Privacy Measures and Storage Technologies for Battery-Based Load Hiding—an Overview and Experimental Study

Vadim Arzamasov
Karlsruhe Institute of Technology
Karlsruhe, Germany
vadim.arzamasov@kit.edu

Rebecca Schwerdt
Karlsruhe Institute of Technology
Karlsruhe, Germany
schwerdt@kit.edu

Shahab Karrari
Karlsruhe Institute of Technology
Karlsruhe, Germany
shahab.karrari@kit.edu

Klemens Böhm
Karlsruhe Institute of Technology
Karlsruhe, Germany
klemens.boehm@kit.edu

Tien Bach Nguyen
Karlsruhe Institute of Technology
Karlsruhe, Germany
tien.nguyen@student.kit.edu

ABSTRACT

Large-scale smart meter roll-outs all over the world are one effect of the ongoing energy transition. This poses a significant risk to consumer's privacy. Battery based load hiding (BBLH)—where an energy storage system is employed to obscure actual demand patterns—is one possibility to still retain privacy. In recent years many different BBLH algorithms have been proposed. But although most of them were assessed with some formally defined privacy measure, the current state of the art sorely lacks any comparability.

We give an overview of privacy measures which were proposed for this scenario, available storage technologies, and datasets used for the assessment of BBLH. Furthermore, we conduct a study of how all of these factors influence the different ratings of several state-of-the-art BBLH algorithms. Our results illustrate the need for standardization as well as further research into meaningful privacy measures. Achieving this is necessary for private households to make an informed decision which BBLH algorithm is best for their specific situation.

KEYWORDS

Privacy Measures, Storage Technologies, Battery-Based Load Hiding, Smart Meters, Comparability

ACM Reference Format:
Vadim Arzamasov, Rebecca Schwerdt, Shahab Karrari, Klemens Böhm, and Tien Bach Nguyen. 2020. Privacy Measures and Storage Technologies for Battery-Based Load Hiding—an Overview and Experimental Study. In *Proceedings of e-Energy (e-Energy'20).* ACM, New York, NY, USA, Article 4, 19 pages. https://doi.org/10.475/123_4

1 INTRODUCTION

In electrical grids, smart meters become more and more widespread. Such devices record power consumption, voltages and currents with high frequency (minutes or seconds) and typically send the data to the utility provider. Smart meters are assumed to be necessary for future electricity systems. This is because they allow monitoring the status of the grid in almost real time. This is important since power generation becomes increasingly volatile with a higher share of renewable and distributed energy resources.

However, high-resolution data collected by smart meters (*load profiles*) can leak sensitive information. For private households, this includes employment status [61], sleep-wake cycles [32, 33, 60, 67, 74, 80, 85, 100], habits [21, 22, 32, 38, 43, 61, 72, 79, 102] and many other characteristics – cf. [8, 9, 11, 16, 18, 20, 30, 36, 41, 45, 50, 53, 54, 57, 62, 65, 86, 87, 91, 92, 98, 99, 104]. The privacy of commercial and industrial costumers can be violated similarly. Numerous methods have been proposed to minimize the leakage of private information. These privacy-enhancing methods either change the consumption measured by the smart meter, or they alter the data after collection but before sending it to the utility provider. However, such a retroactive manipulation of smart meter data is not possible under the legislation of most countries. Germany is an example [3]. It may also be infeasible, since the provider directly controls the metering infrastructure, like in the US [1]. So we only consider methods from the first category in this work.

A common technique to modify consumption is battery-based load hiding (BBLH). They deploy energy storage systems to mask electricity-consumption patterns. Various BBLH algorithms have been proposed and evaluated. But they have rarely been compared in a meaningful way. When comparisons were attempted [22, 100, 102, 104], they lack breath regarding potentially important factors. The following example illustrates this.

Example 1.1. Bob's household has a smart meter installed, and he decides to buy a Li-ion battery to mask his private information. He knows that BBLH Algorithm A has been shown to outperform Algorithm B when tested with (i) specific battery characteristics, on (ii) data with the hourly resolution coming from a three-person household when (iii) 'relative entropy' quantifies the level of privacy. However, Bob does not know what this privacy measure means, whether there are other measures, and, if so, which one might be better. His battery has characteristics beyond the ones used for testing, he lives alone, and his smart meter takes measurements at a higher frequency. So is Algorithm A still preferable for Bob?

In line with the literature, three factors may affect the relative performance of BBLH algorithms:

e-Energy'20, June 23-25 2020, Melbourne, Australia

ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

(i) *Privacy measure.* Various measures were proposed to quantify the privacy level provided by charging algorithms. These measures typically take original and modified load profiles as input and output a value representing the privacy gain or absolute level of achieved privacy.
(ii) *Storage characteristics.* Capacity, charging/discharging rates (power) and round trip efficiency often serve as inputs for BBLH algorithms.
(iii) *Data.* Discretization rate, length of a load profile, household location, behavior of inhabitants and other information associated with a load profile may affect the performance of a BBLH algorithm.

Contributions. To our knowledge, the combined influence of these factors has not been studied before. This paper closes this gap. More specifically: (1) We review privacy measures proposed in the literature on privacy of smart meters. We cover all measures we are aware of and comment on their properties, possible design alternatives and assumptions they rely on. There is a broad variety of measures, so this part is relatively long. (2) We review characteristics of and existing technologies for storage, focusing on their usefulness for privacy protection of private households. (3) We perform an exhaustive experimental study of the relative performance of five BBLH algorithms. We vary privacy measures, storage characteristics and load profiles. Our results suggest that all three factors are significant: Different privacy measures, when evaluated on varying load profiles and storage characteristics, rank the tested BBLH algorithms very differently. In other words, there is no BBLH algorithm which is best for a majority of combinations of these factors. This result underlines the need for standards for the evaluation of privacy algorithms. (4) We take a first step towards such a standard by discussing ways to select data sets for evaluation and choosing the most suitable privacy measure. We also consider other relevant factors one must observe when developing or evaluating BBLH algorithms. (5) We share our implementation of BBLH algorithms, privacy measures and our experiments[1]

Outline. Sections 2–5 review the four components of our study: privacy measures, storage systems, load profiles and BBLH algorithms. Section 6 describes the experimental setting and results. Section 7 discusses the results and possible future work.

2 PRIVACY MEASURES

In this section, we review measures which previous work has proposed or deployed to quantify smart meter privacy. The measures typically compare the consumption before and after modification through BBLH and output a single value representing the degree of privacy. Many measures require additional information such as the respective pricing function, or rely on additional assumptions, e.g., that load profiles follow first-order Markov processes. Depending on the type of additional information, we distinguish between *general* and *specialized* privacy measures. For now we concentrate on general privacy measures and refer to Appendix A for specialized ones. Before going into details, we introduce some notation.

[1]https://github.com/Arzik1987/SaP

2.1 Notation

$x^T = (x_1, \dots, x_T)$ denotes the (finite or infinite) time series of an original load profile, measured at equidistant times $t \in \{1, \dots, T\}$. A value x_t either gives the amount of energy consumed in the time period $[t-1, t)$, measured in kWh, as in [85], or the power in kW needed at time t, as in [100]. The literature suggests the terms *user load* [42] or *demand load* [100] for x^T. Applying BBLH to a user load results in a modified load $y^T = (y_1, \dots, y_T)$, referred to as *grid load* or *external load*. The modification $b_t := y_t - x_t$ ($t \in \{1, \dots, T\}$) from user to grid load is the in-/output of the storage system.

One can think of x^T as a realization of a multivariate random variable $X^T = (X_1, \dots, X_T)$ with a density function f_{X^T}. f_{X_t} denotes the marginal densities of its components. Many privacy measures discussed in this section take random variables X^T or Y^T as input. The distributions of these variables are usually estimated from realizations x^T, y^T. This however is impossible without additional assumptions. This is because x^T, y^T are only single samples from the distributions. In some cases, X^T and Y^T are estimated from sets of realisations $\{x^T\}, \{y^T\}$. A common assumption is that X_t, Y_t are independent from time, i.e., there are functions f_X and f_Y with $f_{X_t} = f_X$ and $f_{Y_t} = f_Y$ for all $t \in \{1, \dots, T\}$. Hence we use X and Y (without sub-/superscript) to denote X_t or Y_t, respectively, whenever they do not depend on the time t. Furthermore, most papers treat the random variables X, Y as discrete and estimate their probability mass functions P_X, P_Y instead of probability density functions. Discreteness is usually gained by quantizing the underlying load profiles x^T, y^T into bins. In most cases these bins are of equal size, except for [85] where the size follows the μ-law [15].

Some authors use the first differences $\Delta x_t = x_t - x_{t-1}$ and $\Delta y_t = y_t - y_{t-1}$, $t \in \{2, \dots, T\}$ instead of the load profiles x^T and y^T respectively. These can be construed as the realization of random variables ΔX^T, ΔY^T, and we extrapolate the notation so far to this case. We use Ω_U to denote the sample space of a random variable U, where U can be any of the variables X^T, Y^T, X, Y etc.

Table 1 lists notation used for all privacy measures. Notation specific to individual measures is explained whenever it appears.

2.2 General Privacy Measures

Table 2 lists all surveyed general privacy measures. It also states whether a paper applies the respective measure to original load profiles (x, y) or to their time differences $(\Delta x, \Delta y)$.

2.2.1 *Cluster Similarity.*

Cluster similarity is calculated as follows:

(1) Determine the number of bins $n \in \mathbb{N}$ for clustering.
(2) Cluster x^T into n clusters $C_i \subseteq \{x_1, \dots, x_T\}$, $i \in \{1, \dots, n\}$. Let the value $c_i \in \mathbb{R}$ be the cluster center of C_i. W.l.o.g. we assume that the sequence $(C_1, \dots, C_n)$ of clusters is ordered according to their corresponding values $c_1 < \dots < c_n$.
(3) Cluster y^T into n clusters $D_i \subseteq \{y_1, \dots, y_T\}$, $i \in \{1, \dots, n\}$, with cluster centers $d_i \in \mathbb{R}$. Again we assume w.l.o.g. that the sequence $(D_1, \dots, D_n)$ of clusters is ordered according to their corresponding values $d_1 < \dots < d_n$.

Notation	Interpretation
$\mathbb{1}$	Indicator function: $\mathbb{1}[\alpha] = 1$ if the statement α is true, otherwise $\mathbb{1}[\alpha] = 0$.
b_t	Battery load: Electricity demand of the storage system at time t.
b^T	$= (b_1, \dots, b_T)$. Battery load time series.
Δ	Difference operator: $\Delta x_t = x_t - x_{t-1}$ and $\Delta x^T = (\Delta x_2, \dots, \Delta x_T)$.
$\mathbb{E}$	Expectation $\mathbb{E}[U]$ of a random variable U.
f_U	Probability density function of a continuous random variable U.
$\mathbb{P}$	Probability $\mathbb{P}[u]$ of the event u occurring.
P_U	Probability mass function of a discrete random variable U.
t	$\in \{1, \dots, T\}$. Timestamp, at which the smart-meter takes a measurement.
T	$\in \mathbb{N} \cup \{\infty\}$. Number of timestamps, at which the smart-meter takes measurements.
x_t	User load: Electricity demand of all household appliances. Either the energy consumed in the period $[t-1, t)$ or the power measured at time t.
x^T	$= (x_1, \dots, x_T)$. User load time series.
X^T	$= (X_1, \dots, X_T)$. Random user load vector with realizations $x^T \leftarrow X^T$.
X	Univariate random variable. Used to denote X_t whenever it is independent of time t.
y_t	Grid load: Electricity demand of all household appliances plus storage at time t.
y^T	$= (y_1, \dots, y_T)$. Grid load time series.
Y^T	$Y^T = (Y_1, \dots, Y_T)$. Random grid load vector with realizations $y^T \leftarrow Y^T$.
Y	Univariate random variable. Used to denote Y_t whenever it is independent of time t.
u^*	Estimate of a value u.
Ω_U	Sample space of a random variable U.

Table 1: Overview of common notation.

(4) Compute the cluster similarity as the ratio of times t where x_t and y_t are classified differently:

$$CS\left(x^T, y^T\right) := \frac{1}{T} \sum_{t=1}^{T} \mathbb{1}\left[\exists i \in \{1, \dots, n\} : x_t \in C_i \wedge y_t \notin D_i\right].$$

The assumption behind this privacy measure is that clustering results for user and grid load will not differ much if they are similar, giving low privacy. Kalogridis et al. use $CS(x, y)$ in [50], $CS(\Delta x, \Delta y)$ in [53], and both in [52]. All these studies exclude the first clusters C_1 and D_1 as they represent periods of negligible electricity demand. The choice of clustering algorithm and of the number of clusters influence the output of this privacy measure. The above mentioned research proposes to use the Clara clustering algorithm [56] (for determinism and scalability reasons) and silhouette maximization [81] to specify n in Step (1). But there are many approaches to perform clustering analysis which have different meta-parameters and rely on different ideas. Alternatives to silhouette are available as well to quantify clustering quality and select the number of clusters [7, 68].

Measure	(x, y)	$(\Delta x, \Delta y)$
Cluster similarity	[50, 52]	[52, 53]
Coefficient of determination	[20, 52]	[50, 52, 53]
Conditional entropy	[86, 101]	[53]
Entropy ratio		[72]
Feature mass		[72, 99, 102]
K-divergence	[51]	
KL divergence	[104]	[53, 54, 98]
Load variance	[85]	
Mutual information	[8, 21, 22, 34, 41, 43, 45, 60, 63, 79, 82, 85–87, 92, 102, 104]	[100, 102]
Removed uncertainty	[61]	
Total variation distance	[50, 104], [11][a]	

[a] It is not clear from the description whether time differences are taken.

Table 2: Overview of general smart-meter privacy measures.

2.2.2 (Pseudo-) Coefficient of Determination. Let y_t^* be a function attempting to predict y_t from x_t without additional information[2].

Let furthermore $\bar{y} := (y_1 + \dots + y_T)/T$. We define:

$$TSS := \sum_{t=1}^{T} (y_t - \bar{y})^2, \quad ESS := \sum_{t=1}^{T} (y_t^* - \bar{y})^2, \quad RSS := \sum_{t=1}^{T} (y_t - y_t^*)^2$$

standing for total sum of squares, explained sum of squares and residual sum of squares. Textbooks in statistics and econometrics [5, 46, 96] define the coefficient of determination either as the average deviation of the predicted values from the mean in relation to the average deviation of the original values:

$$R_1^2 := ESS/TSS \quad \text{or as} \quad R_2^2 := 1 - RSS/TSS.$$

like [10, 48] do. Let us additionally define the pseudo-coefficient of determination[3] in line with Chen et al. [20] as

$$R_p^2\left(x^T, y^T, (y^*)^T\right) = 1 - RSS/(RSS + TSS).$$

If y_t^* is estimated through ordinary least squares (OLS) then $TSS = RSS + ESS$ and $R_1^2 = R_2^2$ [25]. If user and grid loads are similar, then one can construct a (linear) estimator y_t^* which resembles the behavior of y_t well. In this case $RSS \ll TSS$, and all measures take values close to 1, indicating low privacy.

Kalogridis et al. [50, 52, 53] use R_2^2; they align load profiles maximizing their cross-correlation and obtain y^* through OLS[4]. Kalogridis et al. use $\Delta x^T, \Delta y^T$ in [50, 52, 53] and x^T, y^T in [52][5]. Chen et al. [20] compute $R_p^2(x^T, y^T, (y^*)^T)$ with $y_t^* = x_t$.

2.2.3 Entropy Ratio. Entropy quantifies the uncertainty of a random variable. The entropy of U is defined as

$$\mathrm{H}(U) := -\sum_{u \in \Omega_U} P_U(u) \log_b \left[P_U(u)\right].$$

[2] Note that—although it might seem counterintuitive—it is indeed y_t which is estimated from x_t, not vice versa.

[3] The authors also call it "coefficient of determination", although in general it does not coincide with R_1^2 or R_2^2

[4] In fact they use the equation $1 - RSS/(RSS + ESS)$, which coinsides with R_2^2 in their case, as we have just explained.

[5] We observe, that one does not need the OLS estimation in these works to compute coefficient of determination, since it is equal to the cross-correlation squared [46]

For example, an experiment with a highly unfair coin (e.g., falling tails with a probability close to 1) has a less uncertain outcome than one conducted with a fair coin. Consequently it has lower entropy. The common value of b is 2, and the corresponding unit of entropy is bit. McLaughlin et al. [72] use the ratio $ER(\Delta X, \Delta Y) := \mathrm{H}(\Delta Y)/\mathrm{H}(\Delta X)$ to measure privacy. They estimate $P_{\Delta Y}$ and $P_{\Delta X}$ in two ways – using zero values in Δy^T, Δx^T (ER^{z}) and excluding them (ER^{nz}). They assume that a greater reduction of entropy due to BBLH stands for a higher difference between user load and grid load, i.e., better privacy. However, Δx^T and Δy^T can differ significantly, still having the same entropy.

2.2.4 Feature Mass. A "feature" in this context is defined as one ocurrence of the time series value exceeding a predefined threshold *thr*. So the feature mass *FM* of a user load x^T with respect to the threshold *thr* is defined as:

$$FM(x^T) := \sum_{t=1}^{T} \mathbb{1}\left[x_t > thr\right].$$

To be able to compare user and grid load, *CFM* gives the number of points in time where features of both time series coincide:

$$CFM(x^T, y^T) := \sum_{t=1}^{T} \mathbb{1}\left[|x_t| > thr \wedge |y_t| > thr\right].$$

Several proposals for smart meter privacy derived measures from *FM* and *CFM*. McLaughlin et al. [72] use the *relative feature mass*

$$FM_r(\Delta x^T, \Delta y^T) := FM(\Delta y^T)/FM(\Delta x^T)$$

with $thr = 0$; Yang et al. [99] use $FM(\Delta y^T)$ with $thr = 50W$; Zhao et al. [102] introduce *event detection accuracy* with $thr = 50W$:

$$FM_{ed}(\Delta x^T, \Delta y^T) := CFM(\Delta x^T, \Delta y^T)/FM(\Delta y^T).$$

Low values supposedly indicate high privacy.

2.2.5 K-Divergence. One computes the K-divergence K of random variables V and U with joint domain $\Omega_U = \Omega_V$ as

$$K(U\|V) = \sum_{u\in\Omega_U} P_U(u) \log \frac{2P_U(u)}{P_U(u) + P_V(u)}.$$

It is used by [51] where $K(X\|X_{\mathrm{avg}})$ is compared to $K(Y\|Y_{\mathrm{avg}})$ or $K(Y\|X_{\mathrm{avg}})$. X_{avg} (resp. Y_{avg}) denotes a random variable of average consumption estimated from 30 different user loads (or rather different time windows in the same user load) instead of just one. Higher K-divergence $K(Y\|Y_{\mathrm{avg}})$ supposedly indicate higher levels of privacy, as it stands for more variability and therefore more randomness in the time series. In our experiments we use $K(X\|Y)$.

2.2.6 Kullback–Leibler Divergence. This privacy measure is sometimes referred to as *relative entropy*. For two random variables U and V sharing the sample space Ω_U, it quantifies how different the distribution of U is from that of V:

$$KL(U\|V) := \int_{u\in\Omega_U} f_U(U) \log \frac{f_U(u)}{f_V(u)} du$$

For discrete distributions P_U and P_V the integral simplifies to

$$KL(U\|V) := \sum_{\omega\in\Omega_U} P_U(u) \log \frac{P_U(u)}{P_V(u)}.$$

Kullback–Leibler divergence is not symmetric in general, i.e., there are U, V with $KL(U\|V) \neq KL(V\|U)$. Kalogridis and others [53, 54] use $KL(\Delta X\|\Delta Y)$, while Yang and others [98] use $KL(\Delta Y\|\Delta X)$. This measure is also mentioned but not used in [104], where the authors operate on distributions X and Y of non-differenced load profiles.

In order to compute this measure, Kalogridis and others [53, 54] use the last formula and preliminarily quantize load profiles in bins of equal size. In [53] the bin size is set to 6W, whereas in [54] they consider different sizes and recommend 100W. Intuitively, more bins result in better estimates if the length of load profiles is sufficiently large. Otherwise, some bins will have a small number of observations, resulting in an unreliable estimate of Kullback–Leibler divergence. Yang and others [98] do not specify the procedure they use to calculate this measure.

2.2.7 Load Variance. Tan et al. [85] propose the load variance

$$LV(y^T) := \frac{1}{T}\sum_{t=1}^{T}(y_t - E)^2$$

as a privacy measure, where E is a constant picked by the user. They argue that a constant load would yield perfect privacy. Hence a deviation of the grid load from some constant would be a good indicator of privacy. If E is the average grid load $\frac{1}{T}(y_1 + \cdots + y_T)$, this is nothing but the sample variance of the time series y^T.

2.2.8 Mutual Information. Mutual information is one of the most widely used privacy measures for BBLH algorithms. The mutual information of two random vectors U and V is

$$I(U;V) := \sum_{u\in\Omega_U}\sum_{v\in\Omega_V} P_{UV}(u,v) \log \frac{P_{UV}(u,v)}{P_U(u)\,P_V(v)}.$$

One often uses the average mutual information

$$MI\left(X^T, Y^T\right) := \frac{1}{T}\cdot I\left(X^T; Y^T\right)$$

as a privacy measure, also referred to as *information leakage rate*. Since generally only one sample x^T, y^T of the random variables X^T, Y^T is available, further assumptions, like i.i.d. or Markov properties, are often employed to estimate X^T, Y^T from given data and to simplify the calculation of *MI*.

Under the assumption that X_t are i.i.d. and Y_t are i.i.d., like in [21, 22, 34, 41, 43, 45, 100, 102, 104], this formula simplifies to

$$MI\left(X^T, Y^T\right) \overset{\mathrm{i}}{=} I(X;Y).$$

Some authors [85, 100] make the less strict assumption of stationary first-order Markov processes, so the formula becomes

$$MI\left(X^T, Y^T\right) \overset{\mathrm{s}}{=} \frac{(T-1)I(X_t, X_{t-1}; Y_t, Y_{t-1}) - (T-2)I(X_t; Y_t)}{T}.$$

In contrast to [100], Tan et al. [85] only model X^T and (X^T, Y^T) as stationary first-order Markov processes, but not Y^T. In this case the privacy measure MI_s, defined by the right hand side of the above equation, provides a lower bound $MI_s(X^T, Y^T) \leq MI(X^T, Y^T)$ for the mutual information.

Chin et al. [22] propose so-called feature-dependent first-order Markov processes to model consumption $(\tilde{X}^K, \tilde{Y}^K)$, $K \leq T$. To construct $(\tilde{X}^K, \tilde{Y}^K)$, the period $\{1, \ldots, T\}$ is split into K intervals $\{1, \ldots, T_1\} \cup \cdots \cup \{T_{K-1}, \ldots, T_K = T\} = \{1, \ldots, T\}$. Within each

interval $\{T_{k-1}, \dots, T_k\}$, with $k \in \{1, \dots K\}$, the values x_t resp. y_t are assumed to be i.i.d. realizations of the random variables $\tilde{X}_k$ resp. $\tilde{Y}_k$ which form first order Markov processes $\tilde{X}^K$ and $\tilde{Y}^K$. The resulting privacy measure is the mutual information $MI^{\mathsf{m}}(\tilde{X}^K, \tilde{Y}^K)$ of these (not necessarily stationary) first-order Markov processes. Yang et al. [100] use $MI(\Delta X^T, \Delta Y^T)$ with and without assumption of a first-order Markov process; they also apply it to 'binary versions' of Δx^T, Δy^T, i.e., when these load profiles are quantized using two bins to estimate respective probability functions. We call this version MI^{b} in our experiments. Zhao et al. [102] apply mutual information to both differenced and non-differenced load profiles, the other references mentioned above use $MI(X^T, Y^T)$. Other work using or mentioning average mutual information as a measure of smart meter privacy is [8, 60, 63, 79, 82, 86, 87, 92].

2.2.9 Conditional Entropy. Conditional entropy—also referred to as *equivocation*—of a random variable U given V is the difference of the entropy of U and the mutual information of U and V:

$$H(U|V) := H(U) - I(U;V).$$

High relative entropy indicates low similarity of U and V. Conditional entropy is mentioned as a potential privacy measure in [53, 86]. Yao and others [101] use a slightly more specialized variant of conditional entropy: They include a random variable P^T which models the time series of prices and also normalize the measure to

$$\lim_{T\to\infty} H\left(X^T \middle| Y^T, P^T\right) \Big/ \lim_{T\to\infty} H\left(X^T\right).$$

For the evaluation—where they model distributions rather than estimating them from data—they assume X^∞ and T^∞ to be i.i.d. and X_t to be independent of P_t ($t \in \mathbb{N}$). For experiments restricted to fixed demand series X^T, this measure differs from negative mutual information only by a constant. Hence although indication of absolut privacy levels differ between both measures, ranking different BBLH algorithms with conditional entropy and mutual information yields the same results. In our experiments we use $H(X|Y)$.

2.2.10 Removed Uncertainty. Depending on how reversible a BBLH algorithm is, it can be possible to reconstruct original values x_t from y_t. Let x_t^* denote a reconstruction, $\tilde{x}_t := x_t - x_t^*$ and $\tilde{y}_t := x_t - y_t$ for $t \in \{1, \dots, T\}$. Assuming that $\tilde{x}_t$ and $\tilde{y}_t$ are i.i.d. realizations of random variables $\tilde{X}, \tilde{Y}$, Laforet et al. [61] use the standard deviations $\sigma_{\tilde{X}}$ and $\sigma_{\tilde{Y}}$ of $\tilde{X}$ and $\tilde{Y}$ to calculate a privacy measure

$$RU\left(\tilde{X}, \tilde{Y}\right) := 1 - \sigma_{\tilde{X}}/\sigma_{\tilde{Y}}$$

called *removed uncertainty*. If reconstruction is successful (i.e., little privacy), the difference $x_t - x_t^*$ is small and RU takes values close to 1. Note that the formula is similar to the one of R_2^2 in Section 2.2.2. Hence removed uncertainty behaves similarly if one uses OLS to obtain predictions (x_t^* here or y_t^* in Section 2.2.2).

In [61], the authors propose to fit linear regression of y^T on x^T (RU^{r}) or to apply wavelet filtering to y^T, to obtain $(x^*)^T$ (RU^{w}).

2.2.11 Total Variation Distance. This is a distance measure of probability distributions. For two random variables U and V with $\Omega_U = \Omega_V$, the total variation distance, also dubbed *variational distance*, is

$$TVD(U,V) := \frac{1}{2} \sum_{u\in\Omega_U} |P_U(u) - P_V(u)|$$

The interpretation of this measure is similar to that of Kullback–Leibler Divergence (Section 2.2.6). Indeed, according to Pinsker's inequality [69] (Theorem 2.16) we have

$$2 \cdot TVD(U,V) \le \sqrt{KL(U\|V)}.$$

Kalogridis et al. [50] use $TVD(X,Y)$ and assume independence of time. Other work [11, 104] mentions but does not use this measure.

2.3 Other Measures

Many of the above privacy measures quantify the distance or degree of association of two time series. There exist many more approaches to do so. Crooks [24] alone describes more than 50 information-theoretic measures, Giusti at al. and Wang et al. [44, 93] list different (dis)similarity measures for time series. while this section contains an exhaustive list of measures applied to the BBLH setting thus far, one cannot by no means see it as a complete list of possibilities.

3 ENERGY STORAGE CHARACTERISTICS AND TECHNOLOGIES

The level of privacy achievable with BBLH algorithms depends on the characteristics of the storage system, such as the capacity or charging/discharging rate. In this section, we discuss the characteristics important for privacy protection in households at a qualitative level. We review and compare storage technologies for this application. While recommending a specific storage system requires considering many factors including user load and BBLH algorithm, our analysis already allows to rule out certain technologies. Finally, we list some currently available systems and their characteristics, which we use for our experiments in Section 6.

3.1 Characteristics of Energy Storage Systems

Many BBLH assessments assume arbitrary values, e.g., 2 kWh and 2 kW [101], for capacity and maximum (dis)charging rate of the storage system. Here, we discuss why one should carfully choose these values. We do so by closely looking at the practical requirements of using storage systems for privacy protection in households.

Capacity. The capacity of an energy storage system is a critical factor for hiding time-consuming loads and temporal dependencies in the load profile. Higher capacities can lead to higher privacy, but it can be unnecessary or not worthwhile to have a storage capacity above some limit, as shown in [8, 21, 100, 102]. In practice, however, one often chooses the storage capacity proportional to the yearly electrical energy consumption of a household [83]. This value strongly depends on the country, climate, type of residence, the number and consumption behavior of occupants. For instance, a residential consumer in the U.S. uses 10972 kWh annually on average [31], whereas a 3-person household in Germany consumes 4000 kWh per year [17]; 29.5 kWh and 11 kWh per day, respectively.

Charging/Discharging Rate. The maximum charging/discharging rate of the storage system also impacts the level of privacy achieved using BBLH algorithms [21]. Home appliances with high consumption, such as tumble dryers, heat pumps, water heaters and chargers for electric vehicles, can change the overall consumption pattern and cause sharp spikes in the load profile, leaking private information. For instance, a tumble dryer and a water heater can have a

consumption as high as 6 kW and 4.5 kW, respectively [78]. It might be necessary for the energy storage system to have a higher charging/discharging rate (power) than the demand of these appliances to be able to protect privacy of a private household, depending on the sampling rate of the smart meter's recordings.

Lifetime and Cycling Times. BBLH algorithms often mandate a high number of charging cycles of the storage system. Each storage system has a maximum number of useful cycles. This number usually indicates when a significant part of the capacity of the system is lost. Charging and discharging of a storage system can also be optimized, as not all charges and discharges have the same effect on lifetime. Although controlling the rate of charge/discharge in a lifetime-preserving manner can mitigate capacity loss [75], the effect cannot be neglected.

Round-Trip Efficiency. Round-trip efficiency is the ratio of the energy retrievable from a storage system to the energy given as input. The higher the round-trip efficiency, the lower is the cost of privacy protection. Protecting privacy using storage systems calls for a high round-trip efficiency, as the systems are discharged much more frequently compared to applications like backup power.

Cost. The cost of an energy storage system is undoubtedly significant as well. While the price of some storage technologies, such as lithium-ion batteries [90], has been plummeting drastically over the last decade, some are still too costly for private households.

3.2 Energy Storage Technologies

This section gives an overview of energy storage technologies from the perspective of their suitability for privacy protection with a focus on characteristics described above. While many reviews of storage technologies for various applications exist [26, 35, 47, 73], there is none for this perspective. For each technology we comment on the issues which arise when it is used for privacy protection. Besides the following storage technologies, there are many others which we do not address here because we deem them unsuitable for BBLH algorithms in households. We explain this in Appendix C.

3.2.1 Electrochemical Battery Energy Storage Systems. Electrochemical energy storage systems are a common choice for household applications [37]. This is mainly due to their low cost and widespread availability and despite reliability and lifetime concerns.

Lead-Acid Batteries. Lead-acid batteries are the oldest and most mature electrochemical batteries [26, 73, 103]. They offer several significant advantages such as low investment costs, availability, low self-discharge rates (less than 0.1% per day [26]), high efficiency, and ease of transport. But limited lifetime (1200–1800 cycles [103]), reliability issues, low energy and power densities are big drawbacks. High discharge rates and frequent deep discharge cycles further reduce their lifetime. Lead-acid batteries also suffer from sensitivity to ambient temperature. High temperatures ($> 45°C$) reduce their lifetime, low temperatures ($< -5°C$) reduce efficiency. This is due to a thermodynamic effect which increases the resistance in the electrolyte [29, 35]. According to Figgener et al. [37], lead-acid batteries are almost forced out of the market in developed countries.

Due to the limited lifetime of the lead-acid battery and its sensitivity to high discharge rates, it is not well-suited for privacy protection, despite being suggested in the literature [72]. BBLH algorithms often require frequent power cycling, in particular at a partial state of charge, i.e., when the battery is not fully charged. This can lead to premature failure of a lead-acid battery [70].

Lithium-Ion Batteries. Lithium-ion (Li-ion) batteries, commercially produced since the 1990s, have a higher energy density (energy per unit of volume), longer lifetime, higher charging and discharging rates, faster response, and higher cycle efficiency (up to 97% [73]) in comparison to lead-acid batteries. A Li-ion battery can be about one-quarter of the weight and one-half of the volume of a lead-acid battery with the same energy content. Capital costs of Li-ion batteries have plummeted drastically over the last years, falling below 80% of its price in 2010. This could continue to at least 50% of its 2018 price by 2030 and another 25% by 2040 [90]. Deep discharge of the battery can wear out the capacity. But one can avoid this by limiting the maximum rate of discharge.

There are different possible combinations of the materials for electrodes in Li-ion batteries, which can improve one or more characteristics of the technology, including energy density, lifetime, or charging/discharging rate. Among these different types of Li-ion batteries, Lithium Iron Phosphate batteries can be well-suited for privacy protection, because they can have more cycles in their lifetime, higher peak power and a more gradual loss of capacity.

Electric Vehicles (EV). Earlier Electric Vehicles (EV) used lead-acid or other electrochemical batteries, but nowadays almost all EVs are based on Li-ion batteries. Batteries in EVs are usually charged at times when people are home, as more than 80% of EV drivers charge their cars at their own house [2]. If the EV charging device does not have a separate meter (which is required by some grid operators [40]) EV batteries can potentially be used for privacy protection as well. Depending on the car model, today's EV batteries have a capacity between 6 and 100 kWh and power in the range of 2.5 to 40 kW—which is significant. However, using an EV battery to protect privacy has various limitations. The availability of the EV during the day is not guaranteed, and a BBLH algorithm would significantly reduce the lifetime of the EV.

3.2.2 Flow Battery Energy Storage Systems. In contrast to electrochemical batteries, which store the energy in the electrodes, flow batteries, also known as redox flow batteries, store energy in the electrolyte. An advantage of these batteries is that their capacity and power can be scaled independently from each other. The capacity scales with the amount of electrolyte stored in external tanks, while the power depends on the active area of the cell compartment. Therefore, these batteries can be customized in accordance with a household's needs. Flow batteries also have a long lifetime (in the order of 10,000 cycles) and no degradation with deep discharge or overcharge [103]. However, their market share for household applications is negligible [37]. This might be due to their relative complexity, high space requirements, and low efficiency (around 75-85% [19]). Here, we briefly describe two common types of flow batteries which can be used in households.

Vanadium Redox Flow Battery (VRB). VRB is the most mature type of flow batteries. A VRB cell has an efficiency of up to ~85% [19], much lower than the efficiency of the Li-ion batteries. One other

drawback of the VRB is its low power and energy density, requiring considerable space. Vanadium Bromide redox flow batteries, also known as the "Generation 2" VRB, have a higher energy density, but a decreased lifetime [89], an essential characteristic for implementing BBLH algorithms.

Zinc Bromium Flow Battery (ZnBr). ZnBr batteries are not as mature as VRBs. Nevertheless, utility-scale and even small-scale ZnBr batteries for households are now available on the market (Table 4). They have higher energy density in comparison to VRB. However, these batteries require an additional pump for the circulation of bromine complexes [26]. Like with many other technologies, the lifetime of ZnBr batteries has improved significantly over the years.

3.2.3 Comparison. The final choice of a storage system for privacy protection requires comprehensive simultaneous evaluation together with the choice of a BBLH algorithm (cf. Section 7), which is beyond the scope of this paper. We can however qualitatively compare the reviewed technologies according to their characteristics described, see Table 3. From our review, one specific chemistry of Li-ion batteries—the Lithium Iron Phosphate batteries—seems to be a better candidate for privacy protection than the others. This is because they can provide high peak power to cover spikes in the load profile, have a higher lifetime and lower degradation with more cycles. Redox flow batteries can in some cases provide a higher number of cycling times than Li-ion batteries but at a lower round trip efficiency. There is a need for energy storage systems which can provide more cycles with fewer concerns for their lifetimes.

Table 4 shows a comparison of storage systems currently available on the market. All Li-ion batteries in this table are lithium iron phosphate batteries. We provide efficiencies and costs considering both battery and inverter. In the case of the LG Chem and BYD batteries which are sold without an inverter, we assume an inverter efficiency of 95.3% (efficiency of SMA-SB-240 inverter [97]). Next, we assume the cost of an inverter to be half the price for a battery with the same rating.

4 BBLH ALGORITHMS

In our experiments we use five BBLH algorithms which are the most prominent ones in literature. They specify the logic of setting y_t given x_t and storage characteristics.

4.1 Best-Effort

The idea of the best-effort algorithm (BE1) [53], also known as the water-filing algorithm [54], is to keep the grid load constant for as long as possible, given the constraints implied by the storage system. At each point in time t, BE1 chooses the rate $b_t = x_{t-1} - x_t$ of charge/discharge to compensate for the change in user load if the remaining storage capacity and maximal charging/discharging rate allow to do so. Otherwise, BE1 charges/discharges the storage system at the maximal rate. See [53] for the formal description.

Yang et al. [100] propose a slightly different variant of this algorithm (BE2). There, when maintaining the required charging/-discharging rate is not possible given storage capacity and its current state of charge, it is set to $y_t = x_t$.

4.2 Non-Intrusive Load Leveling (NILL)

By default, in a *stable state*, the NILL algorithm [72] uses a battery to keep the external load stable at the value y_{st}. It does so by drawing power from the battery to supply the appliances when the power demand is higher than y_{st} and by charging the battery from the smart grid when the demand is lower than y_{st}. When the battery is relatively empty or full, the algorithm enters the so-called *low recovery state* or *high recovery state* with the new target load values of the grid y_l and y_h, respectively. The value y_l is equal to the maximal affordable charging rate, while y_h is set below (1–5A in [72] or 0.5A in [100]) the most recent measured electrical current (x_t/V, if x_t is power and V is the grid voltage). After certain conditions are satisfied, such as the charge of the battery reaching a certain level or the current dropping sufficiently, the algorithm returns to the stable state. The original paper [72] does not account for the case when the difference between y_t and x_t is greater than the maximal battery charging/discharging rate. Yang et al. [100] describe the algorithm more completely. Our implementation of NILL is based on their paper, incorporating additional actions to avoid errors, which would have occurred with batteries with relatively low capacities.

4.3 Stepping Algorithms

Stepping algorithms [100] set the grid load to multiples $y_t = h_t \cdot \beta$ of the maximal (dis-)charging rate β of the storage system. The factor h_t is either $\lceil x_t/\beta \rceil$ or $\lfloor x_t/\beta \rfloor$. Let s_t be the *charging signal*. I.e., the battery is charging if $s_t = 1$ and discharging if $s_t = 0$. Stepping algorithms are defined by the logic to control this signal.

Lazy Stepping. This algorithm aims at keeping y_t constant. This fails if (1) the battery is full and cannot maintain the previous external load without being overcharged. In this case it will discharge. (2) If the battery is empty and cannot keep up the external load any more, it will charge. (3) If the load demand is so low or so high that $|y_t - x_t| > \beta$, the battery charges if its less than half full and discharges otherwise with Algorithm (LS1). With (LS2), s_t randomly takes value 0 or 1 in this case.

Lazy Charging (LC). The lazy charging algorithm keeps s_t constant whenever possible: If the battery is full, it provides electricity until it is empty. Then it will fully charge again. An advantage of this algorithm is the reduced number of charging cycles.

Random Charging (RC). Here, s_t does not depend on s_{t-1}. The probability of (dis)charging the storage is $P[s_t = 0] = SOC_t/C$, where SOC_t is the state of charge of the battery and C its capacity.

5 LOAD PROFILES

For our experiments we form 37 load profiles from four different sources listed in Table 5. We achieve a high diversity concerning (1) sampling rate, (2) number of residents, (3) day of the week (working day or weekend), (4) employment status, (5) location and (6) time of the year. To obtain different sampling rates, we aggregate high-resolution load profiles. Some data sets have missing measurements. We impute these by repeating the last available measurement. For each load profile we add seven preceding days. We use them for BBLH algorithms to make the result less dependent on initial parameters, like state of charge of the storage, but exclude

Storage Technology	Energy Density	Power Density	Number of Cycles	Round-trip Efficiency	Degradation	Costs	Applicable
Lead-Acid Battery	−	++	−−	++	−−	+ + +	No
Lithium-ion Batteries	+ + +	+ + +	+ + +	++	−	++	Yes
Vanadium Redox Flow Battery (VRB)	−	−−	+ + +	−	+	++	Yes
Zinc Bromium (ZnBr) Batteries	+	+	+ + +	−	+	++	Yes

Table 3: Comparison of storage technologies suitable for privacy protection.

Type	Manufacturer	Model	C (kWh)	CR/DR (kW)	RE (%)	Lifetime (cycles)	ECC (€/kWh)	CpC (cent/kWh)	Price (€)
Li-ion	Tesla[b]	PowerWall 2	13.5	4.6	90	3200	385	16	6850
	Enphase[c]	AC Battery	1.20	0.27	90	3650	1417	39	1700
	Sonnen-Batterie[d]	eco 8.0/4	4	2	89	10000	1949	19.4	7795
		eco 8.0/6	6	3.3	89	10000	1467	14.96	8800
	Solarwatt[e]	Myreserve Pack	2.4	4	92	N/A	729	20	1750
	LG chem[f]	RESU 3.3	3.3	3.6	85.5	6000	1636	27.7	2450
		RESU 6.5	6.5	5	85.5	6000	1060	17.9	6890
		RESU 10	10	5	85.5	6000	7640	12.7	7640
	BYD[g]	B-Box 2.5	2.45	2.5	85.5	6000	954	15.9	2238
		B-Box 5	4.9	5	85.5	6000	815	13.60	3995
		B-Box 7.5	7.35	7.5	85.5	6000	821	13.77	6073
		B-Box 10.0	9.8	10	85.5	6000	798	13.31	7830
VRB	VoltStorage[h]	Smart	6.8	1.5	74.6	10000	882	8.82	6000
Zinc Bromiom	Redflow[i]	Zcell	10	3	75.2	3650	760	20.82	7600
	Schmid EverFlow[j]	Compact	15	5	N/A	≥10000	N/A	N/A	N/A

[b] https://www.tesla.com/de_DE/powerwall?redirectno
[c] https://enphase.com/en-au/products-and-services/storage
[d] https://sonnen.de/stromspeicher/sonnenbatterie-eco/
[e] https://www.solarwatt.com/solar-batteries/myreserve
[f] https://www.lgchem.com/product/PD00000149
[g] https://en.byd.com/energy/b-box-ess/
[h] https://voltstorage.com/en/voltstorage-smart-home-battery/
[i] https://redflow.com/products/redflow-zbm2/
[j] https://schmid-group.com/en/business-units/energy-systems/everflow-energy-storage-solutions/compact-storage

Table 4: Comparison of energy storage technologies. C — capacity, CR/DR — charging/discharging rate, RE — roundtrip efficiency, ECC — energy capital cost, CpC — cost per cycle.

Name	Used in	Resolution
REDD [59]	[32, 34, 36, 61, 85, 101, 102, 104]	1s
CER [39]	[22, 57, 61, 62]	30m
ECO [13]	[9]	1s
Smart* [12]	-	1s

Table 5: Data sources used in experiments.

them when computing privacy measures. The lengths of resulting load profiles are 1–14 (+7) days with sampling rates of {30, 60, 900,-1800}. This is similar to the ranges used by many of the papers mentioned in Section 2 – see Tables 11 and 12 in Appendix D. Table 15 in Appendix F summarizes load profiles we used.

6 EXPERIMENTS

This section contains our experimental setting and results. We study to which extent (1) privacy measures, (2) load profiles and (3) storage characteristics affect relative performance of BBLH algorithms.

6.1 Experimental Setting

The algorithms studied here require specification of the battery capacity, charging/discharging power and initial state of charge. We chose the latter to always be 0.5. We experiment with four sets of capacity, charging and discharging power values, corresponding to 'AC Battery', 'eco 8.0/4', 'Myreserve Pack' and 'RESU 6.5', see Table 4. We apply the five algorithms from Section 4, including two versions of the BE and the LS algorithm, to 37 load profiles (cf. Section 5). For each pair of user load and respective grid load we compute the level of privacy. To this end, we use 11 privacy measures introduced in Section 2.2 and listed in Table 2. If the measure was applied to undifferenced or differenced load profiles (or both) in previous research, we do the same. As mentioned before, we adapt K-divergence and relative entropy for our experimental setting as $K(X\|Y)$ and $H(X\,|Y)$ respectively. We use different variants of entropy ratio (ER^{z} and ER^{nz}), coefficient of determination (R_2^2 and R_p^2), feature mass (FM_r, FM and FM_{ed}), mutual information (MI^{i}, MI_s, MI^{m} and MI^{b}) and removed uncertainty (RU^{r} and RU^{w}). This

Measure	Δ	BE1	BE2	LC	LS1	LS2	NILL	RC
R_2^2	y	2	3	7	4	5	6	1
R_2^2	n	5	7	2	3	6	4	1
MI^{b}	y	2	1	7	6	5	4	3
MI^{i}	y	4	3	6	2	1	7	5
MI_s	y	5	3	6	1	2	4	7
min rank		1	1	1	1	1	1	1
max rank		7	7	7	6	7	7	7

Table 6: Average algorithm ranking with respect to privacy measures (excerpt).

Load profile	BE1	BE2	LC	LS1	LS2	NILL	RC
CER5	3	5	6	4	2	7	1
ECO5	5	2	3	4	6	1	7
REDD5	6	5	3	2	4	1	7
SmartB2	4	2	6	1	3	7	5
min rank	2	1	1	1	1	1	1
max rank	7	7	7	5	6	7	7

Table 7: Average algorithm ranking with respect to load profiles (excerpt).

results in 25 different privacy measures and $37 \cdot 25 \cdot 7 \cdot 4 = 25900$ experiments in total. For the measures which require quantization of load profiles we use 20 equidistant bins as suggested in [21, 22, 101].

Privacy measures take values in different ranges. Some are frequently normalized to bring them to within $[0, 1]$ (see, e.g., [58] for mutual information). However, such a normalization is not straightforward for other measures like load variance. To overcome this issue and gain comparability, we assign ranks $\{1, \ldots, 7\}$ to the algorithms in accordance with the values of privacy measures for each combination (load profile, battery characteristics, privacy measure). Rank 1 corresponds to the best privacy level, rank 7 to the lowest.

6.2 Results

For our results, we build summary tables as follows. Given the factor in question (e.g., privacy measures) we compute the average rank of each algorithm over the remaining factors (load profiles/storage characteristics), and rank the algorithms again according to this average rank. A new rank 1 corresponds to the best, 7 to the worst *average* performance. Due to limited space show only representative results here. Complete results are in Appendix E.

Table 6 shows the results for different privacy measures (using the same abbreviations as in Table 13). The column "Δ" indicates whether a measure is applied to differenced load profiles. Note that even different variants of the same measure (MI^{b}, MI_s or MI^{i}), or the change from differenced to undifferenced load profiles (R_2^2) leads to different rankings. To some extent, this is in line with the findings in [100, 102] (which consider only MI), but the NILL algorithm is not a clear outsider in our experiments. The last two rows show the lowest and the highest average rank of the algorithms.

Table 7 highlights the impact of different data sets. Both Table 6 and Table 7 indicate that almost every algorithm can be best or worst, depending on the chosen load profile and privacy measure.

Table 8 lists results with respect to the battery characteristics. One can see that they also have a significant influence one the relative performance of the algorithms in terms of privacy achieved.

Our findings so far demonstrate that all three factors significantly affect the comparison results. Since there is currently no agreement on which load profiles, storage and privacy measures constitute a "gold standard" for BBLH, no fair relative evaluation of BBLH algorithms is possible at this point. Hence it is crucial to agree on a procedure to assess the quality of BBLH algorithms.

Storage	BE1	BE2	LC	LS1	LS2	NILL	RC
AC Battery	1	2	7	3	4	6	5
eco 8.0/4	7	4	5	1	2	3	6
Myreserve Pack	7	6	3	2	1	4	5
RESU 6.5	7	5	4	3	2	1	6

Table 8: Average algorithm ranking with respect to battery characteristics.

7 DISCUSSION

In our experiments in Section 6, we followed the common procedure to evaluate BBLH algorithms. But since reality is even more complicated than our extensive evaluation framework, some explanations of the inherent simplifications of this evaluation procedure are due. Furthermore, we will discuss costs implied by BBLH and their potential impact on the utility of smart metering; one should not disregard these factors when estimating the quality of BBLH algorithms. Finally, we reason about possible methodologies to agree on load profiles and privacy measures for BBLH evaluation.

Simplifications. In our experiments we applied the same simplification as related research papers: We assume the sampling rate of a smart meter to coincide with the frequency of a BBLH algorithm taking actions. This is a strong assumption since there is no reason for these two rates to be equal. Allowing BBLH to operate at different frequencies may affect the outcome significantly. Next, active power levels are not the only output of smart meters. They also measure reactive power, power factor, voltage and current harmonics [84]. All this can reveal private information as well. Current harmonics, for instance, can indicate types of appliances [4].

Utility of Smart Meters and Costs of BBLH. Even a BBLH algorithm which provides exceptional privacy can be infeasible if it comes at high costs or curbs the utility of smart metering. One should account for this when assessing BBLH algorithms. Reasons given for smart metering include flexible pricing, which encourages consumers to shift demand to off-peak hours [11, 16, 20, 60, 63, 71, 74, 85, 94, 98–101, 104], precise load forecasts [11, 16, 60, 79, 94, 102], support for alternative energy sources [41, 43, 50, 52, 54, 72, 87] and others [8, 18, 21, 22, 30, 36, 38, 53, 57, 62, 64, 65, 67, 91, 92]. Since BBLH affects consumption patterns, it will likely affect the utility of smart meters as well. For instance, nondeterministic algorithms like random charging may decrease the accuracy of load forecasts, while load leveling improves it [23].

Additional costs for consumers who apply BBLH are quite diverse. Obvious costs are associated with buying, maintaining and disposal of the storage system. Table 4 lists the prices for several storage technologies. As mentioned before, imperfect round-trip efficiencies of storage leads to energy loss and respective costs. BBLH algorithms which make use of the storage with high intensity lead to higher purchase and maintenance costs as well as costs due to energy losses. To take these costs into account, we propose to use a realistic simulation of a storage system, modelling its efficiency, aging etc., when evaluating BBLH algorithms. In addition to privacy protection, storage systems can be used for demand response [95]: They store energy when electricity is cheap and discharge it when prices are high. Sacrificing this option for privacy increases opportunity costs. Finally, widespread use of BBLH may result in unpredictable changes of aggregated regional consumption, which in turn could affect electricity prices.

There are several papers considering privacy together with utility [23, 34, 79] or costs [60, 85, 98, 99]. But they are all limited to picking one or few combinations of (1) privacy measures (2) load profiles (3) battery characteristics (4) cost sources and (5) utility measures. Our study has shown that such restrictions have a crucial effect on the result, limiting the utility of these combined analyses.

User Load Data. Our experiments highlight that the choice of data sets to test BBLH algorithms has a big impact on the resulting privacy level. To achieve meaningful comparability, it is imperative to arrive at a commonly used benchmark data set for this scenario. This data set would need to be *diverse* to represent all types of user behaviour, *balanced* not to favour abnormal behaviour and *extensive* to assess privacy levels over long periods of time, where adverse effects of habits/recurring behavioural patterns can be detected. It might also be sensible to assess privacy levels with several different data sets if these accurately depict different categories of users or differently developed energy infrastructures. Such a range of data sets would provide more accurate orientation to choose the best algorithm in each specific situation. Note, however, that in choosing an algorithm taylored to a specific situation, the choice for this algorithm itself might disclose sensitive information.

Choice of Privacy Measure. Different storage characteristics and different types of user loads are inherently present in the real world and impact the level of privacy resulting from a BBLH algorithm. In contrast, the choice of a privacy measure does not have any effect on the actual privacy achieved by BBLH, only on its assessment. Hence the challenge here is very different from those we still face regarding different storage characteristics and testing data. The goal is to arrive at a meaningful privacy measure which accurately quantifies what is intuitively perceived as privacy. As of yet, BBLH research largely seems to skip the vital step of assessing privacy measures, comparing their theoretical properties and interpret the real-world implications of their outputs for the BBLH scenario.

The first and already non-trivial step towards meaningful privacy measures is a rigorous definition of what privacy and private information actually entails. Does the user want to hide specific occurrences (like what they did on a specific Tuesday afternoon), or ongoing lifestyle choices (e.g., never cooking at home), or both? Some privacy measures, like Kullback-Leibler divergence or cluster similarity, for example, do not depend on the joint distribution of user and grid loads and are therefore clearly tailored towards hiding lifestyle choices, habits and fixed characteristics of consumption behavior. These measures will indicate no privacy at all if user and grid load follow the same distribution, even if they are entirely independent of one another, i.e., hide specific occurrences well.

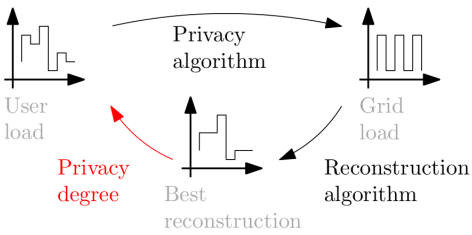

Figure 1: Privacy assessment considering reconstructability.

Regardless of the definition of private information, there are several properties a meaningful privacy measure should always satisfy. Let us illustrate this with the example of reconstructability or robustness to post-processing. Many privacy measures from Section 2 assess a degree of dissimilarity between user and grid load, without taking into account how reversible the BBLH algorithm is, i.e., how well an adversary might be able to reconstruct the user from the grid load. However, we stipulate that a meaningful privacy measure should take reconstructability into account (cp. Figure 1). This means that, if no additional information is needed to get from the grid load to a reconstruction of the user load, a privacy measure should not output a lower degree of privacy between reconstruction and user load than between the grid load and the user load. For a more formal grasp on this property, see Appendix B. Reconstructability, however, is only one property a meaningful privacy measure should have to assess BBLH algorithms and smart meter privacy in general. Identifying a sufficiently complete set of required properties and developing privacy measures which satisfy them is a challenging but ultimately necessary step towards comparable and provable privacy in the context of smart metering.

8 CONCLUSIONS

Smart meters are considered essential in future energy grids. But their frequent measurements pose significant risks to consumer privacy. Battery-based load hiding (BBLH) can mask the actual electricity demand, by charging and discharging an energy storage system. Although a plethora of BBLH algorithms exists, it remains unclear whether some of them are better than others. This is due to a significant lack of understanding on how different factors – (1) load profiles (2) storage characteristics and (3) choice of privacy measure – affect the result of comparisons of BBLH algorithms.

In this paper, we contribute to closing this gap. Firstly, we systematically reviewed privacy measures proposed for BBLH as well as currently available storage technologies. From this review we already obtained valuable insights. For instance, there are several approaches to calculate the same privacy measure which do not agree with its accepted definition (e.g., coefficient of determination), and lead-acid batteries are not well suited for BBLH. Next we conducted an experimental study, evaluating several prominent BBLH algorithms with a variety of load profiles and storage characteristics. We compared the achieved level of privacy with different measures.

Our results suggest that all three factors have a crucial effect on the relative performance of BBLH algorithms. Hence without an agreement on the proper way to evaluate BBLH algorithms, their further development may have limited usefulness. We made our implementation of BBLH algorithms and privacy measures freely available along with the code to reproduce our experiments. Finally, we discussed other factors which must be taken into account when evaluating BBLH algorithms and establishing a theoretical basis for choosing a suitable privacy measure.

ACKNOWLEDGMENTS

This work was supported by the German Research Foundation (DFG) as part of the Research Training Group GRK 2153: Energy Status Data - Informatics Methods for its Collection, Analysis and Exploitation. We thank Pavel Obraztcov for suggesting and summarizing the textbooks we refer to in Section 2.2.2.

REFERENCES

[1] 2007. Energy Independence and Security Act of 2007. Pub. L. No. 110-140, 121 Stat. 1792.

[2] 2015. *Plugged In: How Americans Charge Their Electric Vehicles.* Technical Report. Idaho National Laboratory. 1–24 pages. https://avt.inl.gov/sites/default/files/pdf/arra/SummaryReport.pdf

[3] 2016. Gesetz über den Messstellenbetrieb und die Datenkommunikation in intelligenten Energienetzen (Messstellenbetriebsgesetz - MsbG). Bl. I S. 2034, §60.

[4] Kofi Agyeman, Sekyung Han, and Soohee Han. 2015. Real-time recognition non-intrusive electrical appliance monitoring algorithm for a residential building energy management system. *Energies* 8, 9 (2015), 9029–9048.

[5] Takeshi Amemiya. 1985. *Advanced econometrics.* Harvard university press.

[6] Mathew Aneke and Meihong Wang. 2016. Energy storage technologies and real life applications–A state of the art review. *Applied Energy* 179 (2016), 350–377.

[7] Olatz Arbelaitz, Ibai Gurrutxaga, Javier Muguerza, Jesús M. Pérez, and Iñigo Perona. 2013. An extensive comparative study of cluster validity indices. *Pattern Recognition* 46, 1 (2013), 243–256. https://doi.org/10.1016/j.patcog.2012.07.021

[8] Miguel Arrieta and Inaki Esnaola. 2017. Smart meter privacy via the trapdoor channel. In *2017 IEEE International Conference on Smart Grid Communications, SmartGridComm 2017, Dresden, Germany, October 23-27, 2017.* 277–282. https://doi.org/10.1109/SmartGridComm.2017.8340722

[9] Ramana R. Avula, Tobias J. Oechtering, and Daniel Mansson. 2018. Privacy-preserving smart meter control strategy including energy storage losses. In *2018 IEEE PES Innovative Smart Grid Technologies Conference Europe, ISGT-Europe 2018, Sarajevo, Bosnia and Herzegovina, October 21-25, 2018.* 1–6. https://doi.org/10.1109/ISGTEurope.2018.8571537

[10] AC Ayvazyan and BC Mkhitaryan. 1998. *Applied statistics and basics of econometrics.* Unity, (In Russian).

[11] Michael Backes and Sebastian Meiser. 2013. Differentially Private Smart Metering with Battery Recharging. In *Data Privacy Management and Autonomous Spontaneous Security - 8th International Workshop, DPM 2013, and 6th International Workshop, SETOP 2013, Egham, UK, September 12-13, 2013, Revised Selected Papers.* 194–212. https://doi.org/10.1007/978-3-642-54568-9_13

[12] Sean Barker, Aditya Mishra, David Irwin, Emmanuel Cecchet, Prashant Shenoy, Jeannie Albrecht, et al. 2012. Smart*: An open data set and tools for enabling research in sustainable homes. *SustKDD, August* 111, 112 (2012), 108.

[13] Christian Beckel, Wilhelm Kleiminger, Romano Cicchetti, Thorsten Staake, and Silvia Santini. 2014. The ECO data set and the performance of non-intrusive load monitoring algorithms. In *Proceedings of the 1st ACM Conference on Embedded Systems for Energy-Efficient Buildings, BuildSys 2014, Memphis, TN, USA, November 3-6, 2014.* 80–89. https://doi.org/10.1145/2674061.2674064

[14] Andreas Becker, Hauke Loges, Stefan Kippelt, Alexander Gitis, Ghada Merei, David Echternacht, Marcus Müller, Alexander Zeh, Martin Kleimaier, Matthias Leuthold, et al. 2015. Electricity storage systems in medium-and low-voltage networks. In *International ETG Congress 2015; Die Energiewende-Blueprints for the new energy age.* VDE, 1–8.

[15] Charles W Brokish and Michele Lewis. 1997. A-Law and mu-Law companding implementations using the TMS320C54x. *SPRA163* (1997).

[16] Erik Buchmann, Klemens Böhm, Thorben Burghardt, and Stephan Kessler. 2013. Re-identification of Smart Meter data. *Personal and Ubiquitous Computing* 17, 4 (2013), 653–662. https://doi.org/10.1007/s00779-012-0513-6

[17] Bundeskartellamt Bundesnetzagentur. 2017. Monitoringbericht 2017. Bonn.

[18] Ann Cavoukian, Jules Polonetsky, and Christopher Wolf. 2010. Smartprivacy for the smart grid: embedding privacy into the design of electricity conservation. *Identity in the Information Society* 3, 2 (2010), 275–294.

[19] Haisheng Chen, Thang Ngoc Cong, Wei Yang, Chunqing Tan, Yongliang Li, and Yulong Ding. 2009. Progress in electrical energy storage system: A critical review. *Progress in natural science* 19, 3 (2009), 291–312.

[20] Zhi Chen and Lei Wu. 2013. Residential appliance DR energy management with electric privacy protection by online stochastic optimization. *IEEE Transactions on Smart Grid* 4, 4 (2013), 1861–1869.

[21] Jun-Xing Chin, Tomas Tinoco De Rubira, and Gabriela Hug. 2017. Privacy-Protecting Energy Management Unit Through Model-Distribution Predictive Control. *IEEE Trans. Smart Grid* 8, 6 (2017), 3084–3093. https://doi.org/10.1109/TSG.2017.2703158

[22] Jun-Xing Chin, Giulio Giaconi, Tomas Tinoco De Rubira, Gabriela Hug, et al. 2018. Considering Time Correlation in the Estimation of Privacy Loss for Consumers with Smart Meters. In *2018 Power Systems Computation Conference (PSCC).* IEEE, 1–7.

[23] Jun-Xing Chin, Thierry Zufferey, Etta Shyti, and Gabriela Hug. 2019. Load Forecasting of Privacy-Aware Consumers. In *2019 IEEE Milan PowerTech.* IEEE, 1–6.

[24] Gavin E Crooks. 2017. On measures of entropy and information. *Tech. Note* 9 (2017), v4.

[25] Russell Davidson, James G MacKinnon, et al. 1993. Estimation and inference in econometrics. *OUP Catalogue* (1993).

[26] Francisco Díaz-González, Andreas Sumper, Oriol Gomis-Bellmunt, and Roberto Villafáfila-Robles. 2012. A review of energy storage technologies for wind power applications. *Renewable and sustainable energy reviews* 16, 4 (2012), 2154–2171.

[27] Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. 2006. Our Data, Ourselves: Privacy Via Distributed Noise Generation. In *Advances in Cryptology - EUROCRYPT 2006, 25th Annual International Conference on the Theory and Applications of Cryptographic Techniques, St. Petersburg, Russia, May 28 - June 1, 2006, Proceedings.* 486–503. https://doi.org/10.1007/11761679_29

[28] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam D. Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *Theory of Cryptography, Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4-7, 2006, Proceedings.* 265–284. https://doi.org/10.1007/11681878_14

[29] S Eckroad and I Gyuk. 2003. *EPRI-DOE handbook of energy storage for transmission & distribution applications.* Technical Report. 3–35 pages.

[30] Costas Efthymiou and Georgios Kalogridis. 2010. Smart grid privacy via anonymization of smart metering data. In *2010 First IEEE International Conference on Smart Grid Communications.* IEEE, 238–243.

[31] US EIA. 2019. How much electricity does an American home use.

[32] Günther Eibl and Dominik Engel. 2015. Influence of Data Granularity on Smart Meter Privacy. *IEEE Trans. Smart Grid* 6, 2 (2015), 930–939. https://doi.org/10.1109/TSG.2014.2376613

[33] Dominik Engel. 2013. Wavelet-based load profile representation for smart meter privacy. In *IEEE PES Innovative Smart Grid Technologies Conference, ISGT 2013, Washington, DC, USA, February 24-27, 2013.* 1–6. https://doi.org/10.1109/ISGT.2013.6497835

[34] Murat A Erdogdu, Nadia Fawaz, and Andrea Montanari. 2015. Privacy-utility trade-off for time-series with application to smart-meter data. In *Workshops at the Twenty-Ninth AAAI Conference on Artificial Intelligence.*

[35] Faramarz Faraji, Abbas Majazi, Kamal Al-Haddad, et al. 2017. A comprehensive review of flywheel energy storage system technology. *Renewable and Sustainable Energy Reviews* 67 (2017), 477–490.

[36] Farhad Farokhi and Henrik Sandberg. 2018. Fisher Information as a Measure of Privacy: Preserving Privacy of Households With Smart Meters Using Batteries. *IEEE Trans. Smart Grid* 9, 5 (2018), 4726–4734. https://doi.org/10.1109/TSG.2017.2667702

[37] Jan Figgener, David Haberschusz, Kai-Philipp Kairies, Oliver Wessels, Benedikt Tepe, and Dirk Uwe Sauer. 2018. Wissenschaftliches Mess-und Evaluierungsprogramm Solarstromspeicher 2.0-Jahresbericht 2017. *ISEA Institut für Stromrichtertechnik und Elektrische Antriebe RWTH Aachen: Aachen, Germany* (2018).

[38] Sören Finster and Ingmar Baumgart. 2015. Privacy-Aware Smart Metering: A Survey. *IEEE Communications Surveys and Tutorials* 17, 2 (2015), 1088–1101. https://doi.org/10.1109/COMST.2015.2425958

[39] Commission for Energy Regulation. 2011. Electricity Smart Metering Customer Behaviour Trials (CBT) Findings Report.

[40] Forum Netztechnik / Netzbetrieb im VDE (FNN). 2016. *Anschluss und Betrieb von Speichern am Niederspannungsnetz.* Technical Report. Forum Netztechnik / Netzbetrieb im VDE (FNN), Berlin.

[41] Giulio Giaconi, Deniz Gündüz, and H. Vincent Poor. 2015. Smart meter privacy with an energy harvesting device and instantaneous power constraints. In *2015 IEEE International Conference on Communications, ICC 2015, London, United Kingdom, June 8-12, 2015.* 7216–7221. https://doi.org/10.1109/ICC.2015.7249478

[42] Giulio Giaconi, Deniz Gündüz, and H. Vincent Poor. 2018. Privacy-Aware Smart Metering: Progress and Challenges. *IEEE Signal Process. Mag.* 35, 6 (2018), 59–78. https://doi.org/10.1109/MSP.2018.2841410

[43] Giulio Giaconi, Deniz Gündüz, and H. Vincent Poor. 2018. Smart Meter Privacy With Renewable Energy and an Energy Storage Device. *IEEE Trans. Information Forensics and Security* 13, 1 (2018), 129–142. https://doi.org/10.1109/TIFS.2017.2744601

[44] Rafael Giusti and Gustavo E. A. P. A. Batista. 2013. An Empirical Comparison of Dissimilarity Measures for Time Series Classification. In *Brazilian Conference on Intelligent Systems, BRACIS 2013, Fortaleza, CE, Brazil, 19-24 October, 2013.* 82–88. https://doi.org/10.1109/BRACIS.2013.22

[45] Jesús Gómez-Vilardebó and Deniz Gündüz. 2013. Privacy of smart meter systems with an alternative energy source. In *Proceedings of the 2013 IEEE International Symposium on Information Theory, Istanbul, Turkey, July 7-12, 2013.* 2572–2576. https://doi.org/10.1109/ISIT.2013.6620691

[46] William H Greene. 2003. *Econometric analysis.* Pearson Education India.

[47] Ioannis Hadjipaschalis, Andreas Poullikkas, and Venizelos Efthimiou. 2009. Overview of current and future energy storage technologies for electric power applications. *Renewable and sustainable energy reviews* 13, 6-7 (2009), 1513–1522.

[48] Fumio Hayashi. 2000. Econometrics. *Princeton University Press* (2000).

[49] E Jannelli, M Minutillo, A Lubrano Lavadera, and G Falcucci. 2014. A small-scale CAES (compressed air energy storage) system for stand-alone renewable energy power plant for a radio base station: A sizing-design methodology. *Energy* 78 (2014), 313–322.

[50] Georgios Kalogridis, Rafael Cepeda, Stojan Z. Denic, Tim A. Lewis, and Costas Efthymiou. 2011. ElecPrivacy: Evaluating the Privacy Protection of Electricity

Management Algorithms. *IEEE Trans. Smart Grid* 2, 4 (2011), 750–758. https://doi.org/10.1109/TSG.2011.2160975

[51] Georgios Kalogridis and Stojan Z. Denic. 2011. Data Mining and Privacy of Personal Behaviour Types in Smart Grid. In *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on, Vancouver, BC, Canada, December 11, 2011*. 636–642. https://doi.org/10.1109/ICDMW.2011.58

[52] Georgios Kalogridis, Stojan Z. Denic, Tim A. Lewis, and Rafael Cepeda. 2011. Privacy protection system and metrics for hiding electrical events. *IJSN* 6, 1 (2011), 14–27. https://doi.org/10.1504/IJSN.2011.039630

[53] Georgios Kalogridis, Costas Efthymiou, Stojan Z Denic, Tim A Lewis, and Rafael Cepeda. 2010. Privacy for smart meters: Towards undetectable appliance load signatures. In *2010 First IEEE International Conference on Smart Grid Communications*. IEEE, 232–237.

[54] Georgios Kalogridis, Zhong Fan, and Sagar Basutkar. 2011. Affordable privacy for home smart meters. In *2011 IEEE Ninth International Symposium on Parallel and Distributed Processing with Applications Workshops*. IEEE, 77–84.

[55] Shahab Karrari, Mathias Noe, and Joern Geisbuesch. 2018. High-speed Flywheel Energy Storage System (FESS) for Voltage and Frequency Support in Low Voltage Distribution Networks. In *2018 IEEE 3rd International Conference on Intelligent Energy and Power Systems (IEPS)*. IEEE, 176–182.

[56] Leonard Kaufman and Peter J Rousseeuw. 2009. *Finding groups in data: an introduction to cluster analysis*. Vol. 344. John Wiley & Sons.

[57] Stephan Kessler, Erik Buchmann, and Klemens Böhm. 2015. Deploying and Evaluating Pufferfish Privacy for Smart Meter Data. In *2015 IEEE 12th Intl Conf on Ubiquitous Intelligence and Computing and 2015 IEEE 12th Intl Conf on Autonomic and Trusted Computing and 2015 IEEE 15th Intl Conf on Scalable Computing and Communications and Its Associated Workshops (UIC-ATC-ScalCom), Beijing, China, August 10-14, 2015*. 229–238. https://doi.org/10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.55

[58] Shiraj Khan, Sharba Bandyopadhyay, Auroop R Ganguly, Sunil Saigal, David J Erickson III, Vladimir Protopopescu, and George Ostrouchov. 2007. Relative performance of mutual information estimation methods for quantifying the dependence among short and noisy data. *Physical Review E* 76, 2 (2007), 026209.

[59] J Zico Kolter and Matthew J Johnson. 2011. REDD: A public data set for energy disaggregation research. In *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*, Vol. 25. 59–62.

[60] Jinkyu Koo, Xiaojun Lin, and Saurabh Bagchi. 2012. PRIVATUS: Wallet-Friendly Privacy Protection for Smart Meters. In *Computer Security - ESORICS 2012 - 17th European Symposium on Research in Computer Security, Pisa, Italy, September 10-12, 2012. Proceedings*. 343–360. https://doi.org/10.1007/978-3-642-33167-1_20

[61] Fabian Laforet, Erik Buchmann, and Klemens Böhm. 2015. Individual privacy constraints on time-series data. *Inf. Syst.* 54 (2015), 74–91. https://doi.org/10.1016/j.is.2015.06.006

[62] Fabian Laforet, Erik Buchmann, and Klemens Böhm. 2016. Towards provable privacy guarantees using rechargeable energy-storage devices. In *Proceedings of the Seventh International Conference on Future Energy Systems, Waterloo, ON, Canada, June 21 - 24, 2016*. 7:1–7:14. https://doi.org/10.1145/2934328.2934335

[63] Simon Li, Ashish Khisti, and Aditya Mahajan. 2016. Privacy-optimal strategies for smart metering systems with a rechargeable battery. In *2016 American Control Conference, ACC 2016, Boston, MA, USA, July 6-8, 2016*. 2080–2085. https://doi.org/10.1109/ACC.2016.7525225

[64] Zuxing Li and Tobias J. Oechtering. 2015. Privacy on hypothesis testing in smart grids. In *2015 IEEE Information Theory Workshop - Fall (ITW), Jeju Island, South Korea, October 11-15, 2015*. 337–341. https://doi.org/10.1109/ITWF.2015.7360791

[65] Zuxing Li, Tobias J. Oechtering, and Deniz Gündüz. 2018. Smart Meter Privacy: Adversarial Hypothesis Testing Models. *CoRR* abs/1807.01916 (2018). arXiv:1807.01916 http://arxiv.org/abs/1807.01916

[66] Zuxing Li, Tobias J. Oechtering, and Mikael Skoglund. 2016. Privacy-preserving energy flow control in smart grids. In *2016 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2016, Shanghai, China, March 20-25, 2016*. 2194–2198. https://doi.org/10.1109/ICASSP.2016.7472066

[67] Mikhail Lisovich and Stephen Wicker. 2008. Privacy concerns in upcoming residential and commercial demand-response systems. *IEEE Proceedings on Power Systems* 1, 1 (2008), 1–10.

[68] Yanchi Liu, Zhongmou Li, Hui Xiong, Xuedong Gao, Junjie Wu, and Sen Wu. 2013. Understanding and Enhancement of Internal Clustering Validation Measures. *IEEE Trans. Cybernetics* 43, 3 (2013), 982–994. https://doi.org/10.1109/TSMCB.2012.2220543

[69] Pascal Massart. 2007. Concentration inequalities and model selection. (2007).

[70] Jim McDowall. 2006. Integrating energy storage with wind power in weak electricity grids. *Journal of Power sources* 162, 2 (2006), 959–964.

[71] Eoghan McKenna, Ian Richardson, and Murray Thomson. 2012. Smart meter data: Balancing consumer privacy concerns with legitimate applications. *Energy Policy* 41 (2012), 807–814.

[72] Stephen E. McLaughlin, Patrick D. McDaniel, and William Aiello. 2011. Protecting consumer privacy from electric load monitoring. In *Proceedings of the 18th ACM Conference on Computer and Communications Security, CCS 2011, Chicago, Illinois, USA, October 17-21, 2011*. 87–98. https://doi.org/10.1145/2046707.2046720

[73] Marcelo G Molina. 2017. Energy storage and power electronics technologies: A strong combination to empower the transformation to the smart grid. *Proc. IEEE* 105, 11 (2017), 2191–2219.

[74] Andres Molina-Markham, Prashant J. Shenoy, Kevin Fu, Emmanuel Cecchet, and David E. Irwin. 2010. Private memoirs of a smart meter. In *BuildSys'10, Proceedings of the 2nd ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings, Zurich, Switzerland, November 3-5, 2010*. 61–66. https://doi.org/10.1145/1878431.1878446

[75] Adarsh Nagarajan and Raja Ayyanar. 2014. Design and strategy for the deployment of energy storage systems in a distribution feeder with penetration of renewable resources. *IEEE Transactions on Sustainable Energy* 6, 3 (2014), 1085–1092.

[76] Lucas Pereira and Nuno Nunes. 2018. Performance evaluation in non-intrusive load monitoring: Datasets, metrics, and tools - A review. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.* 8, 6 (2018). https://doi.org/10.1002/widm.1265

[77] Miroslav P Petrov, Reza Arghandeh, and Robert Broadwater. 2013. Concept and application of distributed compressed air energy storage systems integrated in utility networks. In *ASME Power Conference*.

[78] Manisa Pipattanasomporn, Murat Kuzlu, Saifur Rahman, and Yonael Teklu. 2013. Load profiles of selected major household appliances and their demand response opportunities. *IEEE Transactions on Smart Grid* 5, 2 (2013), 742–750.

[79] S. Raj Rajagopalan, Lalitha Sankar, Soheil Mohajer, and H. Vincent Poor. 2011. Smart meter privacy: A utility-privacy framework. In *IEEE Second International Conference on Smart Grid Communications, SmartGridComm 2011, Brussels, Belgium, October 17-20, 2011*. 190–195. https://doi.org/10.1109/SmartGridComm.2011.6102315

[80] Ishtiaq Rouf, Hossen A. Mustafa, Miao Xu, Wenyuan Xu, Robert D. Miller, and Marco Gruteser. 2012. Neighborhood watch: security and privacy analysis of automatic meter reading systems. In *the ACM Conference on Computer and Communications Security, CCS'12, Raleigh, NC, USA, October 16-18, 2012*. 462–473. https://doi.org/10.1145/2382196.2382246

[81] Peter J Rousseeuw. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics* 20 (1987), 53–65.

[82] Lalitha Sankar, S. Raj Rajagopalan, Soheil Mohajer, and H. Vincent Poor. 2013. Smart Meter Privacy: A Theoretical Framework. *IEEE Trans. Smart Grid* 4, 2 (2013), 837–846. https://doi.org/10.1109/TSG.2012.2211046

[83] Wouter L Schram, Ioannis Lampropoulos, and Wilfried GJHM van Sark. 2018. Photovoltaic systems coupled with batteries that are optimally sized for household self-consumption: Assessment of peak shaving potential. *Applied energy* 223 (2018), 69–81.

[84] Konark Sharma and Lalit Mohan Saini. 2015. Performance analysis of smart metering for smart grid: An overview. *Renewable and Sustainable Energy Reviews* 49 (2015), 720–735.

[85] Onur Tan, Jesús Gómez-Vilardebó, and Deniz Gündüz. 2017. Privacy-Cost Trade-offs in Demand-Side Management With Storage. *IEEE Trans. Information Forensics and Security* 12, 6 (2017), 1458–1469. https://doi.org/10.1109/TIFS.2017.2656469

[86] Onur Tan, Deniz Gündüz, and H. Vincent Poor. 2012. Smart meter privacy in the presence of energy harvesting and storage devices. In *IEEE Third International Conference on Smart Grid Communications, SmartGridComm 2012, Tainan, Taiwan, November 5-8, 2012*. 664–669. https://doi.org/10.1109/SmartGridComm.2012.6486062

[87] Onur Tan, Deniz Gündüz, and H. Vincent Poor. 2013. Increasing Smart Meter Privacy Through Energy Harvesting and Storage Devices. *IEEE Journal on Selected Areas in Communications* 31, 7 (2013), 1331–1341. https://doi.org/10.1109/JSAC.2013.130715

[88] Andrei G Ter-Gazarian. 2011. *Energy storage for power systems* (2nd ed.). The Institution of Engineering and Technology. 292 pages. https://doi.org/10.1049/PBPO063E

[89] Stephan Thomas, Tobias Blank, Eva Szczechowicz, Thomas Pollok, Christoph Roggendorf, Ionut Trintis, Rik W De Doncker, Armin Schnettler, Frede Blaabjerg, Stephan Thomas, and Tobias Blank. 2016. *HERMES: Highly Efficient and Reliable Modular Battery Energy Storage Systems*. Technical Report 8.

[90] I Tsiropoulos, D Tarvydas, and N Lebedeva. 2018. Li-ion Batteries for Mobility and Stationary Storage Applications Scenarios for Costs and Market Growth. *Publications Office of the European Union: Luxembourg* (2018).

[91] Valentin Tudor, Magnus Almgren, and Marina Papatriantafilou. 2013. Analysis of the impact of data granularity on privacy for the smart grid. In *Proceedings of the 12th annual ACM Workshop on Privacy in the Electronic Society, WPES 2013, Berlin, Germany, November 4, 2013*. 61–70. https://doi.org/10.1145/2517840.2517844

[92] David P. Varodayan and Ashish Khisti. 2011. Smart meter privacy using a rechargeable battery: Minimizing the rate of information leakage. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2011, May 22-27, 2011, Prague Congress Center, Prague, Czech Republic*. 1932–1935. https://doi.org/10.1109/ICASSP.2011.5946886

[93] Xiaoyue Wang, Abdullah Mueen, Hui Ding, Goce Trajcevski, Peter Scheuermann, and Eamonn J. Keogh. 2013. Experimental comparison of representation methods

and distance measures for time series data. *Data Min. Knowl. Discov.* 26, 2 (2013), 275–309. https://doi.org/10.1007/s10618-012-0250-5

[94] Yi Wang, Qixin Chen, Tao Hong, and Chongqing Kang. 2019. Review of Smart Meter Data Analytics: Applications, Methodologies, and Challenges. *IEEE Trans. Smart Grid* 10, 3 (2019), 3125–3148. https://doi.org/10.1109/TSG.2018.2818167

[95] Zhimin Wang, Chenghong Gu, Furong Li, Philip Bale, and Hongbin Sun. 2013. Active Demand Response Using Shared Energy Storage for Household Energy Management. *IEEE Trans. Smart Grid* 4, 4 (2013), 1888–1897. https://doi.org/10.1109/TSG.2013.2258046

[96] Jeffrey M Wooldridge. 2016. *Introductory econometrics: A modern approach.* Nelson Education.

[97] Weidong Xiao. 2017. *Photovoltaic Power System: Modeling, Design, and Control.* John Wiley & Sons Ltd.

[98] Lei Yang, Xu Chen, Junshan Zhang, and H. Vincent Poor. 2014. Optimal privacy-preserving energy management for smart meters. In *2014 IEEE Conference on Computer Communications, INFOCOM 2014, Toronto, Canada, April 27 - May 2, 2014.* 513–521. https://doi.org/10.1109/INFOCOM.2014.6847975

[99] Lei Yang, Xu Chen, Junshan Zhang, and H. Vincent Poor. 2015. Cost-Effective and Privacy-Preserving Energy Management for Smart Meters. *IEEE Trans. Smart Grid* 6, 1 (2015), 486–495. https://doi.org/10.1109/TSG.2014.2343611

[100] Weining Yang, Ninghui Li, Yuan Qi, Wahbeh H. Qardaji, Stephen E. McLaughlin, and Patrick D. McDaniel. 2012. Minimizing private data disclosures in the smart grid. In *the ACM Conference on Computer and Communications Security, CCS'12, Raleigh, NC, USA, October 16-18, 2012.* 415–427. https://doi.org/10.1145/2382196.2382242

[101] Jiyun Yao and Parv Venkitasubramaniam. 2015. The Privacy Analysis of Battery Control Mechanisms in Demand Response: Revealing State Approach and Rate Distortion Bounds. *IEEE Trans. Smart Grid* 6, 5 (2015), 2417–2425. https://doi.org/10.1109/TSG.2015.2438035

[102] Jing Zhao, Taeho Jung, Yu Wang, and Xiangyang Li. 2014. Achieving differential privacy of data disclosure in the smart grid. In *2014 IEEE Conference on Computer Communications, INFOCOM 2014, Toronto, Canada, April 27 - May 2, 2014.* 504–512. https://doi.org/10.1109/INFOCOM.2014.6847974

[103] Zhibin Zhou, Mohamed Benbouzid, Jean Frédéric Charpentier, Franck Scuiller, and Tianhao Tang. 2013. A review of energy storage technologies for marine current energy systems. , 390–400 pages.

[104] Liehuang Zhu, Zijian Zhang, Zhan Qin, Jian Weng, and Kui Ren. 2017. Privacy Protection Using a Rechargeable Battery for Energy Consumption in Smart Grids. *IEEE Network* 31, 1 (2017), 59–63. https://doi.org/10.1109/MNET.2016.1500292NM

A SPECIALIZED PRIVACY MEASURES

In constrast to the general privacy measures surveyed in Section 2, specialized measures require a lot of additional information, e.g., an hypothesis that is being tested, a description of the probabilistic battery charging strategy or the schedule of each individual appliance. An overview of all specialized privacy measures can be found in Table 9. As with general privacy measures before, the table distinguishes whether a paper applies the respective measure to original load profiles (x, y) or their time differences $(\Delta x, \Delta y)$. For several measures this separation is not applicable or not meaningful, in this case references are centered between the columns.

Measure	(x, y)	$(\Delta x, \Delta y)$
AMBR	[9, 66]	
Classification accuracy loss	[62]	
Conditional K-Divergence	[52]	[52]
Confusability	[62]	
Fisher information	[36]	
NILM-based	[9, 32, 34, 36, 61, 67]	
Re-identification	[61, 62], [57][a]	
Sum of distance	[20]	
Type II error probability	[64], [65][a]	

[a] It is not clear from the description, whether time differences are taken.

Table 9: Overview of specialized privacy measures.

Accumulated (Discounted) Minimal Bayesian Risk. This measure considers appliance vectors $h_t \in \{0,1\}^n$. Each entry $(h_t)_i = 1$ with $i \in \{1, \ldots, n\}$ indicates that household appliance i is turned on at time t. The users consumption behavior is given as a stochastic process H^T of appliance vectors with probability mass function P_{H^T}. This induces probability functions P_{H_t} for individual time slots. An adversarial decision strategy δ outputs appliance vectors $h^* \in \{0,1\}^n$ which indicate the adversary's guess about the actual consumption. The individual adversarial model determines the input upon which this decision is reached. We denote the set of all possible decision strategies by $\mathscr{D}$. A cost function $c_t : (\{0,1\}^n)^2 \to \mathbb{R}_+$ gives the cost $c_t(h_t^*, h_t)$ that incurs to the adversary if he guesses h_t^* while the true consumption behavior at time t was h_t. We can now calculate the *minimal Bayesian risk at time t*:

$$r_t^* = \min_{\delta \in \Delta} \sum_{h_t, h_t^* \in \{0,1\}^n} c_t(h_t^*, h_t) \cdot \mathbb{P}\left[h_t^* \leftarrow \delta \,\middle|\, h_t\right] \cdot P_{H_t}(h_t).$$

The weighted sum of individual minimal Bayesian risks over all times t results in the *accumulated discounted minimal Bayesian risk*:

$$V = \sum_{t=1}^{T} \beta^{t-1} r_t^*.$$

The weight basis β is called *discount factor*. It models the assumption that private information becomes less valuable as it gets older. Although V is defined in terms of expected costs to an adversary, it can be construed as a privacy measure: The higher the costs to the adversary, the more privacy is retained. Note that although this measure does not explicitly take any battery charging strategy as input, user and grid load, battery in-/outputs and charging strategy may all contribute to the inputs of the decision strategy δ. An intuitive adversarial model would, e.g., have $h_t^* := \delta(y_1, \ldots, y_t)$ depending on the previous grid load—which in turn depends on consumption and charging behavior.

In [66], Li et al. model user's consumption as a first order Markov process given by a (time invariant) probability mass function $P_{H|\mathfrak{B}H} = P_{H_{t+1}|H_t}$, where $\mathfrak{B}$ denotes the backshift operator. The stochastic process X^T of the user load deterministically dependends on the consumption behaviour H^T, but only aggregate consumption ΣX^T with $(\Sigma X)_t := X_1 + \cdots + X_t$ is considered. Similarly, grid load and storage in-/outputs are accumulated time series ΣY^T and ΣB^T. The user's battery control strategy is again modeled as a stochastic process given by a time invariant probability mass function

$$P_{\Sigma B|\mathfrak{B}(\Sigma B),\mathfrak{B}(\Sigma X)} = P_{(\Sigma B)_{t+1}|(\Sigma B)_t,(\Sigma X)_t}$$

The grid load ΣY^T is deterministically determined by the stochastic processes of user load and battery charging. [66] uses a rather extensive adversarial model, where decision strategies are defined by a probability mass function $P_{H_t^*|(\Sigma Y)_t}$ that depends on accumulated grid load as well as the adversary's knowledge of $P_{H|\mathfrak{B}H}$ and $P_{\Sigma B|\mathfrak{B}(\Sigma B),\mathfrak{B}(\Sigma X)}$. The cost functions c_t are assumed to be simple: $c_t(h_t^*, h_t) = 1$ whenever $h_t^* \neq h_t$ and $c_t(h_t^*, h_t) = 0$ otherwise.

Avula et al. [9] use a very similar approach, but with undiscounted risk ($\beta = 1$).

Classification Accuracy Loss. Classification accuracy loss measures how much worse a classification algorithm with a specific prediction target performs on the modified grid load compared to the

actual user load. Let $S = \{(x^T, y^T)_i \mid i \in I\}$ be a set of pairs of user and grid loads resulting from the application of a specific battery algorithm. Let furthermore $tg : I \to \mathrm{im}(tg)$ denote a prediction target (e.g., employment/retirement status of household inhabitants) on this set. Let C_{tg} be a deterministic classification algorithm which takes a time series z^T of length T and outputs a classification $C_{tg}(z^T) \in \mathrm{im}(tg)$. We can now define a baseline accuracy

$$\mathrm{acc}_{\mathrm{base}}(tg, S) := \max_{\tau \in \mathrm{im}(tg)} \frac{\#\{i \in I \mid tg(i) = \tau\}}{\#S}$$

for the prediction target tg on S, and the accuracy of C_{tg} regarding user and grid loads in S

$$\mathrm{acc}_{\mathrm{ul}}(C_{tg}, S) := \frac{\#\{i \in I \mid C_{tg}(x_i^T) = tg(i)\}}{\#S}$$

$$\mathrm{acc}_{\mathrm{gl}}(C_{tg}, S) := \frac{\#\{i \in I \mid C_{tg}(y_i^T) = tg(i)\}}{\#S}.$$

The *classification accuracy loss* on S regarding C_{tg} is defined as

$$\mathrm{CAL}_{C_{tg}}(S) := \frac{\mathrm{acc}_{\mathrm{ul}}(C_{tg}, S) - \mathrm{acc}_{\mathrm{gl}}(C_{tg}, S)}{\mathrm{acc}_{\mathrm{ul}}(C_{tg}, S) - \mathrm{acc}_{\mathrm{base}}(tg, S)}.$$

This measure is used by [62] for the prediction targets employment status, number of devices, social class and retirement status.

Conditional K-Divergence. Kalogridis et.al. of [52] model load profiles as (first order) Markov chains and use conditional K-divergence between user and grid load [6]:

$$K_{\omega_t|\omega_{t-1}}(X\|Y) = \sum_{\omega_{t-1}\in\Omega} P_X(\omega_{t-1}) \sum_{\omega_t\in\Omega} P_{X_t|X_{t-1}}(\omega_t|\omega_{t-1}) \cdot \cdot \log \frac{2P_{X_t|X_{t-1}}(\omega_t|\omega_{t-1})}{P_{X_t|X_{t-1}}(\omega_t|\omega_{t-1}) + P_{Y_t|Y_{t-1}}(\omega_t|\omega_{t-1})}.$$

Here P_X is a steady-state probability distribution, with P_X, P_Y estimated from x^T, y^T as well as $\Delta x^T, \Delta y^T$.

Confusability. In this case, the battery algorithm is again assumed to be probabilistic and we look at the random grid load vectors Y^T instead of sampled grid load instances $y^T \leftarrow Y^T$. But in contrast to most other measures, the adversary does not know the complete grid load $y^T \leftarrow Y^T$, but is restricted to the output $q(Y^T) \in \mathrm{im}(q)$ of only one query q. Comparability with other measures could be achieved by using a query which samples the grid load and outputs the resulting time series $q(Y^T) = y^T$ with $y^T \leftarrow Y^T$. For this specific query we can define the pairwise confusability of two random grid load vectors Y_1^T, Y_2^T via the probability density functions $f_{q(Y_1^T)}$ and $f_{q(Y_1^T)}$:

$$\sigma(Y_1^T, Y_2^T) := \int_{y^T \in \mathrm{im}(q)} \min\left\{ f_{q(Y_1^T)}(y^T), f_{q(Y_2^T)}(y^T) \right\} \mathrm{d}y^T.$$

As with classification accuracy loss, this measure needs a prediction target tg on a specific user and grid load set $S = \{(x_i^T, Y_i^T) \mid i \in I\}$, this time with random grid load vectors Y_i^T instead of time series y_i^T. Now for $m \in \mathbb{N}$ and probability $\sigma \in [0, 1]$ the query q is called (σ, m)-*confusable* regarding S, if and only if for every $i \in I$ there are distinct $j_1, \ldots, j_m \in I$ such that

$$tg(x_i^T) \neq tg(x_{j_k}^T) \qquad \text{and} \qquad \sigma(Y_i^T, Y_{j_k}^T) \geq \sigma$$

for every $k \in \{1, \ldots, m\}$. In the above scenario this means the probability to confuse the grid load of household i with that of a different household j is at least σ for at least m households with different target label $tg(x_j^T)$.

Confusability as a measure for BBLH algorithms was introduced and estimated in [62].

Differential Privacy. Differential privacy is not inherently a measure, rather a parameterized privacy property which some BBLH algorithms satisfy, but most do not. Differential privacy ensures that a single person's sensitive data is protected while accurate general information (e.g. statistics about the whole population) can be accessed. Hence application of differential privacy to the BBLH scenario is not straight foreward and requires some explanations. We first give the general definition before we discuss the special BBLH setting.

Differential privacy was first introduced by Dwork et al. in [28][7], with the slightly weaker and more commonly used notion of (ε, δ)-differential privacy introduced in [27]. Both are properties of randomized functions on datasets. Formally, a function $\mathscr{K}$ gives (ε, δ)-differential privacy, if for all neighboring datasets D_1, D_2 and $S \subseteq \mathrm{Range}(\mathscr{K})$ the inequality

$$\mathbb{P}[\mathscr{K}(D_1) \in S] \leq \exp(\varepsilon) \cdot \mathbb{P}[\mathscr{K}(D_2) \in S] + \delta$$

holds. Datasets usually contain one entry per person and are neighboring if they differ in exactly one entry. While the output of $\mathscr{K}$ is to be released, the input dataset is kept secret. The parameter δ is used as the "probability of exception", where ε-differential privacy can not be guaranteed. I.e. unless an improbable event (with probability at most δ) occurs, the data of one individual may only increase the probability of any output of $\mathscr{K}$ by a small factor $\exp(\varepsilon)$. But if changing the data pertaining to one person can only marginally influence the output of $\mathscr{K}$, then releasing this output will not affect any single person's privacy too much.

Differential privacy is commonly achieved by adding noise (e.g. from a suitably parameterized Laplace distribution) to the output one is actually interested in. This noise can be provided by a battery.

In our BBLH scenario, each user's/household's data is considered individually and not mixed with other household's consumption profiles. In this case, each database entry corresponds to the consumption of one application within the household and $\mathscr{K}$ outputs the modified version y_t of the sum x_t of all database entries. Differential privacy was applied to the BBLH setting in this manner by [11, 62, 102]. Since no storage systems can provide unlimited capacity, it is not possible to achieve ε-differential privacy in this way, only (ε, δ)-differential privacy. The exception probability δ is required for cases where the storage system would have to be charged although it is already full (or discharged although it is empty) to provide adequate privacy protection.

[6] The publication [52] introducing conditional K-divergence applied to time series data only provides the general formula, without specifying the meaning of the variables. So this is our interpretation to some extent.

[7] This is the revised version of the original paper from 2006.

Fisher Information. Again we look at time series x^T of user load and the probabilistic outcome Y^T of a BBLH algorithm. With the probability density function f_{Y^T} of this random grid load vector, we can calculate the Fisher information matrix $I(x^T, Y^T)$. Its entries are defined as

$$FI(x^T, Y^T)_{i,j} = \int_{y^T} f_{Y^T}(y^T) \cdot \frac{\partial}{\partial x_i} \log\left(f_{Y^T}(y^T)\right) \cdot \frac{\partial}{\partial x_j} \log\left(f_{Y^T}(y^T)\right) \mathrm{d}y^T$$

for all $i, j \in \{1, \ldots, T\}$. The trace

$$\mathrm{Tr}\left(FI(x^T, Y^T)^{-1}\right)$$

can be used as an estimate of the privacy retained via the BBLH algorithm. This is justified by the fact that for any unbiased user load estimator $(x^*)^T$ this trace gives a lower bound on the expected L_2-norm error

$$\mathbb{E}\left[\left\|x^T - (x^*)^T(y^T)\right\|_2^2 \middle| y^T \leftarrow Y^T\right] \geq \mathrm{Tr}\left(FI(x^T, Y^T)^{-1}\right),$$

as long as mild regularity conditions are satisfied.

Farokhi and Sandberg [36] first proposed to apply this measure to the context of BBLH algorithms, but use it to construct an algorithm rather than to evaluate the result. Calculating this measure for empirical data requires estimation of the function $\log\left(f_{Y^T}\right)$ and its T partial derivatives. As this is impossible without any additional assumptions from a single sample, using this measure for privacy evaluation from empirical data is problematic.

NILM-based Measures. There are multiple non-intrusive load monitoring (NILM) techniques [76] which process the load profiles to extract the schedule of individual appliances. One may argue, that much of sensitive private information can be inferred from such decomposition. Thus, the drop of performance of a NILM technique when the user load x^T is replaced with a grid load y^T can serve as a privacy measure. This performance can be quantified for instance, via accuracy [61, 67], F-score [9, 32, 36], AUC [34] or other statistics [76].

To apply NILM technique's performance drop to measure privacy, one needs to decide which decomposition task to consider. For instance Erdogdu and others [34] assume that the adversary tries to distinguish between on/of states of the microwave, in [9, 32, 36, 61] different appliances are considered. One also needs to choose a specific NILM technique and decide on the data used to train it.

Re-identification. For re-identification the adversary has a database of N records of grid loads y_i^T, $i = 1, \ldots, N$ with personal identifier removed. For each household there are values of several statistics, called queries, as external knowledge. Examples of features are energy consumption between 4 and 8 a.m. or the average bedtime. For each i the adversary computes the values of these features for y_i^T and compares the results to the external knowledge. A result of such comparison is a score. A household j is re-identified if its grid load y_j^T is within n grid loads with lowest scores. The final value of this privacy measure is affected by database, selected features, matching algorithm calculating the score and the value of n.

Re-identification is used by [57, 61, 62]. Even without any privacy algorithm re-identification is usually not perfect. Hence, to asses a BBLH algorithm, comparison of re-identification accuracy before and after modification is necessary. This is done either visually [62] or via the ratio $A_{xx} - A_{xy}/A_{xx}$ [57, 61], where A_{xx}, A_{yy} is the accuracy of re-identification without and with load modification respectively.

Sum of Distance. This metric was proposed in [20]. It takes into account both temporal and spatial distances among the appliances:

$$SOD(x^T, y^T) = \sum_a \sqrt{\sum_{t=1}^{T} (y_{a,t} - \bar{x}_a)^2} + \sum_{t=1}^{T} \sqrt{\sum_a (y_{a,t} - x_t)^2}.$$

Here $y_{a,t}$ is the measured electricity consumption of appliance a at time t, $\bar{x}_a$ is the average power consumption of this appliance throughout the period T. Smaller values of $SOD(x^T, y^T)$ are supposed to reflect better privacy protection.

Type II Error Probability. The measure was proposed in [64, 65]. It estimates privacy against testing of some binary hypothesis $\mathscr{H} = \{h_0, h_1\}$. Under several assumptions on the nature of load profile time series (i.i.d., as required by Chernoff-Stein Lemma), the behavior of adversary and the information available to them, estimation of minimal type II error probability reduces to calculation

$$D(Y|h_0||Y|h_1),$$

where $D(\cdot||\cdot)$ is the Kullback-Leibler divergence. Application of this measure in reality for privacy evaluation requires specifying the hypothesis being tested. This can be, for instance, the model of a dishwasher, as in [65]. The limitations of this use case is application to a single load – the dishwasher. That is, the other appliances should not interfere the consumption and the observation time should be limited to the period when the appliance is on. The possibility to extend this measure to more general case and how a hypothesis could look like then is not obvious.

B RECONSTRUCTABILITY

Let $\mathscr{Z}$ be the set of all possible random processes of loads, and let $\mathscr{X} \subseteq \mathscr{Z}$ be a (benchmark) family of random processes of user loads. For any (not necessarily deterministic) privacy algorithm PA, which takes the realisation $x^T \leftarrow X^T$ of some $X^T \in \mathscr{X}$ and outputs a grid load $y^T \leftarrow PA(x^T)$, we can derive a deterministic privacy algorithm $PA_{\mathscr{X}} : \mathscr{X} \to \mathscr{Z}$ where $X^T \in \mathscr{X}$ maps to the random grid load process $Y^T := PA_{\mathscr{X}}(X^T)$ which results from taking the randomness in the steps $x^T \leftarrow X^T$ and $y^T \leftarrow PA(x^T)$ into account.

Proposition (Reconstructability). *No privacy measure (w.l.o.g.) $PM : \mathscr{X} \times \mathscr{Z} \mapsto [0, 1]$ (where 0 indicates no and 1 indicates perfect privacy) can be "sufficiently good" to assess privacy algorithms if there is a privacy algorithm $PA_{\mathscr{X}}$ and a reconstruction algorithm $RA : \mathscr{Z} \to \mathscr{Z}$ such that*

$$\Pr\left[PM\left(X^T, RA\left(PA_{\mathscr{X}}(X^T)\right)\right) < PM\left(X^T, PA_{\mathscr{X}}(X^T)\right) \middle| X^T \leftarrow \mathscr{X}\right]$$

is non-negligible.

Storage Technology	Energy Density	Power Density	Number of Cycles	Round-trip Efficiency	Degradation	Costs	Applicable
Lead-Acid Battery	−	++	−−	++	−−	+ + +	No
Lithium-ion Batteries	+ + +	+ + +	+ + +	++	−	++	Yes
Vanadium Redox Flow Battery (VRB)	−	−−	+ + +	−	+	++	Yes
Zinc Bromium (ZnBr) Batteries	+	+	+ + +	−	+	++	Yes
Polysulphide Bromide Battery (PSB)	−−	−−	+	−	+	+	No
Metal-Air Batteries	+ + ++	+ + +	− − −−	− − −	− − −−	+ + +	No
Supercapacitors	−−	+ + ++	+ + ++	+ + +	−−	−	No
Superconducting Magnetic Energy Storage	− − −−	+ + +	+ + ++	+ + +	+ + +	− − −−	No
Flywheel Energy Storage Systems	−	+ + +	+ + ++	+ + +	+ + ++	− − −−	No
Compressed Air Energy Storage Systems	− − −	− − −−	++	− − −−	++	+ + +	No
Thermal Energy Storage Systems	++	++	+ + ++	− − −	+ + +	+ + ++	Yes

Table 10: Comparison of storage technologies considering characteristics required for privacy protection.

C OTHER STORAGE TECHNOLOGIES

There are many energy storage technologies not mentioned in Section 3.2 because we deem them unsuitable for privacy protection. We now briefly explain this.

The lifetime of Metal-Air batteries is limited to only 100–300 cycles [6]. For the other type of flow batteries, Polysulphide Bromide Batteries (PSB), literature has so far only reported on showcases for this technology, mostly for grid-scale applications [19]. Supercapacitors and superconducting magnetic energy storage systems cannot provide the capacity (few kWh) required for households in a reasonable space.

Next, there are two types of mechanical energy storage systems which have reached the capacity and power required for households in the last decades. One is compressed air energy-storage systems, which now can provide a power output around 3 kW [77]. But they are still highly inefficient with a round trip efficiency of only 50% [49]. The low energy density of this technology also mandates a large tank to be installed, an issue for many household. The second type of mechanical energy storage system are the high-speed flywheel energy storage systems. With advances in high-speed electrical machines, magnetic bearings, and composite materials, flywheel energy storage systems can now reach few kWh of capacity [55]. However, such flywheels continue much more costly, compare to batteries.

One attractive alternative for privacy protection could be thermal energy storage systems, but only if their source of energy is electricity. Obviously, When gas or other sources of energy provide the demand for thermal energy, we cannot alter the load profile using such thermal energy storage systems. In Germany, electric water heaters take up to 1100 kWh per year [17], more than 25% of the annual electricity consumption. However, thermal energy storage systems are often equipped with separate meters and controlled remotely by the grid operator. An example is the so-called night heating storage system or storage heater found in some households in the UK, Austria and Germany [88]. Grid operators can often control these devices remotely with signals sent using power-line communication carriers. So the grid operator can observe whether the device is being operated or not.

A more comprehensive review of all storage technologies and their applicability for the application of privacy protection in households is given in Table 10. This table also shows the incompatible technologies, as compared to table 3.

D LOAD PROFILE CHARACTERISTICS

Tables 11 and 12 summarize lengths and sampling rates of user load profiles used in the literature evaluating BBLH algorithms. Both the lengths and sampling rates vary a lot, with the range of possible values differing by two and three orders of magnitude.

Length (days)	References
1–13	[21, 50, 53, 54, 98, 101]
24–36	[51, 52, 72, 85, 99, 100, 104]
61–530	[22, 57, 61, 62, 72, 100]

Table 11: Lengths of load profiles used in the other papers.

Sampling rate (s)	References
1–30	[34, 50–52, 67, 72, 100]
60–300	[11, 20, 53, 54, 85, 98, 99]
900–3600	[22, 57, 61, 62, 104]

Table 12: Sampling rates of load profiles used in the other papers.

E COMPLETE EXPERIMENTAL RESULTS

Measure	Δ	BE1	BE2	LC	LS1	LS2	NILL	RC
H	y	4	3	6	2	1	7	5
H	n	7	6	4	2	3	5	1
CS	y	5	7	4	2	1	6	3
CS	n	4	6	2	5	3	7	1
R_2^2	y	2	3	7	4	5	6	1
R_p^2	n	1	3	5	6	7	2	4
R_2^2	n	5	7	2	3	6	4	1
ER^{nz}	y	7	5	3	1	2	4	6
ER^{z}	y	5	1	6	2	3	4	7
FM_{ed}	y	2	4	6	5	3	7	1
FM	y	2	1	6	4	5	3	7
FM_r	y	7	6	4	1	2	3	5
K	n	6	7	1	2	3	5	4
KL	y	1	6	4.5	2	4.5	7	3
KL	n	3	6	1.5	4	1.5	7	5
LV	n	3	2	6	4	7	1	5
MI^{b}	y	2	1	7	6	5	4	3
MI^{i}	y	4	3	6	2	1	7	5
MI_s	y	5	3	6	1	2	4	7
MI^{i}	n	7	6	4	2	3	5	1
MI^{m}	n	7	6	4	2	3	5	1
MI_s	n	7	5	3	1	2	4	6
RU^{r}	n	2	1	6	4	5	3	7
RU^{w}	n	1.5	3	6	4	5	1.5	7
TVD	n	6	7	1	2	3	4	5
min rank		1	1	1	1	1	1	1
max rank		7	7	7	6	7	7	7

Table 13: Average algorithm ranking with respect to privacy measures.

Load profile	BE1	BE2	LC	LS1	LS2	NILL	RC
ECO1	6	5	3	1	2	4	7
ECO2	4	3	6	1	2	5	7
ECO3	4	1	5	2	3	7	6
ECO4	5	1	7	4	3	2	6
ECO5	5	2	3	4	6	1	7
ECO6	7	5.5	4	2	3	5.5	1
ECO7	6	4	5	1	3	2	7
ECO8	6	3	4	1	2	5	7
ECO9	5	1	4	2	3	6	7
ECO10	4	3	6	1	5	2	7
ECO11	7	6	4	3	2	1	5
ECO12	7	6	5	1	2	4	3
ECO13	6	5	4	1	2	3	7
ECO14	6	4	5	2	1	3	7
ECO15	7	5	2	3.5	3.5	1	6
ECO16	2	5	4	1	3	6	7
ECO17	6	5	4	1	2	3	7
ECO18	5	7	3	1	2	6	4
ECO19	5	2	6	3	4	1	7
ECO20	7	5	4	2	1	3	6
ECO21	7	5	1	3	2	4	6
ECO22	6	1	2	5	4	3	7
ECO23	7	6	4	2	3	1	5
ECO24	6	4	5	1	2	7	3
REDD1	5	4	6	2	3	1	7
REDD2	2	5	6	4	3	1	7
REDD3	6	2	1	4	5	3	7
REDD4	2	4	7	5	3	1	6
REDD5	6	5	3	2	4	1	7
SmartB1	4	3	6	1	2	7	5
SmartB2	4	2	6	1	3	7	5
CER1	4	5	6	3	2	7	1
CER2	5	6	4	2	3	7	1
CER3	4	6.5	5	2	3	6.5	1
CER4	4	6	5	1	2	7	3
CER5	3	5	6	4	2	7	1
CER6	4	6	5	2	1	7	3
min rank	2	1	1	1	1	1	1
max rank	7	7	7	5	6	7	7

Table 14: Average algorithm ranking with respect to load profiles.

F LOAD PROFILES USED IN EXPERIMENTS

Table 15 summarizes all load profiles we have used in our experiments. We formed the name of each load profile from the name of the respective dataset (cf. Table 5) plus a number. The sampling rate is the periodicity of measurements given in seconds. This sampling rate does not, in general, coincide with the sampling rate of the respective dataset from Table 5, as we have already explained in Section 5. "# Res." indicates the number of residents in a household and "Household" specifies the household identifier for which we use the data. This table uniquely identifies the data we use and can be used for reproducing our results.

Name	Sample rate	# Res.	Location	Season	Day	Employed	Units	Start date	End date	Household
ECO1	30	4	EU	Summer	Weekend	y	W	15.07.12	15.07.12	household 1
ECO2	30	4	EU	Summer	Weekday	y	W	16.07.12	16.07.12	household 1
ECO3	60	4	EU	Summer	Weekday	y	W	17.07.12	17.07.12	household 1
ECO4	60	4	EU	Summer	Weekend	y	W	22.07.12	22.07.12	household 1
ECO5	900	4	EU	Summer	1 week	y	W	23.07.12	29.07.12	household 1
ECO6	1800	4	EU	Summer	2 weeks	y	W	30.07.12	12.08.12	household 1
ECO7	30	4	EU	Winter	Weekend	y	W	25.11.12	25.11.12	household 1
ECO8	30	4	EU	Winter	Weekday	y	W	03.12.12	03.12.12	household 1
ECO9	60	4	EU	Winter	Weekday	y	W	03.12.12	03.12.12	household 1
ECO10	60	4	EU	Winter	Weekend	y	W	09.12.12	09.12.12	household 1
ECO11	900	4	EU	Winter	1 week	y	W	10.12.12	16.12.12	household 1
ECO12	1800	4	EU	Winter	2 weeks	y	W	17.12.12	30.12.12	household 1
ECO13	30	2	EU	Summer	Weekend	y	W	15.07.12	15.07.12	household 2
ECO14	30	2	EU	Summer	Weekday	y	W	16.07.12	16.07.12	household 2
ECO15	60	2	EU	Summer	Weekday	y	W	17.07.12	17.07.12	household 2
ECO16	60	2	EU	Summer	Weekend	y	W	22.07.12	22.07.12	household 2
ECO17	900	2	EU	Summer	1 week	y	W	23.07.12	29.07.12	household 2
ECO18	1800	2	EU	Summer	2 weeks	y	W	30.07.12	12.08.12	household 2
ECO19	30	2	EU	Winter	Weekend	y	W	09.12.12	09.12.12	household 2
ECO20	30	2	EU	Winter	Weekday	y	W	10.12.12	10.12.12	household 2
ECO21	60	2	EU	Winter	Weekday	y	W	10.12.12	10.12.12	household 2
ECO22	60	2	EU	Winter	Weekend	y	W	16.12.12	16.12.12	household 2
ECO23	900	2	EU	Winter	1 week	y	W	10.12.12	16.12.12	household 2
ECO24	1800	2	EU	Winter	2 weeks	y	W	17.12.12	30.12.12	household 2
REDD1	30	na	US	Spring	Weekday	na	W	26.04.11	26.04.11	house 1
REDD2	30	na	US	Spring	Weekend	na	W	07.05.11	07.05.11	house 1
REDD3	60	na	US	Spring	Weekend	na	W	01.05.11	01.05.11	house 1
REDD4	60	na	US	Spring	Weekday	na	W	27.04.11	27.04.11	house 1
REDD5	900	na	US	Spring	1 week	na	W	26.04.11	02.05.11	house 2
SmartB1	30	4	US	Spring	Weekend	y	W	10.06.12	10.06.12	home B
SmartB2	30	4	US	Spring	Weekday	y	W	11.06.12	11.06.12	home B
CER1	1800	2	EU	Summer	2 weeks	n	kWh	14.07.09	28.07.09	house1015
CER2	1800	2	EU	Winter	2 weeks	n	kWh	01.12.09	15.12.09	house1022
CER3	1800	6	EU	Summer	2 weeks	n	kWh	14.07.09	28.07.09	house1045
CER4	1800	6	EU	Winter	2 weeks	n	kWh	01.12.09	15.12.09	house1097
CER5	1800	6	EU	Summer	2 weeks	y	kWh	14.07.09	28.07.09	house1096
CER6	1800	6	EU	Winter	2 weeks	y	kWh	01.12.09	15.12.09	house1096

Table 15: Load profiles used in experiments.