Niklas Buescher*, Spyros Boukoros*, Stefan Bauregger, and Stefan Katzenbeisser

Two Is Not Enough: Privacy Assessment of Aggregation Schemes in Smart Metering

Abstract: The widespread deployment of smart meters that frequently report energy consumption information, is a known threat to consumers' privacy. Many promising privacy protection mechanisms based on secure aggregation schemes have been proposed. Even though these schemes are cryptographically secure, the energy provider has access to the plaintext aggregated power consumption. A privacy trade-off exists between the size of the aggregation scheme and the personal data that might be leaked, where smaller aggregation sizes leak more personal data. Recently, a UK industrial body has studied this privacy trade-off and identified that two smart meters forming an aggregate, are sufficient to achieve privacy. In this work, we challenge this study and investigate which aggregation sizes are sufficient to achieve privacy in the smart grid. Therefore, we propose a flexible, yet formal privacy metric using a cryptographic game based definition. Studying publicly-available, real world energy consumption datasets with various temporal resolutions, ranging from minutes to hourly intervals, we show that a typical household can be identified with very high probability. For example, we observe a 50% advantage over random guessing in identifying households for an aggregation size of 20 households with a 15-minutes reporting interval. Furthermore, our results indicate that single appliances can be identified with significant probability in aggregation sizes up to 10 households.

Keywords: smart grid, smart meters, privacy, aggregation, measurements, privacy metric

DOI 10.1515/popets-2017-0045
Received 2017-02-28; revised 2017-06-01; accepted 2017-06-02.

***Corresponding Author: Niklas Buescher:**
Technische Universität Darmstadt,
E-mail: buescher@seceng.informatik.tu-darmstadt.de
***Corresponding Author: Spyros Boukoros:**
Technische Universität Darmstadt,
E-mail: boukoros@seceng.informatik.tu-darmstadt.de
Stefan Bauregger: Technische Universität Darmstadt,
E-mail: s.bauregger@gmx.de
Stefan Katzenbeisser: Technische Universität Darmstadt,
E-mail: katzenbeisser@seceng.informatik.tu-darmstadt.de

1 Introduction

The smart grid is an advanced, multi-directional power grid, containing many smart meters and is regarded as the future of energy supply systems. Smart meters allow energy suppliers the permanent monitoring of their customers' energy consumption in order to reduce costs through more efficient and automatized power management. Besides the advantages for energy suppliers, the expected increase in renewable energy, electric cars and prosumers (consumers that also produce) in the following decades requires more reliable and flexible energy networks. In recent years, many countries worldwide introduced laws in order to expedite the use of smart meters in households. An example is the EU Directive 2006/32/EC, which asks all EU member states to provide "individual meters that accurately reflect the final customer's actual energy consumption and that provide information on actual time of use" for energy consumers [1].

Despite the economical and ecological advantages for the involved parties, the widespread information flow from energy consumers to producers is a serious threat to the consumers' privacy. The establishment of smart meters generates sensitive data to an extent that could not be reached using conventional meters. The continuous disclosure of energy consumption data in conjunction with algorithms like Non-Intrusive Appliance Load Monitoring (NALM) [2, 3], helps third parties to figure out daily routines of households, particular appliance uses, individuals' presence in a building or even the movie playing on the television [4]. If marketing agencies collude with energy suppliers, they can gather detailed information regarding household appliances [5]. It's not hard to imagine that these data can be used for targeted advertisement campaigns, new offers, etc. Criminals who are able to tap into a meter's data management system could predict when the occupants of a building will not be present [6]. Therefore, they can orchestrate their illegal activities more accurately. Even worse, mass surveillance is significantly enhanced. With little resources, interested malicious parties can observe the daily routines of millions of households.

These privacy concerns have been known to academia, industry and governmental institutions for years and therefore, a plethora of privacy mechanisms have been proposed to protect the consumers' privacy in the smart grid. The most promising and well researched privacy mechanisms are based on aggregation schemes, e.g., [7][8][9][10][11]. The core idea is to form groups of devices within the smart grid. Then, only the aggregated power consumption of the group is periodically reported to the energy supplier. The aggregate of a group can securely be computed using either a trusted third party, or preferably through cryptographic means, e.g., partial homomorphic encryption, secret sharing or other secure computation techniques. This solution has also been suggested as the-way-to-go by an expert group, set up by the European Commission [12].

Even though secure aggregation is technically solved, a major question has, to the best of our knowledge, barely been addressed. Namely, which aggregation size (number of smart meters in every group) is required to achieve privacy for consumers. During the smart meter roll-out in the United Kingdom, a study conducted by the industrial body "Energy Networks Association" concluded that aggregating the consumption of *only two smart meters* provides sufficient customer privacy [13]. However, this result seems to be elusive. It is not hard to imagine two households, where one person works during day shifts, while the other during night shifts. An aggregate of the two load profiles is protecting neither household because the two individuals will most likely be at home and use their appliances at different times.

In this paper, we shed light on the question of how many smart meters are required to provide privacy in an aggregation scheme.

Contributions

To study the privacy achieved by aggregation schemes, we first define a privacy measure in the form of a cryptographic game, using an indistinguishability notion. This game-based approach is inspired by other works in various domains [14][15], and is widely used in the field of cryptography. In a game played between adversary and challenger, the adversary, who can be of variable strength, has to identify a known load profile in an aggregate. The challenger's task is to ensure privacy by utilizing a smart meter aggregation scheme. The adversary's advantage over random guessing, is used as a measure for the achieved privacy. In contrast to previously proposed privacy metrics for the smart grid, the game can be applied to real world consumption data and offers a strong formalism.

With the newly developed metric, we analyze the privacy of individual households for different aggregation sizes. An application on real energy consumption data with more than 700 households shows that an average household is insufficiently protected in aggregates of two load profiles. On average, an adversary can distinguish two typical load profiles forming an aggregate with very high ($> 80\%$) probability, when reporting energy consumption information every 15 minutes. Even for aggregation sizes of 20 households, the adversarial advantage is (surprisingly!) 50%. We note, that these numbers represent the average advantage for all households. Extreme energy consumptions or additional auxiliary information regarding a household's energy consumption, make individual load profiles even more detectable.

Moreover, we examine the influence of further parameters, e.g., temporal resolution, on the detectability of an household within an aggregate. Finally, we show that single energy-hungry appliances can be detected in the aggregates of up to 10 households with significant advantage.

Outline

We discuss related work in § 2, before introducing our metric in § 3. Then, in § 4 the analyzed datasets and the evaluation approach are described. In § 5 we apply the proposed metric on real energy consumption datasets and present various case studies. Furthermore, in § 6 we study the diversity of energy consumption and the applicability of generated load profiles for privacy research. Finally, we conclude in § 7.

2 Related work

Privacy Mechanisms

A plethora of mechanisms have been proposed for the smart grid and a detailed survey is given by Jawurek et al. [16]. Here, we give an overview of the different directions of the existing solutions. The use of trusted third parties has been proposed in [17] and [18] in order to anonymize consumption data. Kalogridis et al. [19] propose to blur the load signature of individual smart meters – that is the unique patterns of every load profile, in order to achieve privacy, while Chim et al. [20] propose pseudo identities and signatures, using tamper resistant devices. Privacy mechanisms that mask energy

consumption using differential privacy have been proposed in [21][22][23][24]. A promising approach is based on aggregation schemes, where data is securely aggregated and sent to the energy supplier. Variations of homomorphic encryption are used in [8][11][10] to securely aggregate the data. Kursawe et al. [7] presented four different aggregation based privacy mechanisms using various cryptographic approaches, while Lu et al. [9] presented an aggregation scheme with enhanced performance on multidimensional data. Secure data aggregation is a promising approach for achieving anonymity. However, to the best of our knowledge, there is no research examining if the final aggregate offers privacy for individual households.

Privacy Metrics

The authors in [25] and [26] develop privacy metrics based on information disclosure and an attacker's estimation error respectively. Zhao et al. [27] proposed a metric for load signature moderation schemes, while Eibl et al. [28] examined the effect of adding Laplacian noise to aggregated smart meter load profiles. Shankar et al. [29] use the F-Test to measure and compare raw and noisy load profiles.

The privacy metric in this paper is based on the cryptographic game developed by Bohli et al. [15]. The goal of their game is the evaluation of privacy protection meachnism for a group of smart meters. The privacy level provided by the smart meter application is defined as the advantage of an adversary over random guessing, when distinguishing two groups of smart meters and their protected load profiles. In contrast, in this paper we create a cryptographic game to *isolate individuals in aggregation schemes*. In addition, we use real world datasets that have not been sanitized by any privacy mechanism, and measure the privacy gain using different aggregation loads.

3 Aggregation Privacy Model

As described, many cryptographic schemes have been proposed that allow the privacy preserving (provably secure) computation of smart meter aggregates. However, only a few metrics have been proposed that assess the effectiveness of smart grid privacy protection mechanisms in a formal and sound manner.

We propose such a framework, borrowing ideas from the 'Smart Grid Privacy Game', proposed by Bohli et al. [15]. We formalize data aggregation in the smart grid before we iteratively develop our privacy metric.

3.1 Smart Grid Aggregation Model

We make use of the following abstraction, which models the interaction between smart meters and an energy supplier. Informally speaking, when using privacy-preserving aggregation schemes, the energy supplier should learn the aggregated power consumption of groups of smart meters in every measurement period. For simplicity, we reduce our model to a single group of meters. Thus, the model consists of an energy supplier ES and a group (set) of smart meters $S = \{s_1, s_2, \ldots, s_n\}$ with $n > 1$. For practicality, we further assume a virtual party, the aggregator V, which connects all smart meters in S with the ES. In practice, this aggregator can either be instantiated by a trusted third party or by a cryptographic aggregation protocol, run between the smart meters. Moreover, a discrete notion of time $T = \{1, 2, 3, ...\}$ is used. In each time period $t \in T$, every smart meter s_i is attributed with a power consumption value $e_{i,t} \in R$, where R is the set of possible readings from a power consumption meter. Furthermore, we refer to consecutive consumption values as load profile. We denote a load profile of length l for a single smart meter s_i with $\hat{\mathrm{e}}_i(l) = (e_{i,1}, e_{i,2}, \ldots, e_{i,l})$. In every time period, all smart meters report their consumption to aggregator V, who computes the sum of all consumption values $a_t = \sum_{i=1}^{n} e_{i,t}$ and finally reports a_t to ES. We remark that we do not model further knowledge of the ES explicitly, yet consider background knowledge of any malicious adversary implicitly through the metric proposed in the next subsections.

3.2 Requirements of Privacy Notions for Aggregation in the Smart Grid

To assess the privacy protection offered by aggregation schemes in the smart grid, we identify the following requirements. A privacy notion / metric (even though it is not a metric in the mathematical sense) that allows to measure privacy leakage in aggregation schemes, should

- provide a strong formalism that allows reasoning about the provided privacy level, e.g., should allow to compute bounds; and should preferably
- allow to reason about practical attacks, i.e., it should be possible to show that these (with a certain probability) will fail.

Moreover, for a study of the trade-off between utility and privacy of the aggregated data, such a metric should:

– provide an adequate adversarial modeling. Hence, it should consider a powerful adversary. Yet, the adversary's power should not be overestimated in order to achieve realistic assessments and to maximize utility.

Achieving an adequate modeling of the adversary, especially its background knowledge, which defines its strength, is a challenging task, which is discussed in more detail in the next subsections.

3.3 Smart Grid Privacy Model

We define privacy for aggregation schemes using an indistinguishability notion. More precisely, we follow the idea of Bohli et al. [15] and use a game based definition. The core idea is to define privacy as the hardness to distinguish two load profiles known to the adversary in an aggregate. Informally speaking, the better the adversary in distinguishing profiles in aggregates, the weaker the privacy protection of individual households is in the aggregate. The strength of such a game based privacy notion is that it allows the modeling of arbitrary adversarial background knowledge, enabling us to model realistic and powerful attackers.

Formal Privacy Game

The basic game is illustrated in Figure 1. First, challenger and adversary agree on a load profile generator E_{gen}, the number of smart meters in the aggregate m, and the load profiles' length l. E_{gen} can either be a set of load profiles, e.g., from a real world consumption data set, or a sampling function that samples (realistic) load profiles from a probability distribution. After the initial setup phase, the adversary chooses (or samples, as described in the next paragraph) two load profiles $\hat{e}_0$ and $\hat{e}_1$ of length l from E_{gen}, which are then sent to the challenger. The challenger draws a random bit $r \in \{0,1\}$, samples $m-1$ further load profiles $\hat{e}_2, \hat{e}_3, \ldots \hat{e}_m$, and computes their aggregate $\hat{e}_a = \hat{e}_r + \hat{e}_2 + \ldots \hat{e}_m$. The aggregate is sent to the adversary who computes a decision function $f_{dec}(\hat{e}_a, \hat{e}_0, \hat{e}_1)$ that returns a bit $g \in \{0,1\}$, representing the guess whether $\hat{e}_0$ or $\hat{e}_1$ is contained in the aggregate. On a correct guess, the challenger outputs true, and false otherwise. We refer to the game as privacy aggregation game (AggG). The privacy of an aggregation scheme can be measured by the chances of an adversary in winning AggG. As in [15], we formally define the advantage of an adversary A for a given load profile generator E_{gen}, a number of smart meters m and load profile length l as the advantage over random guessing:

$$\mathbf{Adv}^{A}_{\mathrm{AggG}}(E_{gen}, m, l) = \\ |\Pr[\mathrm{AggG}^{A}(E_{gen}, m, l, r=0)=0] \\ - \Pr[\mathrm{AggG}^{A}(E_{gen}, m, l, r=1)=0]|.$$

Practical Privacy Notion

Assuming an adversary with an optimal decision function, the outcome of one instance of the privacy game mainly depends on two aspects. Namely, it depends on the load profiles *chosen by the adversary* and the load profiles *sampled by the challenger*. For example, assuming two load profiles with very distinct (visual) shape chosen by the adversary and load profiles with a flat shape sampled by the challenger, these distinct shape of the chosen load profiles may also become visible in the aggregate and allows a decision with high certainty. Thus, the adversary's advantage in the privacy game noticeably depends on the load profile generator E_{gen}, namely, how distinct the generated load profiles are and how these are distributed, as the advantage is computed over all possible aggregates.

An adversary A maximizes its advantage by choosing load profiles that are the most distinct. Computing the maximum possible advantage allows to determine bounds on the privacy leakage and resembles the scenario for the worst case consumer with a very distinct energy consumption. We refer to this advantage as ($\mathbf{Adv}^{A,max}_{\mathrm{AggG}}$). However, this notation might overestimate the privacy leakage for the average consumer, whose consumption is more similar to the average energy consumption of other consumers. Therefore, we introduce a second interpretations of AggG, which is the average advantage over all combinations of load profiles that can be 'chosen' by A ($\mathbf{Adv}^{A,avg}_{\mathrm{AggG}}$).

General Applicability

Revisiting the requirements of a privacy metric for aggregation, we observe that the AggG provides a strong formalism. Moreover, the applied indistinguishability notion is powerful, as it allows to model arbitrary, yet realistic background knowledge (load profiles are chosen from E_{gen}) of the adversary.

To illustrate the applicability of AggG, we consider the following exemplary privacy violation. The question whether it is possible to infer from a given aggregate that a consumer is at home during daytime can be modeled in AggG by choosing a load profile representing this consumption as $\hat{e}_0$ and a different typical load profile where the consumer is not at home as $\hat{e}_1$. A significant advantage in AggG indicates that a malicious energy supplier is able to answer this question with some certainty. A further practical attack, which can be modeled with the AggG is illustrated in Section 5, where we show that individual appliances can be detected in an aggregate with their signature.

On a first glance the game based definition with the precise knowledge of $\hat{e}_0$ and $\hat{e}_1$ might seem as overestimating the adversaries capabilities. However, in practice energy suppliers have access to a significant amount of external information that can be very close to the knowledge of precise load profiles. For example, suppliers have knowledge about:

- households contained in an aggregate (technical requirement for most schemes)
- past load profiles of all aggregators
- current and past montly billing information for every smart meter and specific time charges
- weather conditions, etc.,

Moreover, we note that the adversary A in the AggG is given almost no background information on the energy consumption of the other households contained in the aggregate. A only knows that the aggregate is sampled from a subset of realistic load profiles. In practice, it is not unreasonable to assume, that a malicious supplier has further background information, e.g., to the average power consumption of multiple households contained in the aggregate, because these are also customers that periodically report their consumption for billing purposes.

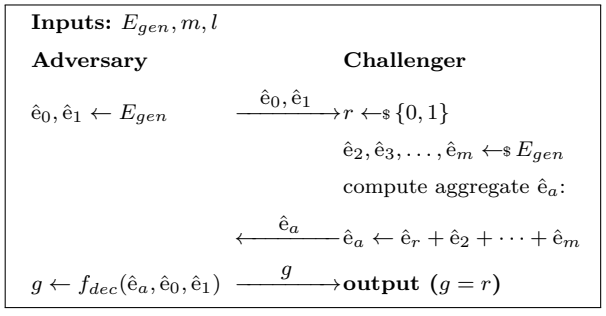

Fig. 1. Basic privacy game for aggregation schemes (AggG) in the smart grid.

Due to these reasons, we consider the proposed metric as well suited to show which aggregation sizes are insufficient and risk the loss of privacy.

Further Application - Membership Disclosure

To further illustrate the versatility of the proposed metric, we describe how the indistinguishablity based notion can be used to evaluate membership disclosure, i.e., to answer the question whether household x is contained in an aggregate or not. This can be evaluated by adapting the AggG, such that the adversary only samples one load profile $\hat{e}_0 \leftarrow E_{gen}$ and the challenger samples the other profile $\hat{e}_1 \leftarrow E_{gen}$, which is consequently unknown to the adversary. The rest of the game can be left unchanged, and the adversarial advantage is computed as the advantage over random guessing, whether $\hat{e}_0$ or $\hat{e}_1$ is contained in the aggregate.

We note that the membership game is at least as hard as the indistinguishability game. Given an adversary that can win the membership game, we can construct an adversary that is able to win the indistinguishability game with the same advantage. In order to decide which of two known profiles has been used in the aggregate of the indistinguishability game, an adversary could use the membership distinguisher to decide whether $\hat{e}_0$ is contained in the aggregate or not. The probability that $\hat{e}_0$ is in the aggregate is the same in both games. Hence, the adversary's advantage is identical to the advantage of the adversary in the membership game. Furthermore, given the fact that inverse reduction is impossible[1], the membership game is strictly harder than the indistinguishability game. However, we observe that it is possible to construct a practical heuristic $f^{\text{mem}}_{dec}(\hat{e}_a, \hat{e}_0)$ for the membership game, given a decision function $f^{\text{ind}}_{dec}(\hat{e}_a, \hat{e}_0, \hat{e}_1)$ for the indistinguishability game. Even without access to $\hat{e}_1$, an adversary in the membership game can repeatedly, i.e., a number of times k, invoke $f^{\text{ind}}_{dec}(\hat{e}_a, \hat{e}_0, \hat{e}_r)$ with a new randomly sampled $\hat{e}_r$. Using a majority voting $\sum_{i=0}^{k-1} f^{\text{ind}}_{dec}(\hat{e}_a, \hat{e}_0, \hat{e}_r) > k/2$, adversary A decides for $\hat{e}_1$ or otherwise for $\hat{e}_0$. We give an experimental evaluation of this privacy question and heuristic in Section 5.

1 Constructing a counter example is trivial: Assuming that all but one load profiles, which can be generated from E_{gen}, are flat, i.e., constant, then an adversary in the indistinguishability game has roughly twice the advantage to observe the non-constant load profile than an adversary for the membership game, who only gets to see a single load profile.

4 Methodology

4.1 Smart Grid Datasets

To identify aggregation sizes that provide sufficient privacy, we apply the privacy game on multiple real world energy consumption datasets. In Table 1, an overview of the datasets used in this work is given. The datasets have mainly been made available for energy disaggregation research. To the best of our knowledge, these are the largest publicly available datasets regarding the number of load profiles. We observe that the datasets have different geographical origins, as well as different measurement set ups, e.g., resolutions. Moreover, we remark that the datasets use different types of power measurement including active, reactive and apparent power [30]. Therefore, in most case studies we distinguish between datasets and study them separately.

Some datasets, e.g., *Dataport* and *UMASS*, contain several hundreds of households, whereas others, e.g., *AMPds*, focus on a single household for a large period of time. Unfortunately, only the *Dataport* and *GOVAU* dataset contain consumption data for more than 6 smart meters over multiple days.

Furthermore, most datasets require preprocessing, as they contain up to 10% incomplete or unusable (e.g., NAN) load profiles due to the experimental nature of energy consumption recording [31]. We consider a load profile to be complete if at least one sample is recorded in every sampling period required for a case study. Incomplete load profiles have been removed from all studies. The difference in the number of load profiles between complete and incomplete data is shown in Table 2. Note that the number of (complete) load profiles for each building in the same dataset may differ, therefore the total number of load profiles is given.

Name	Origin	Households	LPs/Hh	Resolution
Dataport [32]	US	707	647	15 min
Redd [31]	US	5	7	1 s
AMPds [33]	Canada	1	726	1 min
ECO [34, 35]	Switzerland	6	192	1 min
UCI [36]	France	1	1358	1 min
GOVAU [37]	Australia	31	406	30 min
UMASS [38]	US	376	1	1 min

Table 1. Datasets used for the analysis. Presented are the geographical origin, the number of households measured in each dataset, the average number of load profiles that have been recorded for each household, and the sampling resolution.

4.2 Evaluation Approach

To identify an aggregation size that protects the consumer's privacy, an implementation on the privacy game was created. To handle most of the datasets, we rely on the NILMTK framework [39], which has been developed to study energy disaggregation algorithms (NALM). NILMTK provides converters for most of the aforementioned datasets into a consistent data representation. The adversary is modeled in the form of a decision function f_{dec} that decides between two chosen load profiles $\hat{e}_0$ and $\hat{e}_1$. Different decision functions, which use a variety of heuristics, are introduced in the next section. For all case studies presented in Section 5, our implementation applies the following algorithm:

1. For the analysis, a dataset, an aggregation size m, a temporal resolution σ (the sampling frequency, e.g., $\sigma = 15$ min), an adversarial strategy (decision function f_{dec}), and a number of iterations N (e.g., $N = 5000$) are chosen.
2. The dataset is loaded. A dataset consists of multiple households with continuous load samples over one or multiple time periods. The load samples are grouped in load profiles of fixed start and end time. If not stated otherwise, each load profile starts at midnight with a duration of 24 hours in all experiments.
3. Next, all incomplete load profiles, i.e., load profiles that do not have at least one load sample per sampling period, are removed.
4. If the input dataset is more granular than the chosen resolution, the resolution of all load profiles is reduced, by temporal aggregation of consecutive load samples.
5. Two different households are selected from the dataset uniformly at random. From the two house-

-	Buildings		Load profiles		
Name	complete	total	complete	total	usable
Dataport	707	729	458048	474523	96.52%
Redd	6	6	53	236	22.45%
Ampds	1	1	726	730	99.45%
ECO	6	6	1196	1337	89.45%
UCI	1	1	1405	1440	97.56%
GOVAU	31	31	12606	12917	97.59%
UMASS	377	377	367	377	97.34%

Table 2. The number of buildings in each dataset that have at least one complete load profile and the total number of (complete) load profiles per dataset for a sampling resolution of 15min. The fraction of the usable against the total number of load profiles is displayed.

holds, one load profile is sampled for each household. The sampled load profiles are labeled as $\hat{e}_0$ and $\hat{e}_1$. This process ensures that even though different households might have a different number of load profiles, all households are represented equally in the result.

6. Analogously, $m-1$ load profiles are selected from the remaining households. A random bit $r \in \{0,1\}$ is sampled and the $m-1$ load profiles are summed up and added to $\hat{e}_r$ to create an aggregated load profile $\hat{e}_a$.
7. The decision function f_{dec} is evaluated on $\hat{e}_a, \hat{e}_0$ and $\hat{e}_1$.
8. If f_{dec} decided correctly (i.e., $f_{dec}(\hat{e}_a, \hat{e}_0, \hat{e}_1) = r$) a correct guess is recorded.
9. Steps 5-8 are repeated N times. Afterwards, the adversarial advantage values are computed as:

$$\mathbf{Adv}_{\mathrm{AggG}}^{f_{dec},avg}(m) = \left| \frac{correct\ guesses}{N} - 0.5 \right| \cdot 2$$

4.3 Decision Functions

For two given load profiles $\hat{e}_0$ and $\hat{e}_1$, an adversary in AggG has to decide which of the two is more likely contained in the aggregated load profile $\hat{e}_a$. In practice, finding an optimal decision is a hard computational problem, as an optimal distinguisher has to decide according to the maximum likelihood over all possible combinations of load profiles. Therefore, we focus on studying four heuristics and show in Section 5 that the described (comparably simple) heuristics are sufficient to identify load profiles in the aggregate. For better comparison, the aggregated load profiles are first normalized by the aggregation size: $\hat{e}_a \leftarrow \hat{e}_a/m$. The chosen decision functions are based on i) the *Mean Squared Error (MSE)*, ii) the *Pearson correlation*, iii) *peak detection* and iv) a *combined method* based on Pearson correlation and peak detection. These heuristics have been chosen, as they all allow to measure a distance between two time series and follow different approaches.

In i) the MSE is computed as the pairwise squared difference between load samples, hence, $f_{dec}^{\mathrm{MSE}}(\hat{e}_a, \hat{e}_0, \hat{e}_1)$ decides for $\hat{e}_0$ if $\mathrm{MSE}(\hat{e}_0, \hat{e}_a) < \mathrm{MSE}(\hat{e}_1, \hat{e}_a)$.

The Pearson correlation also considers the trend of the compared load profile and ii) is decided by the higher correlation, hence, $f_{dec}^{\mathrm{corr}}(\hat{e}_a, \hat{e}_0, \hat{e}_1)$ decides for $\hat{e}_0$ if $\mathrm{corr}(\hat{e}_0, \hat{e}_a) > \mathrm{corr}(\hat{e}_1, \hat{e}_a)$.

In iii) the relative peaks of each load profile $\hat{e}_a, \hat{e}_0, \hat{e}_1$, are determined and $f_{dec}^{\mathrm{peak}}(\hat{e}_a, \hat{e}_0, \hat{e}_1)$ decides according the most common peaks between $\hat{e}_0$ and $\hat{e}_a$, or $\hat{e}_1$ and $\hat{e}_a$. Peak detection is a promising approach, as it considers the most significant features of a load profile that (in our expectation) could also be visible in an aggregate. For the peak extraction we follow a simple approach, where a window around every sample of length ± 1 is selected. If a sample has a value higher than its neighbors it will be considered as a peak. For illustration, in Figure 2 a slice of an exemplary load profile $\hat{e}_0$ and an aggregated load profile $\hat{e}_a$, the identified peaks, and the windows of size three around each peak are shown. In the shown slice the load profiles only share one peak at 01:00 o'clock.

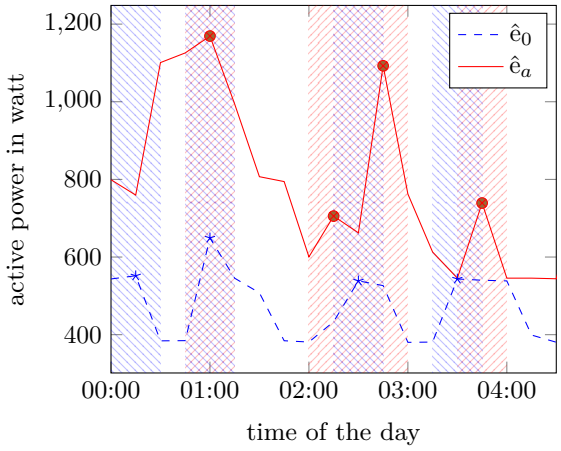

Fig. 2. Shown is a slice of two exemplary load profiles ($\hat{e}_0$, $\hat{e}_a$) and marked are the identified peaks of each load profile. Moreover, a window of size 3 is drawn around each load profile, which has been used to identify the peaks (cf., Section 4.3). The time resolution is 15 minutes.

The decision function iv) combines peak detection and correlation with the idea that the shape of the load profile surrounding the peaks carries more information than the peak itself. Therefore, in iv) all peaks of $\hat{e}_0, \hat{e}_1$ and $\hat{e}_a$ are computed. Then, the union of the peaks between of $\hat{e}_0$ ($\hat{e}_1$) and $\hat{e}_a$ is formed. Afterwards, the Pearson correlation is computed for a surrounding window of a fixed length of samples around every peak, e.g. We identified a window of ± 5 (i.e., windows size is 11) as the best heuristic for 15 minute readings (a detailed analysis on the window size is given in the next section). The decision function $f_{dec}^{comb}(\hat{e}_a, \hat{e}_0, \hat{e}_1)$ decides according to the higher *mean correlation* between all windows of $\hat{e}_0$ and $\hat{e}_a$ or $\hat{e}_1$ and $\hat{e}_a$.

5 Case Studies

To analyze the privacy protection offered by aggregation schemes, we perform multiple case studies. First, we show for multiple datasets that the simple decision functions are sufficient to identify load profiles within aggregates of sizes ranging from two to hundreds of buildings. Moreover, we study the impact of temporal resolution, load profile length and daytimes on the distinguishing advantage. Then, we show that single appliances can be detected in aggregates consisting of load profiles from multiple households. Finally, we investigate membership disclosure in aggregates.

How effective are decision functions in identifying load profiles in an aggregate?
We evaluate the effectiveness of the proposed decision functions by comparing them in the privacy game over $N = 5000$ simulations with different power measurements and time resolutions. A decision function is effective, if the advantage over random guessing is significant. Goal of the decision functions, as described in Section 4.2, is to identify the correct profile ($\hat{e}_0$ or $\hat{e}_1$) contained in the aggregate. First, we compare the average advantage of all proposed decision functions on the Dataport dataset with a sampling resolution of 15 minutes, shown in Figure 3. We observe that all heuristics can identify the correct load profile for small aggregation sizes with significant advantage. More precisely, for only 2 load profiles, all methods have an advantage of more than 75%. The Pearson correlation and peak detection heuristics perform similar over all evaluated aggregation sizes, whereas the proposed combination is the most powerful distinguisher. For aggregation sizes larger than 10, its advantage is than twice the best advantage of the other three heuristics.

As already described in Section 4.3, the combined method computes the Pearson correlation for a window of load sample around all detected peaks. The window size, which influences the distinguishing advantage, is empirically evaluated in Figure 4. Plotted is the averaged advantage for different window sizes for the combined method on the Dataport dataset for three different sampling resolutions (15 min, 60 min, and 120 min) over aggregation sizes from 2 to 30. We observe that the best results are achieved for a moderately sized window, e.g., 10 load samples for a 15 minute reading. Moreover, we observe that a more granular sampling resolution requires more load samples to be contained in the window to achieve the best advantage.

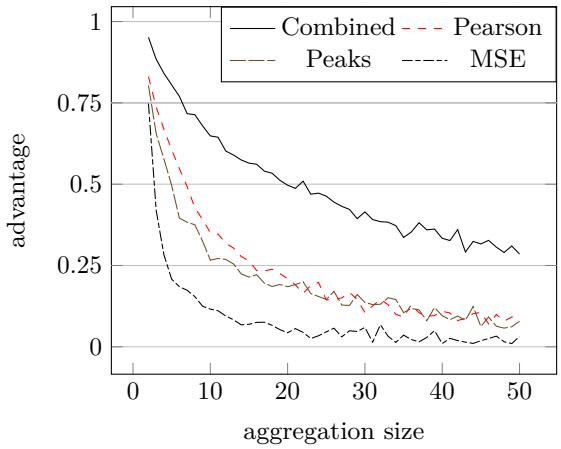

Fig. 3. Comparison of the four decision functions, based on averaged adversarial advantage for different aggregation sizes (Dataport, 15 min resolution).

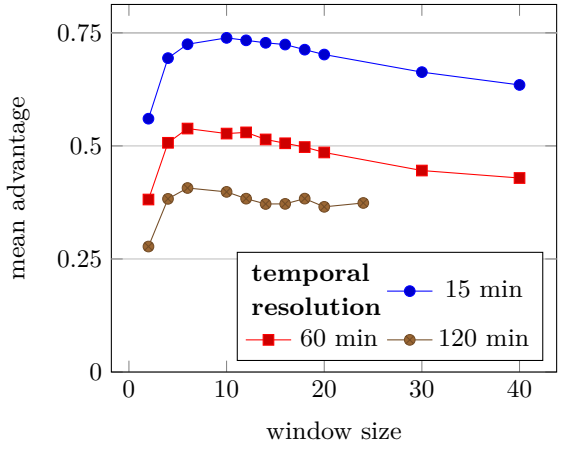

Fig. 4. Evaluating the window parameter of the Combined decision function (Dataport, $m = \{2, ..., 30\}$, 15 min resolution).

Which parameters influence the privacy game?
Dataset dependency. The results of an empirical analysis commonly depend on the dataset used. To show that the difference in adversarial advantage is rather small between the datasets, we compare the distinguishing advantage between the datasets Dataport and the other two largest datasets (GOVAU and UMASS), for the combined decision function in Figure 5. We observe that the power consumption in UMASS is noticeable more distinguishable by the combined decision function than the GOVAU and Dataport dataset, which share a very similar (in-)distinguishability for increasing aggregation sizes.

Furthermore, we can illustrate a similar behavior of all four decision functions on a union of all load profiles from all datasets. To sample a load profile in this experiment, we first sample a dataset, then a household

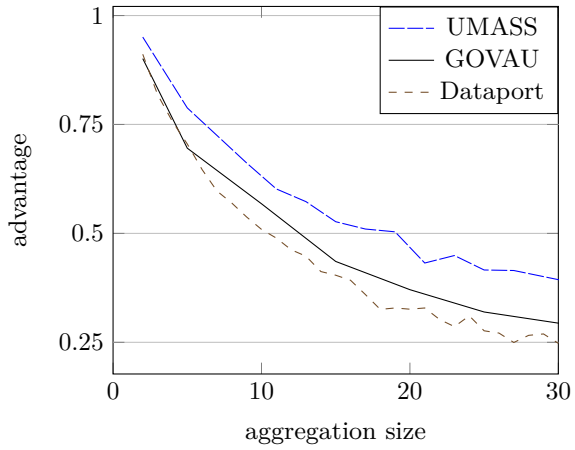

Fig. 5. Comparison of the averaged distinguishing advantage between the UMASS, GOVAU, and Dataport dataset for different aggregation sizes, when using the combined decision function (30 min resolution).

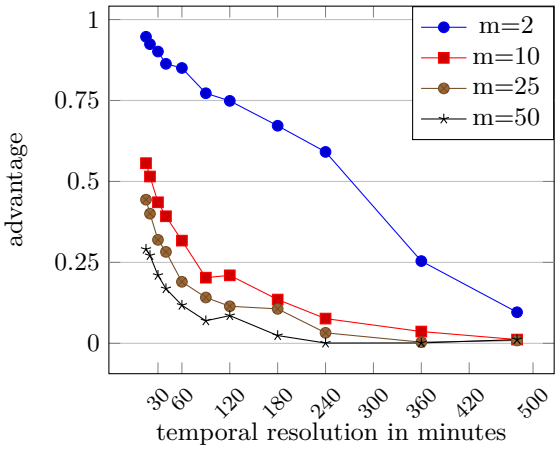

Fig. 7. Comparison of the impact of different temporal resolutions when using different aggregation sizes m. Measured is averaged advantage using the combined heuristic (Dataport).

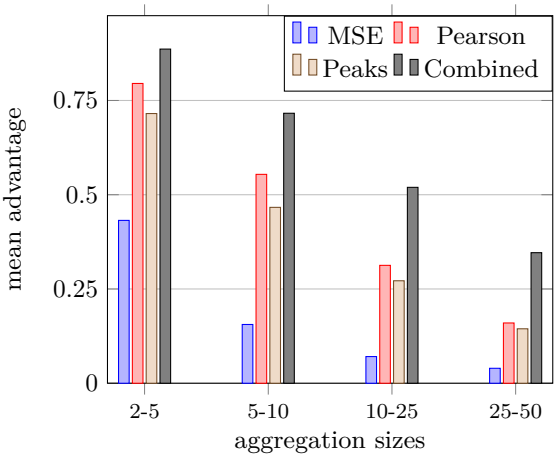

Fig. 6. Comparison of the decision functions based on the averaged adversarial advantage, using different aggregation sizes, using profiles from all datasets (30 min resolution).

(uniformly among the dataset) and then a load profile (uniformly among the household). This guarantees an equal representation of datasets and households. We acknowledge that consequently some load profiles have more impact on the results than others, unfortunately the limited number of large datasets does not allow for a better experimental setup. The distinguishing advantages in this experiment are shown for different groups of aggregation sizes in Figure 6. We observe a similar distinguishing advantage as when studying datasets independently. Moreover, as before the combined decision function outperforms the others in every scenario, and hence will be used as the main decision function in all remainder of this section.

In summary, the datasets show diversity in their load profiles, which is also visible in the AggG. However, the difference in the results between the datasets (also influenced by the empirical nature of our approach) is only marginal when deriving qualitative statements on the individuals privacy.

Temporal Resolution. The rate at which each smart meter reports its values, is a crucial factor for privacy. More frequent reports enable NALM algorithms to work with higher accuracy and extract more information. Hence, we expect that higher sampling resolutions are less privacy friendly. In Figure 7, we present the advantage for different aggregation sizes with different sampling resolutions when using the combined decision function on the Dataport dataset. Clearly, the advantage is higher for more frequent reports and smaller aggregation sizes. In aggregations with 10 or more load profiles ($m \geq 10$), the advantages differ only by a small factor independent of the temporal resolution. When only two households are aggregated, a significant advantage of 50% is observed for a temporal resolution as low as 4 hours. This confirms what we intuitively expected, namely that distinguishability increases with more frequent reports. In addition, it is clear from measurements that the aggregate of two load profiles is not enough to provide privacy, even for a very low sampling resolution.

Influence of Different Daytimes. In previous evaluations, we studied load profiles of 24h length . In this section, we examine if *different daytimes* affect the model's accuracy. The load profiles of the Dataport dataset were split in four parts according their daytime. Those were, night time (0:00-6:00), mornings (6:00-12:00), af-

ternoons (12:00-18:00) and evenings (18:00-24:00). We study aggregation sizes ranging from 2 to 50 households, and fix the sampling resolution to 15 minutes. We performed $N = 5,000$ simulations of the privacy game for each period. In Table 3, the average advantage for the four different daytimes, as well as for the whole day is presented. In summary, we observe only marginal differences between the different daytimes, but as expected, a 24 hour load profile allows for better distinguishability than isolated daytimes.

Daytime	$m = 2$	$m = 5$	$m = 10$	$m = 30$	$m = 50$
Night	**0.811**	**0.574**	**0.411**	**0.233**	**0.184**
Morning	**0.836**	**0.603**	**0.436**	**0.232**	**0.157**
Afternoon	**0.792**	**0.566**	**0.418**	**0.237**	**0.157**
Evening	**0.800**	**0.566**	**0.410**	**0.234**	**0.152**
Day - 24h	**0.947**	**0.793**	**0.634**	**0.396**	**0.29**

Table 3. Average advantage when distinguishing load profiles of 6h length, using the Dataport dataset (15 min resolution), compared with the advantage when distinguishing a load profile of 24h length.

How many households are required to achieve privacy?

Bigger aggregation sizes lead to better privacy for individual households. However, an arbitrary increase in aggregation size defeats the purpose of smart meters, which should be able to monitor and predict the consumption in order to distribute energy more efficiently. Thus, an upper bound exists on how many households should be in an aggregate report in order for the smart grid to retain some utility. Unfortunately, we have no (reasonable) measure of utility, yet we can identify a marginal utility on the privacy protection. Applying the AggG with the combined heuristic on the Dataport dataset, which provides the largest number of load profiles and households, for aggregation sizes of up to 700, we can infer, which aggregation size is needed to achieve a certain level of privacy (distinguishing advantage over random guessing) shown in Figure 8 for a 15-minute sampling resolution. We denote with m the size of the aggregate and with δ the average adversarial advantage. The shape of the curve can be used to analyze the marginal utility. The curve is very steep up to a privacy level of $\delta = 0.5$, which is reached in the experiment with an aggregation size of $m = 23$. At a privacy level of $\delta = 0.2$ ($m = 92$) the curve significantly starts to flatten out with only marginal improvements in privacy after $m = 200$.

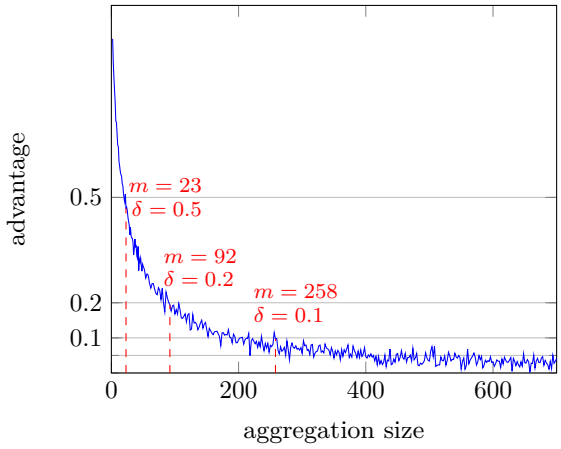

Fig. 8. Adversarial advantages achieved with the combined decision function for different aggregation sizes m. Marked with δ are interesting advantage levels (Dataport, 15 min resolution).

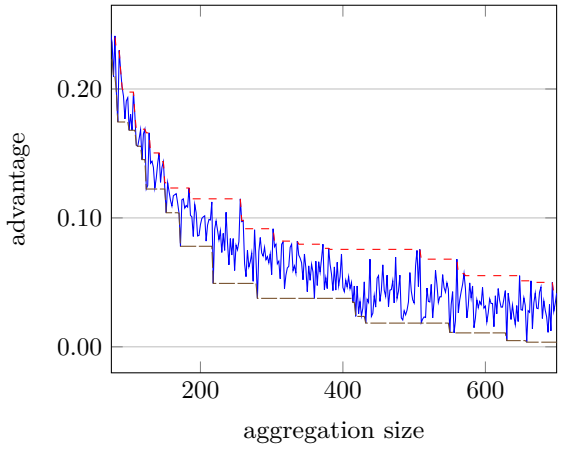

Fig. 9. Illustration of the imprecision in the experimental computation of the distinguishing advantage for aggregation sizes $m > 75$ over simulations with $N = 5000$ runs (Dataport, 15 min resolution).

We remark, that results for the distinguishing advantage for larger aggregation sizes, e.g., above 100, should be studied with a grain of salt. Even though, each data point is computed via a simulation over $N = 5000$ trials, it contains a noticeable error for larger aggregation sizes, which is illustrate in Figure 9. We observe that very similar aggregation sizes can show a noticeable variance in the distinguishing advantage.

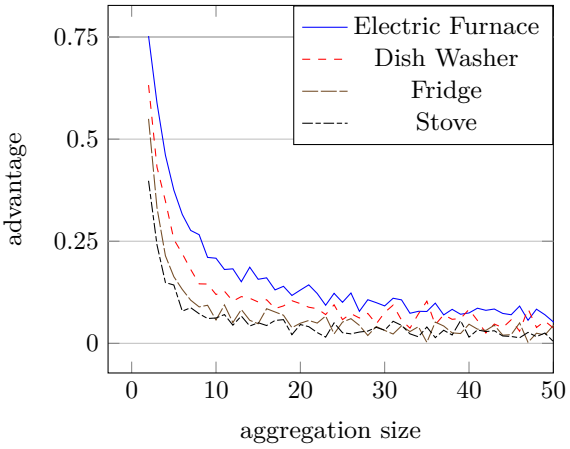

Fig. 10. The detectability of single appliances in an aggregate. Shown is the distinguishing advantage for the combined decision function (Dataport, 15 min resolution).

Are particular appliances detectable in an aggregation scheme?

Specific appliances create unique patters in load profiles. NALM algorithms can extract specific devices' usage in single households, by detecting those patterns. In order to examine if specific devices can be detected in the aggregation, we adjust the adversary's choices in the AggG. The first load profile $\hat{e}_0$ is sampled from the dataset, the second load profile is 'generated' by subtracting from the individually measured load profile $\hat{e}_0$ a single appliance. Hence, $\hat{e}_0$ and $\hat{e}_1$ only differ in the energy consumption of a single appliance.

Using the Dataport dataset, we study the AggG for different aggregation sizes and household appliances. In Figure 10, we present the average adversarial advantage over random guessing, for an electric furnace, a dish washer, a fridge and a stove. The results demonstrate that specific appliances, e.g., electric furnace, are detectable with significant advantage even in aggregates of size $m > 10$. As expected, the detection is more powerful when aggregation is small. A study of further devices and the resulting adversarial advantage is given in the appendix in Figure 18, where we illustrated the adversarial advantage when detecting various devices in the Dataport dataset, for aggregation sizes of $m = 5$, 10 and 25.

To identify specific patterns that make a load profile (of an appliance) distinguishable in an aggregate, we study various properties of appliances in the dataset, namely: mean load (when switched on), maximum load, the number of peaks, as well as the daily uptime and average load per peak. The results are illustrated in Table 4. The correlation between the characteristic properties and the detectability, using the Pearson correlation between the properties and the advantage per aggregation size is depicted in Figure 11.

Appliance	Mean load(W)	Max load (W)	#Peaks	Daily uptime	Load/Peak
Dish washer	**43.5**	**887.1**	**6.1**	**9%**	**504.5**
Electric furnace	**135.7**	**603.7**	**18.7**	**87%**	**305.8**
Fridge	**77.1**	**344.5**	**23.2**	**50%**	**143.9**
Stove	**55.0**	**1110.8**	**9.6**	**27%**	**440.7**

Table 4. Characteristic properties of particular appliances (average values).

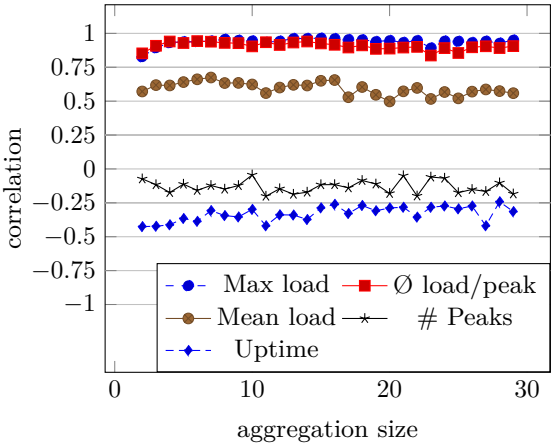

Fig. 11. Correlation between characteristic properties and detectability of appliances for different aggregation sizes.

We observe that the detectability of an appliance shows the largest correlation with the maximum load, followed by the average load per peak. The correlation between average mean load and the detectability is significantly lower, while the properties average daily uptime and number of peaks are negatively correlated to the privacy level. To conclude, not only households but also individual appliances that show consumption patterns with high peaks can be detected with minimal effort in aggregates of smaller size.

Can membership in an aggregate be disclosed using the same decision functions?

A further application of AggG is outlined in Section 3, namely whether an adversary could identify the existence of a single profile in an aggregate rather than distinguishing two known profiles. Hence, shifting the focus from an indistinguishability notion to a membership disclosure question. As described, the privacy game has

to be adapted in the following way; instead of having the adversary select two profiles and then try to distinguish which one is in the aggregate, he only samples one ($\hat{e}_0$). Then the challenger randomly samples a second one ($\hat{e}_1$), unknown to the adversary, and by flipping a coin decides which one of the two will used in the aggregate sent to the adversary. The adversary has then to guess, whether this profile is part of the aggregate or not. Using a similar experimental setup as before, we studied this questions for the Dataport dataset with aggregation sizes from $m = 2$ to 20 and $N = 1000$ simulations per aggregation sizes. Moreover, we used the randomized decision functions as described in Section 3 with $k = 100$ iterations. In Figure 12 the advantage for correct answering the membership question with the help of the two most effective heuristics, i.e., peak detection and the combined method is presented. In addition, the advantage in the indistinguishability notion for the same aggregation sizes is given.

Even though the adversary has less power in this game, and consequently, its advantage decreases compared to the indistinguishably game, we observe that the advantage remains significant for all aggregation sizes. A more surprising result is that that peak detection and the combined methods perform similarly for membership disclosure, in contrast with the cases previously examined. This again indicates that the peaks are the most robust feature to distinguish load profiles.

In summary, the AggG is very suited to also examine privacy under a different view point, i.e., membership disclosure, with small modification. Yet, even in the membership based privacy notion, very simple heuristics are able to achieve a significant advantage over random guessing for larger aggregation sizes.

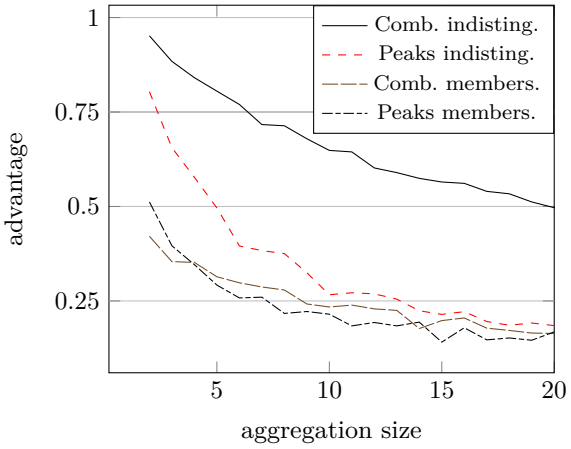

Fig. 12. Comparison of the modified privacy game where the adversary has knowledge of only one load profile (membership disclosure) vs when he distinguishes two profiles (indistinguishability notion). For each experiment, the distinguishing advantage for the two most effective heuristics (peak detection and combined method) for aggregation sizes of up to 20 households is displayed. (Dataport, 15 min resolution).

6 Dataset Analysis

Studying the detectability of individual load profiles with the help of the aggregation game, a question arises, whether a common 'universal' load profile exists. The existence of a universal load profile could be used to only consider the relative changes to the universal load profile as privacy relevant and thus, demand a reformulation of the privacy game. Therefore, in this section, we first study the differences between individual load profiles and their average from the dataset. Second, to overcome the very limited availability of real world energy consumption datasets, we study the applicability of load profile generators in privacy research for the smart grid.

Universal Load Profile

While datasets from different countries presumably differ in their average load profile due to differences in cultural and climatic preconditions, this does not apply to load profiles from similar climate zones and cultural environments. Unfortunately, the available data is insufficient to present an exhaustive analysis. Yet, when comparing the average load profile of the Dataport (707 households) and GOVAU (31 households) and UMASS(376 households), shown in Figure 13, similarities in shape can be identified. For example, comparatively low consumption values during night, and consumption peaks in the morning and evening hours are visible. We note that the different energy measurements (re-/active power) lead to noticeable differences in the individual consumption values and should therefore not be compared by their absolute value. Moreover, we observe significant more variance in the UMASS datasets, which only provides one load profile for every household.

Generally, we observe that individual load profiles can be quite distinct from the average of a dataset. This is illustrated in Figure 14, where the distribution of the mean squared error (MSE) between all individual load profiles and their average is illustrated for the Dataport dataset. Similarly, in Figure 15 the Pearson correla-

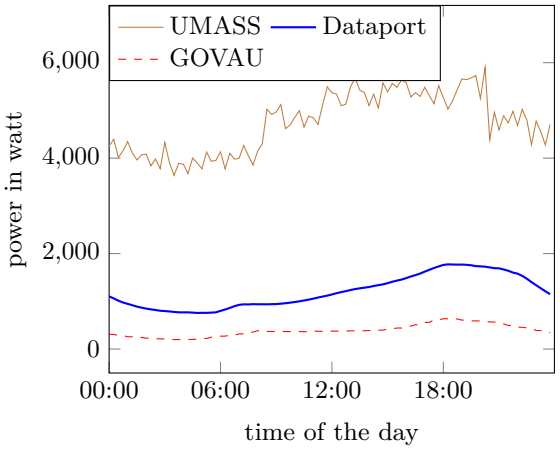

Fig. 13. Mean load profiles of the datasets UMASS, Dataport, and GOVAU.

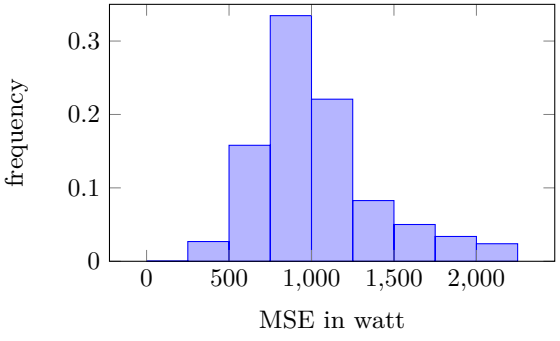

Fig. 14. Distribution of Mean Squared Error between individual load profiles and the average load profile in Dataport.

tion coefficient between individual load profiles and the dataset's average is plotted for the GOVAU and ECO dataset. Even though, there is a noticeable correlation to the datasets' averages, we also observe numerous outliers in both datasets, that are very different to the average. Thus, load profiles carry significantly more information than only small relative differences to the datasets average load profile. Moreover, Figure 15 also shows the correlation of each load profile to the household's average. This correlation is (expectable) higher than the correlation to the mean of the according dataset, yet also shows a significant variance between the load profiles from the same household.

In summary, we observe that (background) knowledge on the average load profile of a region or household could be used to improve the detectability of load profiles in an aggregate. Yet, the significant variance between load profiles illustrate that the protection of only small changes to an average is insufficient.

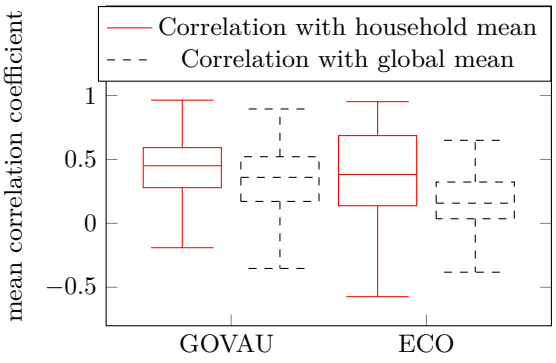

Fig. 15. Distribution of correlation between single load profiles and the datasets (GOVAU, Eco) or households average.

Load Profile Generators for Privacy Research

Load profile generators are tools that are able to simulate the energy consumption for a certain period based on an underlying model. The area of application ranges from studies concerning the effects of new technology on the energy consumption of households to forecasts like the determination of the national energy demand [40]. A natural question is if these generators are adequate for a privacy analysis and could be considered in subsequent work. In this work, we focus on studying energy consumption of individual households, and hence a load profile generator based on the so called bottom-up approach is promising. Bottom-up load profile generators aim at simulating the behavior of inhabitants of a household, that is modeled in the use of household appliances, e.g., cooking, heating or television or other activities. According to the simulated usage of appliances and (pre-recorded) appliance specific demand profiles, load profiles for households are generated. The model can be enhanced by external influences like temperature, holidays and geographic circumstances. Various bottom-up load profile generator have been proposed in [41]. Using the implementation of the *Loadprofile Generator* [42], we created a dataset containing 266 households with 365 load profiles per household, which is studied in the following paragraphs.

In Figure 16 the mean load profile of the generated dataset is shown. When comparing the dataset's mean to the mean load profiles of the datasets depicted in Figure 13 qualitatively, it is clear that all datasets share similarities and prominent features like the peaks at about 06:00 and after 18:00 can be found in both.

Similar to the previous analysis, we applied the privacy game to the generated dataset using a resolution of 15 minutes on aggregates of size $m = 2$ to 50. The results for all decision functions are illustrated in Figure 17. We

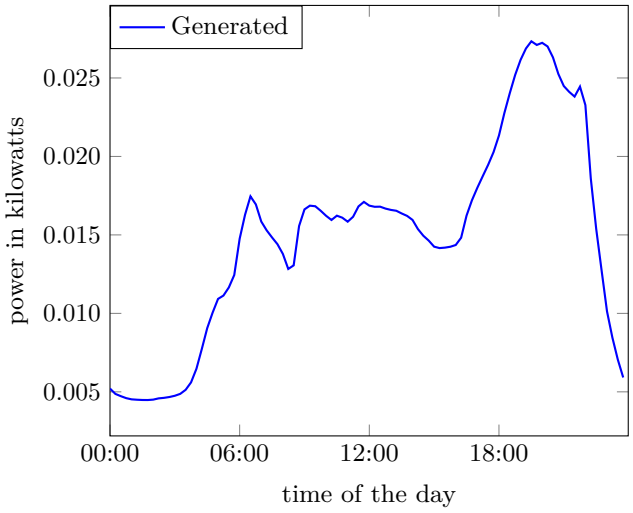

Fig. 16. Mean load profile for the generated dataset consisting of 266 households.

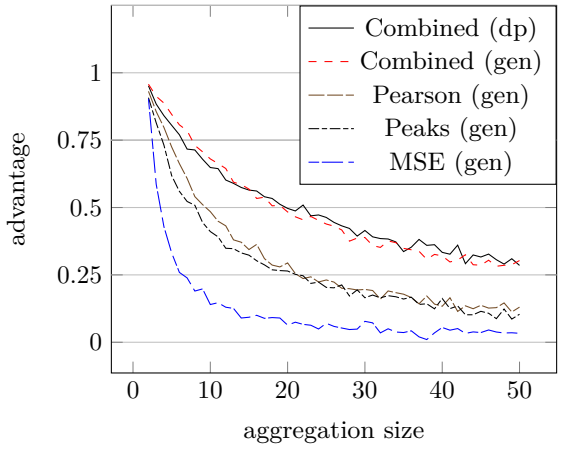

Fig. 17. Comparison of the decision functions for the generated dataset (gen). For comparison, the advantage of the combined decision function applied to the Dataport (dp) dataset is also plotted (15 min resolution).

observe that all decision functions show a similar behavior as on the real data sets, i.e., the combined decision function outperforms the rest, whereas the MSE shows the least distinguishing advantage. For better comparison, we plotted the results of the combined decision function when applied on the Dataport dataset. Also here, we observe no significant differences (in the distinguishing advantage) between the generated dataset and the real dataset. Hence, we conclude for future privacy studies on real world data, generated datasets seem to be a very promising alternative to real world datasets, whose availability is very limited.

7 Conclusions

In this work, we studied the privacy of single load profiles contained in an aggregate. Even though fixing an acceptable privacy loss (advantage) is more a philosophical question rather than a purely technical, it becomes obvious that an aggregation size in the single digit range seems to be far from being sufficient to provide privacy when assuming a 15 minute reporting interval. Even worse, it is safe to say that the privacy leakage is notably higher in practice than in our model. This is due to the fact that energy suppliers continuously record consumption information. Consequently, periodical behavior of households inhabitants (e.g., sleep cycle) will inevitably leak to the supplier.

We are convinced the secure aggregation is a powerful mechanism to protect privacy in the smart grid. However, the observation that load profiles can be detected in aggregates of more than 100 meters with significant advantage (>10%) demands that all parameters that influence the trade-off between utility and privacy, such as temporal resolution and aggregation sizes, should carefully be studied, before being blindly accepted as secure.

It remains an open question, to quantify utility in aggregation schemes. A utility measure would allow to study the utility and privacy trade-off, and thus practicality of aggregation, in more detail.

Acknowledgements

We would like to thank Alfredo Rial and all anonymous reviewers for their very helpful and constructive comments. This work has been funded by the DFG as part of project A1 within the RTG 2050 “Privacy and Trust for Mobile Users”, and within the CRC 1119 “CROSSING”.

References

[1] European Parliament and Council of the European Union, “Directive 2006/32/ec of the european parliament and of the council,” 2006.

[2] G. W. Hart, “Nonintrusive appliance load monitoring,” *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[3] H. Lam, G. Fung, and W. Lee, “A novel method to construct taxonomy electrical appliances based on load sig-

naturesof," *IEEE Transactions on Consumer Electronics*, vol. 53, no. 2, pp. 653–660, 2007.

[4] U. Greveler, B. Justus, and D. Loehr, "Forensic content detection through power consumption," in *Proceedings of IEEE International Conference on Communications, ICC 2012, Ottawa, ON, Canada, June 10-15, 2012*, pp. 6759–6763, 2012.

[5] C. Laughman, K. Lee, R. Cox, S. Shaw, S. Leeb, L. Norford, and P. Armstrong, "Power signature analysis," *IEEE power and energy magazine*, vol. 1, no. 2, pp. 56–63, 2003.

[6] J. I. Lerner and D. K. Mulligan, "Taking the'long view'on the fourth amendment: Stored records and the sanctity of the home," *Stanford Technology Law Review (STLR)*, vol. 3, 2008.

[7] K. Kursawe, G. Danezis, and M. Kohlweiss, "Privacy-friendly aggregation for the smart-grid," in *International Symposium on Privacy Enhancing Technologies Symposium*, pp. 175–191, Springer, 2011.

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

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

[10] F. Li, B. Luo, and P. Liu, "Secure information aggregation for smart grids using homomorphic encryption," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pp. 327–332, IEEE, 2010.

[11] F. D. Garcia and B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," in *International Workshop on Security and Trust Management*, pp. 226–238, Springer, 2010.

[12] Expert Group for Regulatory Recommendations for Privacy, Data Protection and cyber-security in the Smart Grid Environment, "Essential regulatory requirements and recommendations for data handling, data safety, and consumer protection," 2011.

[13] Energy Networks Association, "Smart meter aggregation assessment final report," 2015.

[14] S. Vaudenay, *On Privacy Models for RFID*, pp. 68–87. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007.

[15] J. M. Bohli, C. Sorge, and O. Ugus, "A privacy model for smart metering," in *2010 IEEE International Conference on Communications Workshops*, pp. 1–5, May 2010.

[16] M. Jawurek, F. Kerschbaum, and G. Danezis, "Privacy Technologies for Smart Grids - A Survey of Options," tech. rep., Microsoft Research - Tech Report - 2012 - 119, 2012.

[17] C. Efthymiou and G. Kalogridis, "Smart grid privacy via anonymization of smart metering data," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pp. 238–243, IEEE, 2010.

[18] H. S. Fhom, N. Kuntze, C. Rudolph, M. Cupelli, J. Liu, and A. Monti, "A user-centric privacy manager for future energy systems," in *Power System Technology (POWERCON), 2010 International Conference on*, pp. 1–7, IEEE, 2010.

[19] G. Kalogridis, C. Efthymiou, S. Z. Denic, T. A. Lewis, and R. Cepeda, "Privacy for smart meters: Towards undetectable appliance load signatures," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*, pp. 232–237, IEEE, 2010.

[20] T. W. Chim, S.-M. Yiu, L. C. Hui, and V. O. Li, "Pass: Privacy-preserving authentication scheme for smart grid network," in *Smart Grid Communications (SmartGridComm), 2011 IEEE International Conference on*, pp. 196–201, IEEE, 2011.

[21] P. Barbosa, A. Brito, and H. Almeida, "A technique to provide differential privacy for appliance usage in smart metering," *Information Sciences*, vol. 370, pp. 355–367, 2016.

[22] M. Backes and S. Meiser, "Differentially private smart metering with battery recharging," in *Data Privacy Management and Autonomous Spontaneous Security*, pp. 194–212, Springer, 2014.

[23] Z. Zhang, Z. Qin, L. Zhu, J. Weng, and K. Ren, "Cost-friendly differential privacy for smart meters: exploiting the dual roles of the noise," *IEEE Transactions on Smart Grid*, 2016.

[24] G. Ács and C. Castelluccia, "I have a dream!(differentially private smart metering)," in *International Workshop on Information Hiding*, pp. 118–132, Springer, 2011.

[25] M. A. Lisovich, D. K. Mulligan, and S. B. Wicker, "Inferring personal information from demand-response systems," *IEEE Security & Privacy*, vol. 8, no. 1, pp. 11–20, 2010.

[26] R. Dong, A. A. Cárdenas, L. J. Ratliff, H. Ohlsson, and S. S. Sastry, "Quantifying the utility-privacy tradeoff in the smart grid," *arXiv preprint arXiv:1406.2568*, 2014.

[27] J. Zhao, T. Jung, Y. Wang, and X. Li, "Achieving differential privacy of data disclosure in the smart grid," in *IEEE INFOCOM 2014-IEEE Conference on Computer Communications*, pp. 504–512, IEEE, 2014.

[28] G. Eibl and D. Engel, "Differential privacy for real smart metering data," *Computer Science - Research and Development*, pp. 1–10, 2016.

[29] L. Sankar, S. R. Rajagopalan, and S. Mohajer, "Smart meter privacy: A theoretical framework," *IEEE Transactions on Smart Grid*, vol. 4, no. 2, pp. 837–846, 2013.

[30] IEEE, "The authoritative dictionary of ieee standards terms, seventh edition," *IEEE Std 100-2000*, pp. 1–1362, Dec 2000.

[31] J. Z. Kolter and M. J. Johnson, "Redd: A public data set for energy disaggregation research," in *Workshop on Data Mining Applications in Sustainability (SIGKDD), San Diego, CA*, vol. 25, pp. 59–62, Citeseer, 2011.

[32] P. S. Inc., "Dataport Dataset." https://dataport.pecanstreet.org/. [Online; accessed July-2016].

[33] S. Makonin, B. Ellert, I. V. Bajic, and F. Popowich, "Electricity, water, and natural gas consumption of a residential house in Canada from 2012 to 2014," *Scientific Data*, vol. 3, no. 160037, pp. 1–12, 2016.

[34] W. Kleiminger, C. Beckel, and S. Santini, "Household occupancy monitoring using electricity meters," in *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 975–986, ACM, 2015.

[35] C. Beckel, W. Kleiminger, R. Cicchetti, T. Staake, and S. Santini, "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*, pp. 80–89, ACM, 2014.

[36] M. Lichman, "UCI machine learning repository," 2013.
[37] A. Government, "GOVAU Dataset." http://data.gov.au/dataset/sample-household-electricity-time-of-use-data. [Online; accessed July-2016].
[38] S. Barker, A. Mishra, D. Irwin, E. Cecchet, P. Shenoy, and J. Albrecht, "Smart*: An open data set and tools for enabling research in sustainable homes," *SustKDD, August*, vol. 111, p. 112, 2012.
[39] N. Batra, J. Kelly, O. Parson, H. Dutta, W. J. Knottenbelt, A. Rogers, A. Singh, and M. B. Srivastava, "NILMTK: an open source toolkit for non-intrusive load monitoring," in *The Fifth International Conference on Future Energy Systems, e-Energy '14, Cambridge, United Kingdom - June 11 - 13, 2014*, pp. 265–276, 2014.
[40] L. G. Swan and V. I. Ugursal, "Modeling of end-use energy consumption in the residential sector: A review of modeling techniques," *Renewable and sustainable energy reviews*, vol. 13, no. 8, pp. 1819–1835, 2009.
[41] N. D. Pflugradt, *Modellierung von Wasser und Energieverbräuchen in Haushalten*. dissertation, Technische Universität Chemnitz, 2016.
[42] www.loadprofilegenerator.de. [Online; accessed August 2016].

Appendix

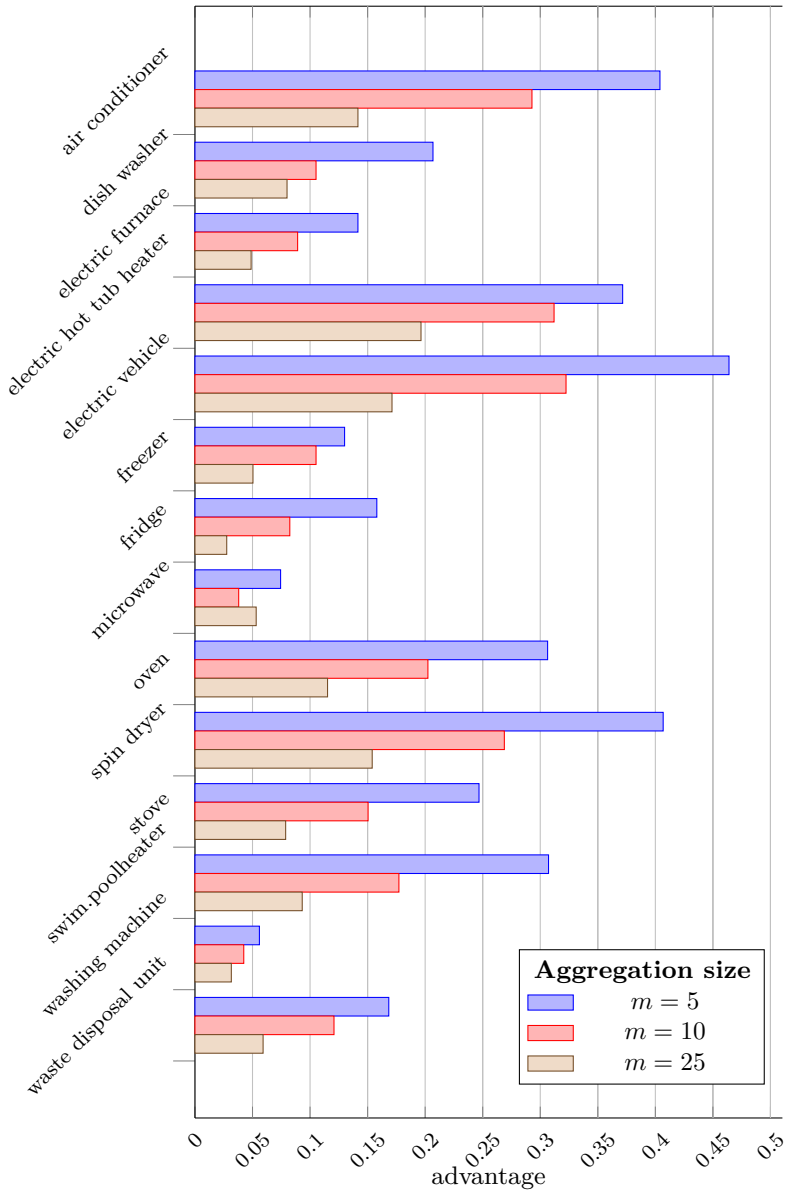

Fig. 18. Average adversarial advantage for particular appliances. Three different aggregation sizes tested(m=5,10,25) (Dataport, 15 min resolution)