energy consumption readings for 5567 London households that took part in Low Carbon London project. Data collection occurred from November 2011 to February 2014; however, since many measurements in the years 2011 and 2014 as well as the first six months of 2012 were missing, we focused on the 18-month time period from July 2012 to December 2013. We also removed households whose readings contained null values, resulting in 4369 remaining households. Finally, to ensure consistency with our adversary model and remaining datasets, we aggregated the data in (Datastore, 2023) into monthly consumptions per household.

Solar: This dataset was provided by Ausgrid, an electricity distribution company in Australia. We downloaded the Solar Home Monthly Data from (Ausgrid, 2023), which contains electricity consumption data for 2657 solar households with rooftop solar systems installed in their houses. The data is provided for the time period between January 2007 and December 2014. After inspecting the Solar dataset, we observed 589 null (missing) values. We used backward interpolation to eliminate them.

Non-solar: Similar to the Solar dataset, the Non-solar dataset was also provided by Ausgrid. It can be accessed through the same URL. The dataset contains monthly energy consumption of 4064 households that had never installed a solar system. The Non-solar dataset also covers the time period between January 2007 and December 2014.

4.2 UR and AAD Results

In this section, we present Uniqueness Ratio (UR) and Average Anonymity Degree (AAD) results under different s and ℓ parameters. We use the plots in Figure 2 to show how UR results change according to s and ℓ. Interestingly, we observed very similar trends in all three datasets. First, as expected, as we increase s from 0 to 3, UR values decrease. A significant reduction is obtained even when s is changed from 0 to 1. Furthermore, UR values are very close to 0 when $s \geq 2$, yielding good privacy for households. This shows that reducing the precision of the adversary's knowledge, e.g., through *generalization*, is a possible solution against the adversary model considered in this paper. Second, we observe that as we increase ℓ from 1 to 5, UR values increase substantially. Although UR values are lower than 20% when $\ell = 1$, they jump to $\geq 80\%$ when $\ell = 2$ (assuming $s = 0$ in both cases). In other words, the fact that the attacker knows two consumption readings of the victim household rather than one reading increases the probability of uniquely identifying that household roughly 4 times. In general, UR values are alarmingly high when $\ell \geq 3$ and $s \leq 1$. In many cases, almost all households become uniquely identifiable (UR $\simeq 100\%$). In order to prevent identification and keep UR low, substantial reduction in precision must be present (i.e., $s = 2$ or 3).

We use the heatmaps in Figure 3 to report the changes in AAD values according to the s and ℓ parameters. In general, across all three datasets, anonymity levels decrease as ℓ increases and s decreases. Note that AAD values on the Solar dataset are relatively lower than the other datasets, but this is because the Solar dataset is smaller than the others. Interestingly, AAD values are quite high when $s \geq 2$, meaning that households can remain reasonably anonymous even when the adversary knows $\ell =$ 5 readings. On the other hand, a large difference is observed between $s = 1$ and $s = 2$. When $s = 1$, as long as $\ell \neq 1$, AADs are low (such as 1, 2, 4, 7) which means that households become almost unique in terms of their consumptions. When $s = 0$, low AADs are observed across all ℓ, i.e., all AADs converge to 1.

Combining the results presented in this section, in addition to confirming our expectations regarding the AAD and UR impacts of ℓ and s, we empirically observe that: (i) precise knowledge of consumption readings ($s = 0$) can indeed cause high risk of unique identification for households, (ii) knowing more than $\ell = 1$ consumption readings greatly increases the risk of identification and greatly decreases the degree of anonymity, and (iii) high amount of imprecision, such as $s = 2$ or 3, must be introduced in order to prevent adversarial inference effectively.

4.3 Impact of Seasonality

Next, in order to examine the impact of *seasonality*, we measure how UR results change from month to month in a calendar year. To do so, we construct knowledge $\mathbf{K}$ for each month separately and compute

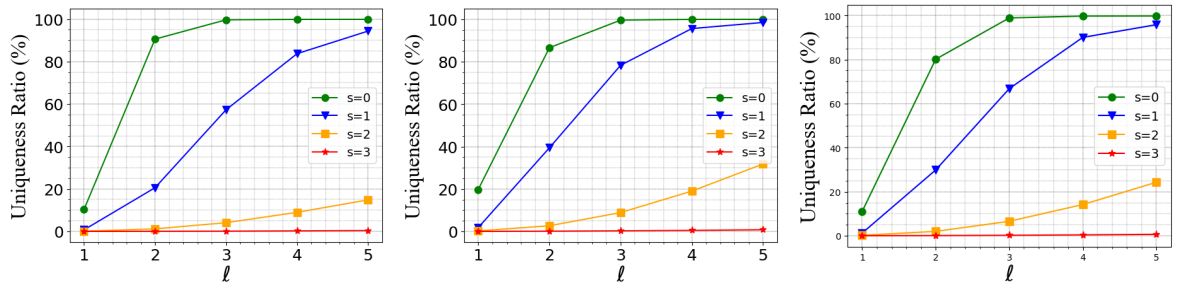

Figure 2: UR results on London, Solar, and Non-solar datasets respectively

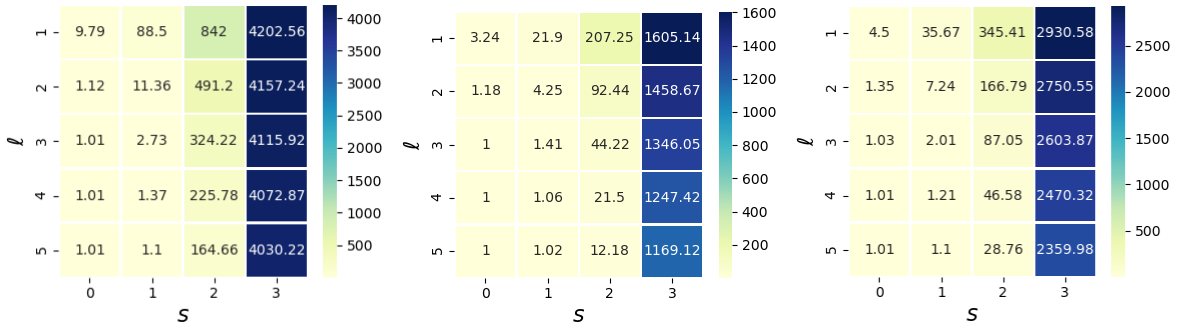

Figure 3: AAD results on London, Solar, and Non-solar datasets respectively

UR values in each month. The results are given in Figures 4, 5 and 6. On the London dataset, we perform this experiment only in year 2013, since 2013 is the only full year of data available. Three different years (2007, 2008, 2009) are considered for the Solar and Non-solar datasets. In all results, $\ell = 1$ and $s = 0$ are used.

An interesting observation from the London dataset is that UR results are lower between months June-September, whereas they are considerably higher between months January-March (e.g., 4% UR vs 8% UR). In other words, uniqueness drops in summer months whereas it rises in the winter months. A similar behavior is observed on Solar and Non-solar datasets, but since these datasets are from Australia, winter months are close to June-August whereas summer months are closer to January-March. Nevertheless, UR results are considerably higher in winter months compared to summer months. Furthermore, across different years and different datasets (2007 to 2009), this observation holds consistently. Thus, we find that uniqueness of households may indeed be impacted by seasons and weather conditions. Households typically consume higher energy in cold weather, and since households have different characteristics, they are likely to be reflected by the increased uniqueness ratios during winter.

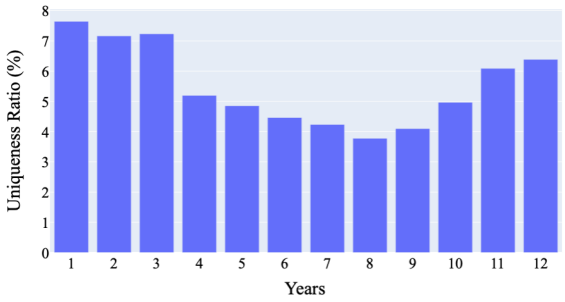

Figure 4: Seasonality on London dataset (year: 2013)

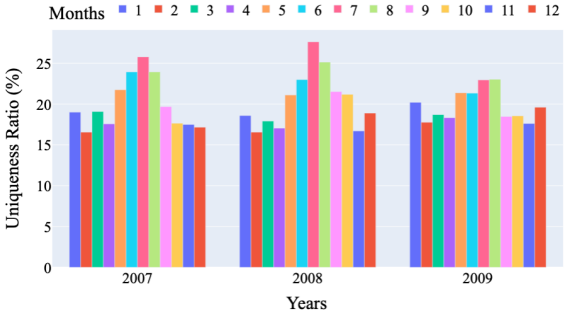

Figure 5: Seasonality on Solar dataset

5 Application of Local Differential Privacy (LDP)

Various approaches can be developed to address privacy risks concerning energy consumption data.

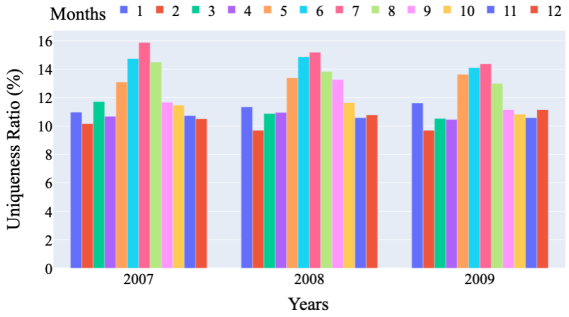

Figure 6: Seasonality on Non-solar dataset

Our results in the previous section showed that generalization-based methods constitute one possible option to reduce UR and AAD; however, large amount of generalization is necessary for them to be effective (e.g., $s = 2$). Another option is the use of differential privacy (DP) such that the full dataset is D is collected by the data collector in plaintext; yet, only a noisy access interface is provided to researchers and/or adversaries. However, the shortcomings of this approach are twofold: (i) it remains susceptible to side channel attacks, and (ii) it assumes that households inherently trust the data collector to hold their data in plaintext. Instead, in this paper we advocate for the use of *local differential privacy (LDP)* which recently emerged as a state-of-the-art notion for privacy protection.

5.1 LDP and LDP Protocols

In LDP, there exist several households (clients) and a data collector (server). To ensure LDP, each client's true consumption is encoded and perturbed by a randomized algorithm Ψ on the client side, and the perturbed output is sent to the data collector. Thus, the collected dataset D consists of perturbed readings, and households' true consumption readings are never revealed in D. Consequently, LDP can be used in scenarios where the data collector is untrusted by the households. Formally, LDP is defined as follows.

Definition 1 (ε-LDP)**.** A randomized algorithm Ψ satisfies ε-local differential privacy (ε-LDP), where $\varepsilon > 0$, if and only if for any two inputs v_1, v_2 in universe $\mathcal{U}$, it holds that:

$$\forall y \in Range(\Psi): \quad \frac{\Pr[\Psi(v_1) = y]}{\Pr[\Psi(v_2) = y]} \leq e^{\varepsilon} \qquad (1)$$

where $Range(\Psi)$ denotes the set of all possible outputs of Ψ.

After collecting perturbed readings from many clients, the data collector needs to perform *estimation* to recover statistics pertaining to the client population. For value $v \in \mathcal{U}$, let $C(v)$ denote the true count of v, i.e., number of times v is actually observed in the population. Let $\bar{C}(v)$ denote the estimated count of v, i.e., the count estimated by the server after LDP. The difference between $C(v)$ and $\bar{C}(v)$ is called the *estimation error*. Several LDP protocols were developed in the literature for minimizing estimation error under various settings. In this paper, we will use three popular LDP protocols: GRR, RAPPOR, and OUE.

Generalized Randomized Response (GRR) is a generalization of the randomized response survey technique introduced in (Warner, 1965) to support non-binary $\mathcal{U}$ and arbitrary ε. Denoting by v the client's true value, the perturbation algorithm Ψ_{GRR} perturbs v and outputs $y \in \mathcal{U}$ with probability:

$$\Pr[\Psi_{\text{GRR}}(v) = y] = \begin{cases} p = \frac{e^{\varepsilon}}{e^{\varepsilon}+|\mathcal{U}|-1} & \text{if } y = v \\ q = \frac{1}{e^{\varepsilon}+|\mathcal{U}|-1} & \text{if } y \neq v \end{cases} \qquad (2)$$

where $|\mathcal{U}|$ denotes the size of the universe. This satisfies ε-LDP since $\frac{p}{q} = e^{\varepsilon}$. The client sends y to the server.

On the server side, upon receiving perturbed outputs from all clients, to perform estimation for some value $v^* \in \mathcal{U}$ the server first first finds $\widehat{C}(v^*)$: total number of clients who reported v^* as their perturbed output. Then, estimate $\bar{C}(v^*)$ is computed as:

$$\bar{C}(v^*) = \frac{\widehat{C}(v^*) - |\mathcal{L}| \cdot q}{p - q} \qquad (3)$$

where $|\mathcal{L}|$ denotes the number of clients in the population.

RAPPOR was originally developed by Google and implemented in Chrome (Erlingsson et al., 2014; Fanti et al., 2016). While the original version of RAPPOR relies on Bloom filters for string encoding, in this paper we leverage a variant of RAPPOR which uses unary encoding, similar to (Wang et al., 2017; Gursoy et al., 2019).

Client initializes a bitvector B with length $|\mathcal{U}|$. The client sets $B[v] = 1$ and for all remaining positions $j \neq v$, those positions are set as: $B[j] = 0$. Then, the perturbation step of RAPPOR takes as input B and outputs a perturbed vector B'. Perturbation algorithm Ψ_{RAP} considers each bit in B one by one, and either keeps or flips the bit with probabilities:

$$\forall_{i \in [1,|\mathcal{U}|]}: \quad \Pr[B'[i] = 1] = \begin{cases} \frac{e^{\varepsilon/2}}{e^{\varepsilon/2}+1} & \text{if } B[i] = 1 \\ \frac{1}{e^{\varepsilon/2}+1} & \text{if } B[i] = 0 \end{cases} \qquad (4)$$

The client sends perturbed bitvector B' to the server.

The server receives perturbed bitvectors B' from all clients in the population. To perform estimation for value v^*, $Sup(v^*)$ is computed as the total number of received bitvectors that satisfy: $B'[v^*] = 1$. Then,

the estimate $\bar{C}(v^*)$ is computed as:

$$\bar{C}(v^*) = \frac{Sup(v^*) + |\mathcal{L}| \cdot (\alpha - 1)}{2\alpha - 1} \quad (5)$$

where α is the bit keeping probability: $\alpha = \frac{e^{\varepsilon/2}}{e^{\varepsilon/2}+1}$.

Optimized Unary Encoding (OUE) has the same encoding phase as RAPPOR with unary encoding, but its bit keeping and flipping probabilities are different. It treats the 0 and 1 bits asymmetrically to improve accuracy of server-side estimation (Wang et al., 2017; Jia and Gong, 2019).

Client initializes bitvector B with length $|\mathcal{U}|$ such that $B[v] = 1$, and for all remaining positions $j \neq v$, $B[j] = 0$. Perturbation algorithm Ψ_{OUE} takes as input B and produces perturbed bitvector B' such that:

$$\forall_{i \in [1,|\mathcal{U}|]}: \quad \Pr[B'[i] = 1] = \begin{cases} \frac{1}{2} & \text{if } B[i] = 1 \\ \frac{1}{e^{\varepsilon}+1} & \text{if } B[i] = 0 \end{cases} \quad (6)$$

The client sends perturbed bitvector B' to the server.

The server receives perturbed bitvectors B' from all clients in the population. To perform estimation for value v^*, $Sup(v^*)$ is computed as the total number of received bitvectors that satisfy: $B'[v^*] = 1$. Then, the estimate $\bar{C}(v^*)$ is computed as:

$$\bar{C}(v^*) = \frac{2 \cdot \left((e^{\varepsilon}+1) \cdot Sup(v^*) - |\mathcal{L}|\right)}{e^{\varepsilon} - 1} \quad (7)$$

5.2 Enforcing LDP on Energy Consumption Data

GRR, RAPPOR and OUE protocols assume a discrete universe $\mathcal{U}$ and perform perturbation within this discrete universe. Since households' energy consumption readings are typically numeric and continuous, these protocols are not directly applicable to energy consumption data. A straightforward solution for this problem would be to discretize each numeric consumption reading by rounding it to the nearest integer. However, this is also not a desirable solution, since it causes $\mathcal{U}$ to be extremely large. To exemplify, recall that the real-world datasets from Section 4 contained 4369, 2657 and 4064 households respectively, meaning that the number of households is in the order of 1000s. Considering that minimum consumption reading is typically 0 and maximum consumption reading can be quite large (e.g., 10000s or more), our $\mathcal{U}$ would be in the order of 10000s which is a magnitude larger than the number of households. This would cause the resulting statistics (i.e., number of households per unique reading) to be extremely sparse, which negatively impacts estimation utility. Additionally, large $\mathcal{U}$ also causes efficiency problems because the computational complexities of many protocols such as GRR, RAPPOR and OUE are at least linear in terms of $\mathcal{U}$, i.e., they are $\Omega(|\mathcal{U}|)$.

Algorithm 1: Our application of LDP

Input : Household population $\mathcal{L}$, privacy parameter ε, bucket range size R, number of buckets N, month j
Output: Estimates $\bar{C}(\cdot)$ for month j

```
   /* Client-side perturbation                    */
1  U ← {0,1,2,...,N−1}
2  for i ∈ L do
3  |  v ← T_i[j]//R
4  |  x ← Perturb v using LDP protocol with
   |       parameters ε and U // use Eqn 2,
   |       4 or 6 for GRR, RAPPOR or OUE
5  |  Send x to the server
6  end
   /* Server-side estimation                     */
7  for v* ∈ [0,N−1] do
8  |  C̄(v*) ← Estimate using LDP protocol
   |       with parameters ε and U // use Eqn
   |       3, 5 or 7
9  end
10 return C̄(v*) for all v*
```

Motivated by the above problems, we propose the following solution for applying LDP to energy consumption readings via *bucketization*. Given the range size (bucket size) R and the number of buckets N, we first construct buckets: $[0,R)$, $[R,2R)$, $[2R,3R)$, ..., $[NR-R,NR]$. When household i would like to perturb his/her consumption reading at month j, denoted by $T_i[j]$, the household first computes his/her true value as $v = T_i[j]//R$, where $//$ denotes integer division. This way, the household's consumption is assigned to one of the N buckets, and the true value v of the household becomes the number of that bucket. Considering there are N buckets, the universe $\mathcal{U}$ is now limited to: $\mathcal{U} = \{0,1,...,N-1\}$. After the household's true value v is determined and $\mathcal{U}$ is known as above, v can be fed into one of GRR, RAPPOR or OUE protocols, and the perturbed output can be obtained. The household sends the perturbed output to the data collector. After collecting perturbed outputs from all households, the data collector (server) performs estimation to find how many households are estimated to have consumption readings in each bucket.

The overall process is summarized in Algorithm 1. As explained above, first the universe of buckets $\mathcal{U} \leftarrow \{0,1,2,...,N-1\}$ is initialized on line 1 of the algorithm. Then, between lines 3-6, each household bucketizes his/her true value (line 3) and then perturbs it using an LDP protocol (line 4) such as GRR, RAP-

POR or OUE. Outcome of the perturbation is sent to the server (line 5). After the server receives perturbed outputs from all households, the server performs estimation (lines 7-9). Since values have been bucketized, the estimation of the server needs to be performed bucket-by-bucket. Thus, for each bucket (line 7), the estimation algorithm of the corresponding LDP protocol is used (line 8) such as GRR, RAPPOR or OUE. Results of the estimations are produced as the output of Algorithm 1 (line 10).

5.3 Quantifying Estimation Error

Due to the perturbation of LDP, the estimates recovered by Algorithm 1 will be imperfect, i.e., they will contain error. We propose two error metrics to measure estimation error: Consumption Histogram Error (CHE) and Total Consumption Error (TCE).

Consumption Histogram Error (CHE): Algorithm 1 recovers noisy LDP estimates $\bar{C}(v^*)$ for $v^* \in [0, N-1]$, which can be viewed as a histogram of number of households that have value $v^* = 0, 1, ..., N-1$. In contrast, let $C(v^*)$ denote the true number of households that have value v^* which would have been learned if all households' consumption readings were observed in plaintext (i.e., no privacy protection). CHE is computed as the average difference between $\bar{C}(v^*)$ and $C(v^*)$, i.e.:

$$\text{CHE} = \frac{\sum_{v^*=0}^{N-1} |\bar{C}(v^*) - C(v^*)|}{N}$$

Total Consumption Error (TCE): The *total* energy consumption of all households in the population is important to preserve, since it is important for demand and capacity planning in a city or an electricity grid. Let φ be the true total energy consumption in the current month, which would have been computed if all households' consumption readings were observed in plaintext. On the other hand, using the output of Algorithm 1, the expected total energy consumption under LDP, denoted by $\bar{\varphi}$, can be computed as:

$$\bar{\varphi} = \sum_{v^*=0}^{N-1} \bar{C}(v^*) \times \left(v^* R + \frac{R}{2}\right)$$

Then, TCE is computed as:

$$\text{TCE} = \frac{|\bar{\varphi} - \varphi|}{\varphi} \times 100\%$$

Multiplication by 100% is performed to turn TCE into a percentage and thereby increase interpretability.

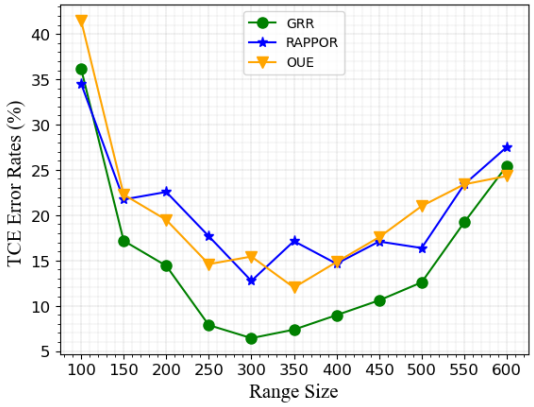

Figure 7: TCE results on London dataset ($\varepsilon = 1$)

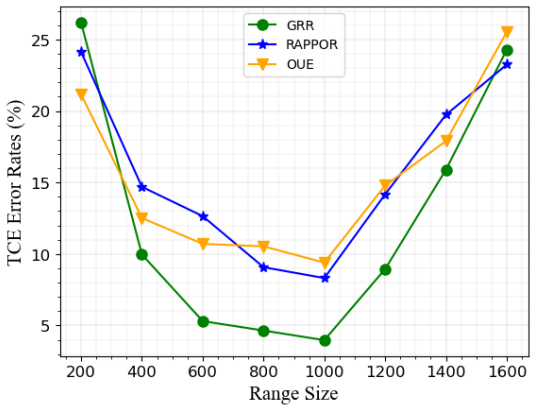

Figure 8: TCE results on Non-solar dataset ($\varepsilon = 1$)

5.4 LDP Experiment Results

In order to measure the potential real-world impact of our LDP solution via simulation, we perform experiments using the same datasets and setup from Section 4. We focus on the impacts of two parameters: privacy budget ε and range size R. We measure error amounts using TCE and CHE metrics. We perform experiments separately for different months in each dataset, compute the values of TCE and CHE metrics, and then take their average across all months.

In Figures 7 and 8, we fix $\varepsilon = 1$ and vary R, in order to analyze the impact of varying R on TCE. On both datasets and all three LDP protocols, the results show a U-shaped curve trend. In other words, in Figure 7, starting from $R = 100$ and gradually increasing R, we observe that TCE results decrease until R is between the 250-350 range, and afterwards TCE results start to increase as R is further increased to $R > 350$. Similarly, in Figure 8, starting from $R = 200$ and gradually increasing R, TCE results decrease until R is between 600-800, and afterwards TCE starts to increase as R is further increased. These figures

and observations yield three key insights. First, our proposed solution using bucketization is indeed more effective than simply rounding consumption readings to the nearest integer and then applying LDP. Rounding to the nearest integer is equivalent to setting $R = 1$, which, according to these results, would be expected to yield much higher TCE. Second, GRR, RAPPOR and OUE protocols all agree in the U-shaped curve trend, and they approach minimal error amounts for similar R. Due to their consistency, it is possible to choose R in a protocol-agnostic manner, and an R value that yields good results for one protocol is likely to yield good results also for other protocols. Third, assuming the use of a good R value, it is possible to estimate total consumptions with TCE $\leq 5\%$ or 10% even with a reasonably strict privacy budget of $\varepsilon = 1$. This shows that our proposed solution is feasible and can yield good utility in practice.

In Tables 2 and 3, we fix R and vary ε, in order to analyze the impact of varying ε on TCE and CHE. In this experiment, we chose near-optimal values of R, that is: $R = 300$ for London dataset, $R = 800$ for Non-solar dataset, and $R = 500$ for Solar dataset. Results with the TCE metric are reported in Table 2 and results with the CHE metric are reported in Table 3. We observe that the GRR protocol has lower error in terms of both TCE and CHE compared to RAPPOR and OUE across many ε settings. $\varepsilon = 0.1$ is the strictest privacy budget we use, and indeed, we observe that errors are quite high in this case. As we increase ε from 0.1 to 0.5, errors are substantially reduced. As we increase ε further to 1, 2, 4 and 6, errors are further reduced, although at a lower speed compared to the reduction from 0.1 to 0.5. When $\varepsilon \geq 4$, TCE values are $\leq 1\%$ and CHE values contain single digit, negligible errors (in case of GRR).

6 Conclusion

In this paper, we made mainly two contributions at the intersection of privacy and energy consumption data. First, we proposed an adversary model in which an adversary observes a pseudonymized energy consumption dataset and knows a limited number of consumption readings of a target household. The knowledge of the adversary is characterized by a length parameter ℓ and a precision parameter s. Using three real-world datasets and UR and AAD metrics, we experimentally showed the effectiveness of such an adversary's re-identification ability. Second, we proposed a LDP and bucketization-based solution for protecting the privacy of households' consumption readings. We measured the estimation error caused by our solution under various settings and parameter choices using CHE and TCE metrics. Results showed that our solution is able to achieve low estimation error under reasonably strong privacy guarantees such as $\varepsilon = 1$.

There are several directions for future work. First, we plan to study the results of the UR and AAD metrics after our LDP and bucketization-based solution from Section 5 is applied to the data. Second, we plan to integrate several additional LDP protocols to our work, which can reduce client-server communication cost (e.g., hashing-based protocols such as BLH and OLH) in bandwidth-constrained environments. Third, recall from Section 5 that bucket sizes are equal for all buckets (R). We plan to investigate the utility and privacy impacts of uneven bucket sizes, e.g., whether uneven bucket sizes can help improve estimation utility, and whether small bucket sizes may increase the risk of identifiability. Furthermore, we plan to investigate whether bucket sizes can be automatically learned from the underlying data.

ACKNOWLEDGEMENTS

We gratefully acknowledge the support from The Scientific and Technological Research Council of Turkiye (TUBITAK) under project number 121E303.

REFERENCES

Anderson, B., Lin, S., Newing, A., Bahaj, A., and James, P. (2017). Electricity consumption and household characteristics: Implications for census-taking in a smart metered future. *Computers, Environment and Urban Systems*, 63:58–67.

Ausgrid (2023). Solar home electricity data. https://www.ausgrid.com.au/Industry/Our-Research/Data-to-share/Solar-home-electricity-data.

Beckel, C., Sadamori, L., and Santini, S. (2013). Automatic socio-economic classification of households using electricity consumption data. *Proceedings of the fourth international conference on Future energy systems*.

Beckel, C., Sadamori, L., Staake, T., and Santini, S. (2014). Revealing household characteristics from smart meter data. *Energy*, 78:397–410.

Benitez, K. and Malin, B. (2010). Evaluating re-identification risks with respect to the hipaa privacy rule. *Journal of the American Medical Informatics Association*, 17:169–177.

Buchmann, E., Böhm, K., Burghardt, T., and Kessler, S. (2012). Re-identification of smart meter data. *Personal and Ubiquitous Computing*, 17:653–662.

ε	GRR	RAPPOR	OUE
0.1	95.64	135.12	171.72
0.5	16.05	27.99	27.24
1	6.59	15.55	13.24
2	2.80	5.02	5.32
4	1.00	3.46	2.71
6	0.63	1.69	1.77

ε	GRR	RAPPOR	OUE
0.1	81.93	108.77	105.50
0.5	11.52	20.54	19.07
1	4.04	10.29	10.12
2	1.63	5.57	4.27
4	0.88	2.37	2.18
6	0.71	1.38	1.64

ε	GRR	RAPPOR	OUE
0.1	53.46	84.39	80.45
0.5	8.29	19.39	19.42
1	3.86	9.22	9.78
2	1.38	4.67	4.61
4	0.66	2.01	2.61
6	0.45	1.27	2.02

Table 2: TCE results on London (left), Non-solar (mid), Solar (right) datasets

ε	GRR	RAPPOR	OUE
0.1	664.32	701.64	806.13
0.5	142.19	170.14	164.30
1	67.86	95.65	89.50
2	28.36	39.76	45.54
4	8.38	21.73	23.09
6	3.05	11.69	20.64

ε	GRR	RAPPOR	OUE
0.1	695.97	769.16	779.02
0.5	139.16	167.92	161.46
1	55.83	86.07	87.68
2	26.12	43.90	43.48
4	8.69	19.75	26.68
6	3.17	11.32	20.19

ε	GRR	RAPPOR	OUE
0.1	532.10	539.31	602.10
0.5	108.97	139.72	150.01
1	50.65	73.46	70.67
2	20.84	35.68	36.92
4	6.72	16.15	21.75
6	2.36	9.20	18.28

Table 3: CHE results on London (left), Non-solar (mid), Solar (right) datasets

Cleemput, S., Mustafa, M. A., Marin, E., and Preneel, B. (2018). De-pseudonymization of smart metering data: Analysis and countermeasures. *2018 Global Internet of Things Summit (GIoTS)*.

Cormode, G., Jha, S., Kulkarni, T., Li, N., Srivastava, D., and Wang, T. (2018). Privacy at scale: Local differential privacy in practice. In *Proceedings of the 2018 International Conference on Management of Data*, pages 1655–1658. ACM.

Czétány, L., Vámos, V., Horváth, M., Szalay, Z., Mota-Babiloni, A., Deme-Bélafi, Z., and Csoknyai, T. (2021). Development of electricity consumption profiles of residential buildings based on smart meter data clustering. *Energy and Buildings*, 252:111376.

Datastore, L. (2023). Smartmeter energy consumption data in london households. https://data.london.gov.uk/dataset/smartmeter-energy-use-data-in-london-households.

de Montjoye, Y.-A., Hidalgo, C. A., Verleysen, M., and Blondel, V. D. (2013). Unique in the crowd: The privacy bounds of human mobility. *Scientific Reports*, 3.

Eibl, G., Burkhart, S., and Engel, D. (2019). Insights into unsupervised holiday detection from low-resolution smart metering data. In *International Conference on Information Systems Security and Privacy (ICISSP)*, pages 281–302. Springer.

Eibl, G. and Engel, D. (2014). Influence of data granularity on smart meter privacy. *IEEE Transactions on Smart Grid*, 6(2):930–939.

Emam, E., Dankar, F. K., Neisa, A., and Jonker, E. (2013). Evaluating the risk of patient re-identification from adverse drug event reports. *BMC Medical Informatics and Decision Making*, 13.

Erlingsson, Ú., Pihur, V., and Korolova, A. (2014). Rappor: Randomized aggregatable privacy-preserving ordinal response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, pages 1054–1067. ACM.

Fanti, G., Pihur, V., and Erlingsson, Ú. (2016). Building a rappor with the unknown: Privacy-preserving learning of associations and data dictionaries. *Proceedings on Privacy Enhancing Technologies*, 2016(3):41–61.

Fung, B. C., Wang, K., Chen, R., and Yu, P. S. (2010). Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys (CSUR)*, 42(4):1–53.

Gai, N., Xue, K., Zhu, B., Yang, J., Liu, J., and He, D. (2022). An efficient data aggregation scheme with local differential privacy in smart grid. *Digital Communications and Networks*, 8(3):333–342.

Gursoy, M. E., Liu, L., Chow, K.-H., Truex, S., and Wei, W. (2022). An adversarial approach to protocol analysis and selection in local differential privacy. *IEEE Transactions on Information Forensics and Security*, 17:1785–1799.

Gursoy, M. E., Tamersoy, A., Truex, S., Wei, W., and Liu, L. (2019). Secure and utility-aware data collection with condensed local differential privacy. *IEEE Transactions on Dependable and Secure Computing*, 18(5):2365–2378.

Jia, J. and Gong, N. Z. (2019). Calibrate: Frequency estimation and heavy hitter identification with local differential privacy via incorporating prior knowledge. In *IEEE International Conference on Computer Communications (INFOCOM)*, pages 2008–2016. IEEE.

Khwaja, A. S., Anpalagan, A., Naeem, M., and Venkatesh, B. (2020). Smart meter data obfuscation using correlated noise. *IEEE Internet of Things Journal*, 7(8):7250–7264.

Kleiminger, W., Beckel, C., and Santini, S. (2015). Household occupancy monitoring using electricity meters. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 975–986.

Marks, J., Montano, B., Chong, J., Raavi, M., Islam, R., Cerny, T., and Shin, D. (2021). Differential privacy applied to smart meters: a mapping study. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, pages 761–770.

McDaniel, P. and McLaughlin, S. (2009). Security and pri-

vacy challenges in the smart grid. *IEEE Security & Privacy Magazine*, 7:75–77.

Ohm, P. (2009). Broken promises of privacy: Responding to the surprising failure of anonymization. *UCLA Law Review*, 57:1701.

Ou, L., Qin, Z., Liao, S., Li, T., and Zhang, D. (2020). Singular spectrum analysis for local differential privacy of classifications in the smart grid. *IEEE Internet of Things Journal*, 7(6):5246–5255.

Pal, R., Hui, P., and Prasanna, V. (2018). Privacy engineering for the smart micro-grid. *IEEE Transactions on Knowledge and Data Engineering*, 31(5):965–980.

Parker, K., Hale, M., and Barooah, P. (2021). Spectral differential privacy: Application to smart meter data. *IEEE Internet of Things Journal*, 9(7):4987–4996.

Radovanovic, D., Unterweger, A., Eibl, G., Engel, D., and Reichl, J. (2022). How unique is weekly smart meter data? *Energy Informatics*, 5.

Samarati, P. (2001). Protecting respondents identities in microdata release. *IEEE Transactions on Knowledge and Data Engineering*, 13(6):1010–1027.

Sweeney, L. (2000). Simple demographics often identify people uniquely. *Health (San Francisco)*, 671(2000):1–34.

Sweeney, L. (2002). k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570.

Tang, G., Wu, K., Lei, J., and Xiao, W. (2015). The meter tells you are at home! non-intrusive occupancy detection via load curve data. In *2015 IEEE International Conference on Smart Grid Communications (SmartGridComm)*, pages 897–902. IEEE.

Tudor, V., Almgren, M., and Papatriantafilou, M. (2015). A study on data de-pseudonymization in the smart grid. *Proceedings of the Eighth European Workshop on System Security*.

Wang, T., Blocki, J., Li, N., and Jha, S. (2017). Locally differentially private protocols for frequency estimation. In *Proc. of the 26th USENIX Security Symposium*, pages 729–745.

Warner, S. L. (1965). Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69.

Yin, L., Wang, Q., Shaw, S.-L., Fang, Z., Hu, J., Tao, Y., and Wang, W. (2015). Re-identification risk versus data utility for aggregated mobility research using mobile phone location data. *PLOS ONE*, 10:e0140589.