Deploying and Evaluating Pufferfish Privacy for Smart Meter Data

Stephan Kessler, Erik Buchmann, and Klemens Böhm
Karlsruhe Institute for Technology (KIT), Germany
stephan.kessler@kit.edu, erik.buchmann@kit.edu, klemens.boehm@kit.edu

***Abstract*— Realizing a Smart Grid without sacrificing the privacy of consumers is a challenging problem. Data-centric approaches like Pufferfish ensure privacy by transforming data so that certain user-specified information, so-called secrets, cannot be inferred. Deploying Pufferfish on smart-meter data requires application-specific decisions, i.e., a general definition of secrets in time series. We investigate how to perturb energy consumption data in this manner, and we quantify the tradeoff between privacy and utility.**

I. Introduction

Designing a smart grid electricity-supply infrastructure is an important issue. This is because it promises to reduce CO_2 emissions and to guarantee supply at affordable prices. The smart grid initiative requires the installation of smart meters in private households. These devices measure the power consumption in short time intervals, e.g., every 15 minutes. Thus, they produce time series that contain the sum of the energy consumption of all electrical devices active during such a time interval. Various applications require access to the this data. Think of demand-side management or local energy markets [7], an efficient way of allocating renewable energy. However, privacy regulations and individual privacy preferences prevent arbitrary parties from accessing smart-meter data. Such data contains personal information [18, 3, 11, 6], e.g. on devices running and on the presence of residents. Obligations such as the European Directive 95/46/EC allow the disclosure of data only if it is non-personalized or if the individuals have consented.

Any smart grid service must deal with the tradeoff between the usefulness of data disclosed and the privacy of individuals. Which information actually is considered private depends on the individual. Thus, processing time series while protecting privacy requires privacy constraints that one can define individually. Libraries of constraints to are conceivable as well. The information to be hidden is referred to as *secrets*. Potential secrets go well beyond aggregated values from several households approaches such as [2] have exclusively focused on so far. De-personalization of such data ('anonymization') is not applicable in many cases either: Work on re-identification [6] shows that it is very difficult to remove all relationships to individuals from smart-meter data while preserving utility, and use cases such as demand-side management require data with identifiers.

Example 1: Bob is willing disclose his smart-meter data if it does not contain certain information. Suppose that Bob has a flow heater which starts when he begins showering, stops when he finishes and does not consume power otherwise. To keep the presentation simple, this heater will be our running example. Bob wants to keep private when he is showering on weekends and in the morning during weekdays. This defines the secrets. An adversary should not be able to learn whether the flow heater is starting or stopping between 8:00 and 11:00 on a weekday by inspecting the disclosed data. On weekends, the data should be so noisy that inferring the time when the heater is working is unlikely. To this end, one has to know how the time series reflect the heater usage and hide this on a weekday and detect when the heater starts and stops on a weekend. Approaches such as differential privacy applied on smart meter data [2] do not help with this kind of secret, because they do not allow such a detailed specification of the information to remain private. Finally, to preserve utility the data should still contain information that Bob does not explicitly want to hide. □

Individuals might allow the disclosure of their smart-meter data if their privacy preferences are strictly respected. Each individual should have the option to specify such private information. The Pufferfish privacy framework [15] guarantees that certain sensitive information is removed from a data set. Pufferfish supports the definition of intuitively understandable privacy requirements and their semantics, and it also considers correlations within the data set.

Example 2: Let $f(A), f(B), f(C)$ be smart-meter time series of Alice, Bob and Carl's household. $f(B)[t]$ is the total power consumption of Bob's household at time slot t. Differential privacy approaches [2, 22] publish the privacy-enhanced sum at each time slot of the households considered, i.e., $f(B)[t] + f(A)[t] + f(C)[t] + \ldots$: If there is not any correlation of the consumptions of Bob, Alice and Carl, an adversary cannot infer the actual consumption of one of them. However, there also are correlations when looking at each time series in isolation: Suppose that Alice, Bob and Carl each have a flow heater (for the shower) and bath lighting. $f(B)^1[t]$ is Bob's flow heater consumption and $f(B)^2[t]$ the one of the bath lighting. $f(B)[t]$ is the sum of all appliances in Bob's household: $f(B)[t] = f(B)^1[t] + f(B)^2[t] + \ldots$. Privacy cannot be guaranteed in the same way as for the sum of $f(B)[t], f(A)[t]$ and $f(C)[t]$: The flow heater and the bath lighting obviously have correlations Differential Privacy does not deal with [14]. □

Pufferfish is an abstract framework that has received much attention from many communities. However, (i) its deployment to smart-meter data requires challenging conceptual work, and (ii) a quantitative evaluation does not exist, i.e.,

it remains unclear which price one has to pay for Pufferfish privacy guarantees. The challenges are to represent private information in smart-meter data, to perturb the aggregated data according to Pufferfish guarantees, to ensure generality and to evaluate utility and coverage of privacy requirements. Regarding (ii), we examine the tradeoff between privacy and utility in a smart-grid scenario.

A run of a specific device results in a sequence of power-consumption values added to the total consumption. Such sequences corresponding to runs of the same device may vary in the actual values. This is because (a) appliances have a slightly different consumption each time they run, and (b) the smart meter may measure their consumption together with the ones of other devices. A first challenge is to find an abstracted representation of time series flexible enough to cover this uncertainty and specific enough to have a meaning for the secret in question. We call a single value of such an abstracted representation coefficient. This abstraction must have a clear-cut semantics, and the transformation of the time series to this representation must be well-defined. The goal of the abstraction is to have coefficients allowing to formulate specific secrets: One should choose transformations whose results correspond to potential secrets.

Example 3: In Example 1 the coefficients have to allow conclusions regarding the heater. Suppose that a heater consumes $25kW$ when running and $0W$ otherwise. Thus, a difference of the power consumption at point of time t to $t+1$ of around $25kW$ possibly indicates a starting flow heater. Exactly this can be subject of a privacy requirement. A meaningful abstraction then has a coefficient representing this kind of change. While the start of the flow heater results in two successive consumption values, other devices will create more complex sequences. For example, a washing machine carries tasks like heating or spinning. □

Pufferfish requires to adapt the data that represents a secret. It is not straightforward to devise perturbations fulfilling Pufferfish requirements [15]. In particular, perturbing an aggregate of several appliances is difficult, since it requires a decomposition. Next, we must take into account that different appliances in the decomposed representation may have correlations. Our approach is to deal with such time series individually per appliance.

Quantifying the usefulness of the perturbed smart meter data is not obvious, too: General distance measures for time series do not necessarily quantify utility. It is more conclusive to compare the performance of a real-world application using perturbed and unperturbed smart-meter data. Sometimes, the application may need information that is supposed to be secret. Think of an application depending on the aggregate sum of the consumption over some time, and exactly this is the Pufferfish secret. In general however, the needs of the application and the secrets may be orthogonal to each other. Because Pufferfish does *not* allow to specify characteristics of the data the perturbation must *not* affect, the impact of Pufferfish on the utility of the data is unclear.

Next, the evaluation of utility requires meaningful user-defined privacy requirements. Finding realistic requirements is challenging since many individuals are not yet aware of the privacy risks of the smart grid. Thus, an objective source of requirements is needed for a meaningful evaluation.

We address all these challenges as follows: Since the kinds of possible secrets are broad, we carefully select different abstracted representations together with adequate transformations for each of them. We illustrate this using the wavelet transformation as example; it covers several kinds of possible secrets. Privacy is guaranteed by the decomposition of the aggregated power signal into several channels on a conceptual level and the application of noise following the ϵ-Pufferfish principle [15] . Before publication, a time series is transformed back to the original, time-based representation.

In our evaluation, we show that this transformation principle is general enough to cover a wide range of requirements. We arrive at objective privacy requirements by looking at the outcome of various information-extraction methods from literature, i.e., features of smart metering data that others have deemed relevant. In this article, we define secrets covering features used for re-identification [6]. An extended version [13] features further secrets, with similar results. Next, in a local energy market [7], the utility of participants depends on the accuracy of the description of their demand; using perturbed data instead of the real one is expected to curb utility. Here, utility can be quantified as welfare, an established notion from economics. Welfare allows to quantify the cost of privacy as a real currency amount, which is more intuitive than, say, measures for prediction quality. The impact of privacy guarantees on utility is relatively low, while hiding realistic secrets: Even with secrets that require modifying the entire time series, the market welfare is reduced by 26% only in one setting we have examined.

Paper outline: We start with related work (Section II) and then introduce our way of applying Pufferfish (Section III). Section IV evaluates our approach, and Section V concludes. – Note that there exists an extended version of this article, containing a more detailed description of Pufferfish and the wavelet transformation, proofs of the lemmas and material that complements the evaluation [13].

II. Fundamentals

Having defined a common notation in Section II-A, we review well-known privacy-protection approaches in Section II-B. The Pufferfish Framework is explained in Section II-C. The wavelet transformation (Section II-D) is a technique to process and analyze time series, which we use as well. Other related work in turn is discussed in Section IV.

A. Notation

To support different abstract representations of time series, we have chosen a vector-based representation. The

coefficients of each vector defined on a basis express a finite linear combination of this basis. In other words, the basis defines the meaning of the coefficients. Vectors also allow to change the basis, resulting in other meanings of the coefficients. The standard representation of a time series is a mapping between points of time and the value domain, e.g., consumption values measured. Thus we define the time domain $\mathcal{T}$ first and then define a time series as a vector.

Definition 1 (Time domain $\mathcal{T}$): $\mathcal{T}$ is the standard domain of the time series considered. We assume that it is discrete and of finite length, i.e., $\|\mathcal{T}\| \leq \infty$.

Definition 2 (Time series): A time series is an n-dimensional vector with the basis B, referred to as f_B. To refer to its t-th element, we write: $f_B[t]$.

The common vector notation requires a standard basis consisting of canonical unit vectors e_i. For a given $\mathcal{T}$, we define the relationship of a time series f to each $t \in \mathcal{T}$: Let $[t_1, ..., t_n]$ be the ordered list of all $t_i \in \mathcal{T}$. Then $f_B[t_i] = f_B^\top \cdot e_i$ is the electricity consumption at time slot t_i. In other words, e_i represents the ith ordered element of $\mathcal{T}$, and $B = \{e_i | i = 1 \dots n\}$ forms the standard basis.

Definition 3 (Vector space): $\mathcal{V}_B$ is the vector space containing all linear combinations of basis B.

B. Privacy-Protection Approaches

Next to Pufferfish (cf. Section II-C), which serves as the framework for this current work, there is further related work. Differential Privacy provides provable privacy guarantees for statistical databases [9] and has been applied to smart meter data [2] and time series [22]. Example 2 has illustrated the limitations. Other approaches for time series disclose only aggregated results [5, 24] or build on k-anonymity [1, 19]. In contrast to such approaches, we are not limited to one specific information-extraction goal. Pufferfish features a more general approach, namely hiding user-defined secrets. Additionally, [5, 24, 1, 19] do not give provable guarantees. The approach evaluated here in turn allows for arbitrary queries over the disclosed data.

A perturbation method which handles each individual time series in isolation is to add random noise. However, there exist several methods to de-noise time series and to recover the original values, see [20]. As a counter-measure to de-noising techniques, the perturbation scheme in [20] transfers the time series to a Fourier or wavelet representation and then adds noise to coefficients exceeding a threshold. However, a data owner cannot decide what exactly is perturbed. This may result in unnecessarily perturbed information and in sensitive information still present.

C. The ϵ-Pufferfish Framework

Pufferfish [15] is a generalization of Differential Privacy providing provable privacy guarantees and utility [14]. It requires the following constituents: (a) A set of potential secrets $\mathcal{S}$ describing *which* information can be hidden. It is a domain for $\mathcal{S}_{pairs}$. (b) The discriminative pairs of secrets $\mathcal{S}_{pairs}$, describing *how* a piece of information should be hidden. (c) Finally, data-evolution scenarios $\mathcal{D}$. Data-evolution scenarios contain assumptions on how the data has been generated. This is background knowledge of an adversary. Technically speaking, $\mathcal{D}$ is a set of probability distributions over the possible database instances $\mathcal{I}$. Each $d \in \mathcal{D}$ corresponds to the background knowledge of an attacker on how the data has been generated. For example, $P(\mathcal{D}ata = \{x_1, ..., x_n\} | d_p) = p(x_1) \cdot ... \cdot p(x_n)$ if the probabilities of each record in $\mathcal{I}$ are independent. $P(\mathcal{D}ata = \{x_1, ..., x_n\} | d_p)$ is the conditional probability that $\mathcal{D}ata$ is $\{x_1, ..., x_n\}$ under d_p.

A privacy mechanism $\mathcal{M}$ is a method for transferring a data set $\mathcal{D}ata$ into a perturbed and privacy-enhanced representation $\mathcal{M}(\mathcal{D}ata)$. It guarantees the ϵ-Pufferfish privacy criterion if it fulfills the following definition:

Definition 4 (ϵ-Pufferfish Privacy): Given a set of secrets $\mathcal{S}^{\mathcal{P}}$, a set of discriminative pairs ${\mathcal{S}^{\mathcal{P}}}_{pairs}$, data-evolution scenarios $\mathcal{D}$ and a privacy parameter $\epsilon > 0$, a privacy mechanism $\mathcal{M}$ satisfies ϵ-Pufferfish$(\mathcal{S}, \mathcal{S}_{pairs}, \mathcal{D})$-Privacy if, for all outputs of $\mathcal{M}$, all pairs $(s_i, s_j) \in \mathcal{S}_{pairs}$ and all distributions $d \in \mathcal{D}$ the following holds:

$$P(\mathcal{M}(\mathcal{D}ata) = o | s_i, d) \leq e^\epsilon \cdot P(\mathcal{M}(\mathcal{D}ata) = o | s_j, d)$$
$$P(\mathcal{M}(\mathcal{D}ata) = o | s_j, d) \leq e^\epsilon \cdot P(\mathcal{M}(\mathcal{D}ata) = o | s_i, d)$$

$P(\mathcal{M}(\mathcal{D}ata) = o | s_j, d)$ is the probability that the output of $\mathcal{M}$ is o if s_j holds, and the data distribution is d.

The intuition is best explained with the following equation that is directly computed from Definition 4:

$$e^{-\epsilon} \leq \frac{P(s_i | \mathcal{M}(\mathcal{D}ata) = o, d)}{P(s_j | \mathcal{M}(\mathcal{D}ata) = o, d)} \Big/ \frac{P(s_i | d)}{P(s_j | d)} \leq e^\epsilon$$

If an adversary thinks that s_i is α times as likely as s_j, then, after having access to the privacy enhanced output of $\mathcal{M}$, he may only believe that s_i is at most $e^\epsilon \alpha$ times and at least $e^{-\epsilon} \alpha$ as likely as s_j.

D. Wavelet Transformation

We use the wavelet-transformed representation as an example, in order to express secrets and to hide them. The following is a concise review, see for instance [21] for a comprehensive introduction. Note that our study is not limited to the wavelet transformation, see Section III-D.

Definition 5 (Wavelet): A wavelet $w[t]$ is a finite time series with properties: $\int_{-\infty}^{+\infty} w[t] = 0$ and $\int_{-\infty}^{+\infty} w[t]^2 = 1$.

Definition 6 (Wavelet Transform): A wavelet transformation is an orthonormal basis transform to a wavelet basis. Each element of the basis is a development over time.

To cover the n-dimensional vector space, the wavelet transform results in multiple levels, reflecting different horizontally stretched representations of $w[t]$. Further, the wavelet transformation is invertible. The coefficient at the

highest level, the *scaling coefficient*, is not a multiple of the wavelet $w[t]$; it represents the y-position of the time series.

To ease presentation, we include all the information for the transformation in w. In our example, w contains the Haar wavelet $w[t]$ together with the transformation. An example Haar wavelet transform of the time series on Figure 3 is displayed in Figure 4. A value smaller than zero corresponds to an increasing power consumption. Depending on the position of the increase, the change influences the first or the second level.

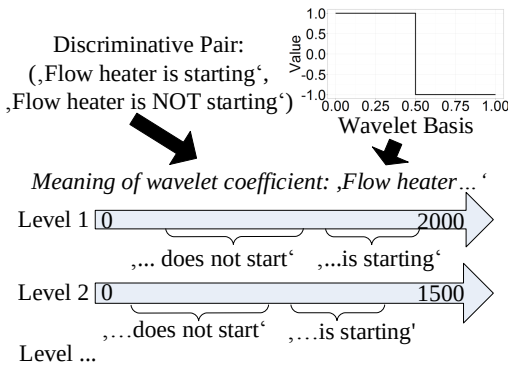

Figure 1. Example: Meaning of wavelet coefficients

Using wavelets requires specifying which elements in f_w are relevant for the individual: Switching on the flow heater results in a strong increase of the power consumption. In the Haar wavelet domain this leads to high coefficients on lower levels. When the flow heater is switched off, this has an analogous effect. This allows the distinction whether Bob starts/stops to shower or not, cf. Figure 1.

III. Provable Privacy for Smart Meter Time-series

We now explain our instantiation of the Pufferfish mechanism $\mathcal{M}$ for smart-meter data. $\mathcal{M}(f)$ reconstructs a time series f resulting in one that guarantees ϵ-Pufferfish privacy. We conduct the steps listed in Figure 2. To ease presentation, we assume a single pair of discriminative secrets s_{pair} and a single time series f. This is not a restriction since each element of $\mathcal{S}_{pairs}$ is handled in isolation for each time series. When speaking of an aggregate, we always mean $f[t]$, the aggregate consumption of all running appliances.

For further explanations see Algorithm 1. It contains of three steps: At first, we transform a time series f to an abstracted representation f_w. Second, secrets determine the perturbation of the abstracted time series according to Pufferfish guarantees. Third, we transform the modified time series back to a time based representation f'. We now explain these steps in detail.

Input: time series f
Input: Set of discriminative pairs $\mathcal{S}_{pairs}$ of secrets $\mathcal{S}$, (Inverse) Transformation Mechanism $\mathcal{C}^{trans}_{B'}$, $\mathcal{IC}_{B'}$ and basis B'
Input: Data evolution scenarios $\mathcal{D}$
Input: Privacy parameter ϵ
Result: Time series with privacy guarantees f'
foreach $s_{pair} \in \mathcal{S}_{pairs}$ **do**
  // Step 1: Transformation;
  $f_{B'} = \mathcal{C}^{trans}_{B'}(f)$;
  // Step 2: Perturbation;
  Determine $\mathcal{N}_\epsilon$ to fulfill ϵ-Pufferfish Privacy based on $\mathcal{D}$ and s_{pair};
  Set p^{coeff} according to s_{pair};
  $f'_{B'} = \mathcal{P}(f_{B'}, \mathcal{N}_\epsilon, p^{coeff})$;
  // Step 3: Inverse Transformation;
  $f' = \mathcal{IC}_{B'}(f'_{B'})$
end
return f';

Algorithm 1: Pufferfish Privacy Mechanism $\mathcal{M}$

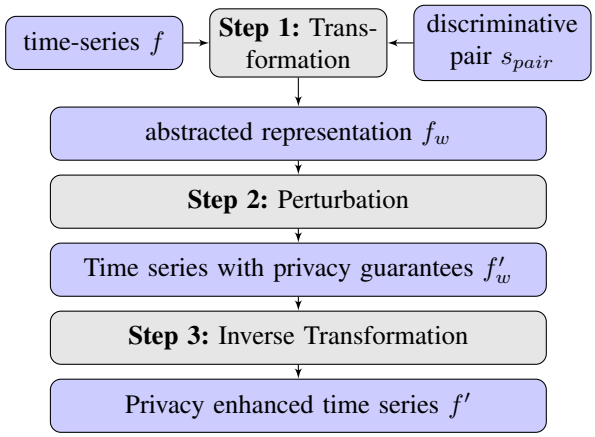

Figure 2. Privacy preservation for s_{pair}

A. Step 1: Transformation

This step transforms a given time series to an abstracted representation where each value carries a specific meaning in relation to secrets (and not necessarily to a point of time). Secrets are geared to specific transformations. Thus we first need to define the transformation mechanisms (Section III-A1), before formulating secrets respectively discriminative pairs for smart meter data (Section III-A2).

1) Transformation: Abstract representations of time series are numerous [8]. The right choice depends on the privacy requirements. Thus, we define requirements on transformation approaches to be applicable with our approach.

Definition 7 (Transformation Mechanism): Let B be the standard basis and B' a different basis of a vector space. A transformation mechanism $\mathcal{C}_{B'}$ is a function of

type $\mathcal{V}_B \to \mathcal{V}_{B'}$ that converts a time series from the time-based representation f to an abstracted representation $f_{B'}$ with basis B' and fulfills the following properties:

1) The transformation is invertible, i.e., there exists an inverse of $\mathcal{C}_{B'}$ We refer to it as $\mathcal{IC}_{B'} : \mathcal{V}_{B'} \to \mathcal{V}_B$.
2) $\mathcal{C}_{B'}$ is an endomorphism for the $+$-operator. Let f, g be time series, then: $\mathcal{C}_{B'}(f + g) = \mathcal{C}_{B'}(f) + \mathcal{C}_{B'}(g)$

Suppose that the time series is an aggregate of power consumptions. The endomorphism property simplifies the perturbation: Noise can be added to certain parts of the aggregate as well as to the aggregate, yielding the same result. Section III-A2 explains the importance of this property.

The invertibility property implies the following: First, if $f_{B'}$ is invertible, any information of f is present in $f_{B'}$. Thus, any information of f is also included in the abstracted representation. Second, invertibility requires well-defined semantics of every element in $f_{B'}$. Consequently, such clear semantics also hold for secrets dependent on the coefficients, i.e., each coefficient has a specific meaning in relation to a secret. Note that we do not make any restriction on the length of $f_{B'}$ in comparison to f; so the transformation output may also have a higher dimensionality than f.

Haar-Wavelet example transformation. The wavelet transformation as described in Section II-D satisfies Definition 7. This transformation for the Haar basis is invertible and an endomorphism for addition. See Lemma 1. Additionally, the wavelet transformation keeps the time location; each value in $f_{B'}[x]$ corresponds to a specific number of entries in $f[t]$. We refer to the wavelet-transformation mechanism with the Haar basis as $\mathcal{C}_h^{Wave}$.

Lemma 1: *The Haar wavelet transformation is invertible and an endomorphism for the $+$-operator*

2) Secrets in Smart-Meter Data: Possible secrets $\mathcal{S}$ an individual may want to hide range from relatively simple ones like *'The dishwasher is running'* to rather complex ones involving several appliances like *'There is cooking activity'*. Other examples are *'There is activity in the kitchen'*, *'The fridge is running'* or *'Someone is watching a certain TV program in the morning'*.

The power-consumption data of a household, usually monitored by a smart meter installed at the main power connection, is the aggregate of all appliances. However, only parts of it typically are relevant for certain secrets. Hence, it is important to be able to examine parts of the aggregate in isolation. Looking at the smart meter time series as a signal, it is the aggregate of several channels. For example, the consumption of the television is one channel $f^1[t]$, the dishwasher is another one, $f^2[t]$.

Definition 8 (Signals and channels): A signal is the complete power consumption measured at the smart meter of the household and is represented as a vector $f[t]$. A channel is a part of the signal, referred to with a superscript, e.g., $f^i[t]$. We see a signal as the sum of n channels: $f[t] = f^1[t] + \cdots + f^n[t]$

Even on channels only containing the consumption of individual devices, a sequence of consumption values is still required in many cases to gain interesting information. From non-intrusive appliance load monitoring (NIALM) approaches [11, 17, 16, 10, 4] it is well-known that a sequence of time-value pairs identifies appliances and their state, and appliances tend to be detectable in f.

The connection between values of a time series (even if it is an abstraction) and intuitive descriptions of possible secrets is not obvious. Thus, we define the following.

Definition 9 (Description of a Secret): A description of a secret is a triple

$$s = (s^{Base}, s^{Trans}, s^{Coeff})$$

where s^{Base} is the basis for a transformation mechanism s^{Trans}. s^{Coeff} is the formal description of the coefficients in the abstracted representation $f_{s^{Base}}$ that make s true. We write $f_w[t] \in s^{Coeff}$ if an element of the transformed time series makes the secret true.

We do not require a specific language to describe the coefficients. However, the description has to be non-ambiguous.

A description of a secret reflects what should be hidden, but not how. It rather is necessary to have discriminative pairs of secrets. Thus, Pufferfish requires a description of discriminative pairs of secrets on smart-meter time series.

Definition 10 (Description of a Discriminative Pair of Secrets): A description of a discriminative pair of secrets s_{pair} is a pair of descriptions of secrets $s_{pair} = (s_1, s_2)$, so that the following holds:

- The base as well as the transformation method are the same ($s_1^{Base} = s_2^{Base}$ and $s_1^{Trans} = s_2^{Trans}$).
- The secrets are mutually exclusive but do not need to be exhaustive, i.e., there may exist values in the range of a coefficient that neither make s_1 nor s_2 true.
- The coefficients in question for s_1 and for s_2 are non-overlapping: $s_1^{Coeff} \cap s_2^{Coeff} = \emptyset$.

Typically, only parts of the entire signals are relevant for secrets and discriminative pairs.

Definition 11 (Relevant Channel): For a given signal f consisting of $i \in [1 \dots n]$ channels and for a discriminative pair $s_{pair} = (s_1, s_2)$, we call the channel that contains the information whether s_1 or s_2 is true the relevant channel r. We refer to the corresponding time series as f^r. The decomposition partitions the signal. Formally:

$$f[t] = f^1[t] + \cdots + f^r[t] + \cdots + f^n[t]$$

There typically are correlations between channels. They depend on the actual discriminative pair and the assumptions contained in $\mathcal{D}$ regarding an adversary. In Example 2, the lighting f^2 is correlated with the heater f^1. But the lighting consumption is not part of the relevant channel, since it is not directly related to the showering activity.

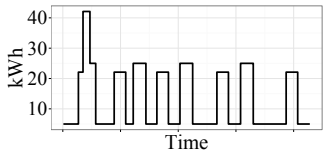

Figure 3. Example of a starting/stopping flow heater

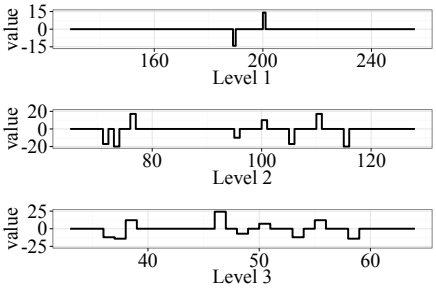

Figure 4. Haar Wavelet decomposition of a flow heater (three levels)

Correlations result in different data-evolution scenarios and require a different distribution of the noise applied. The specifics are part of the Pufferfish Framework [15]. The following example illustrates the description of the secrets in smart-meter time series.

Example 4: Bob wants to hide whether secret s_1 'The heater is starting/stopping' or secret s_2 'The heater is not starting/stopping' is true. The wavelet transform with the Haar basis reflects 'switch on' respectively 'switch off' events and is suitable for the discriminative pair $s_{pair} = (s_1, s_2)$. Let h be the Haar wavelet basis, then $s_1^{Trans} = s_2^{Trans} = \mathcal{C}_h^{Wave}$. For the sake of simplicity, we assume that the heater power-consumption function is of rectangular shape over time, as illustrated in Figure 3 (generated with the model of [23]). Figure 4 contains $\mathcal{C}_h^{Wave}(f)$ of the time series illustrated in Figure 3: The x-axis in Figure 4 shows the time location and the y-axis the 'intensity' of the Haar basis. Coefficients in Level 1 and 2 reflect the starting and stopping of the heater, as explained in Section II-D. To include small inaccuracies, we define s_1^{Coeff} to cover coefficients of Level 1 if their value is in $[13, 17]$ or $[-17, -13]$ and Level 2 if their value is in $[18, 22]$ or $[-22, -18]$. Consequently s_2^{Coeff} contains all values of coefficients on Level 1 except for $[13, 17]$ and $[-17, -13]$ and Level 2 except for $[18, 22]$ and $[-22, -18]$. s_1 and s_2 qualify as a discriminative pair s_{pair} since $s_1^{Trans} = s_2^{Trans}$ and $s_1^{Coeff} \bigcap s_2^{Coeff} = \emptyset$. In this example, the channel relevant for s_{pair} only contains the heater consumption. □

For different transformations or for different bases the determination of coefficients works in the same way, as long as the proposed specification of coefficients holds. Using a different transformation or basis allows to cover other requirements, see Section III-D.

B. Step 2: Perturbation

This section explains how we have ensured Pufferfish privacy in time series of smart meter data. One common method explicitly illustrated in the following is to apply additive Laplace noise to aggregates [15]. If different channels are correlated, the noise should follow other distributions, see [15]. However, this does not affect the following description. As explained in Section III-A2, a smart meter signal is an aggregate of different appliances, but noise is only required for some channels. Identifying the channels and the noise distribution applicable is not obvious.

1) Perturbation of Time Series: We explain our approach for perturbing a time series of smart meter data in the transformed representation. The perturbation must have a noise distribution. We refer to the transformed version with mechanism s^{Trans} and basis s^{Base}, where w consists of s^{Trans} and s^{Base}, as f_w. The perturbed time series is f'_w. The perturbation also requires the selection of the coefficients to add noise to. This leads to the following definition.

Definition 12 (Perturbation Mechanism for a Discriminative Pair): A perturbation mechanism $\mathcal{P}$ is a function that takes a time series f_w in abstracted representation, the noise $\mathcal{N}_\epsilon$ to be applied dependent on the privacy parameter ϵ and a formal definition of the coefficients to be perturbed p^{coeff}. It returns the privacy-enhanced time series in the transformed representation, referred to as f'_w.

$$f'_w = \mathcal{P}(f_w, \mathcal{N}_\epsilon, p^{coeff})$$

2) Noisy elements: p^{coeff} specifies the elements of f'_w to be perturbed. Similarly to the definition of secret descriptions, we leave aside the language for selecting these coefficients. Examples for p^{coeff} are as follows:

- **All:** This is the most simple strategy. Additive noise is applied to all coefficients.
- **Trigger dependent:** Since coefficients in a certain range have a defined meaning, they are perturbed. This is similar to [20]. However, the ranges and the noise have a well-defined meaning (c.f. Figure 1), guaranteeing a certain level of privacy. Note that it is now possible to define the noise relative to $f_w[x]$.
- **Time dependent:** The user specifies coefficients to be perturbed (e.g., from t_1 to t_2 etc.), independent of the value. However, this only works if the transformation mechanism keeps the time location.
- **Trigger and time dependent:** This combines both possibilities just mentioned.

3) Noise Distribution: $\mathcal{P}$ used with noise according to Pufferfish and to the discriminative pair $s_{pair} = (s_1, s_2)$ guarantees privacy.

Lemma 2: *Let f be a time series of smart meter data, $s_{pair} = (s_1, s_2)$ the information an individual wants to hide, $\mathcal{C}_{s^{Base}}$ a transformation mechanism suitable for s_{pair} and*

$\mathcal{P}$ a perturbation mechanism. There exists a distribution of noise $\mathcal{N}_\epsilon$ with $\mathcal{P}$ for $\mathcal{C}_{s^{Base}}(f)$ that satisfies the ϵ-Pufferfish Privacy Definition.

The following example illustrates how to choose noise for the starting flow heater appropriately.

Example 5: Bob wants to hide the pair $s_{pair} = (s_1, s_2)$ from Example 4. To do so, we carry out the proposed wavelet transformation $\mathcal{C}_w^{Wave}$ with the Haar basis w. Let f^r be the relevant channel for s_{pair}. To ease presentation, suppose that the channels are statistically independent. The coefficients in question for s_1 and s_2 correspond to non-overlapping intervals by definition. For instance, let $f_w[x]$ be a value of Level 1 of the wavelet-transformed representation. If $f_w^r[x] \in [y-k, y+k]$, s_1 is true for $y = 15$ with an imprecision interval of $k = 2$, otherwise s_2. For Level 2 s_1 is true for $y = 20$ and $k = 2$. In this case, we want to prevent an adversary from learning the value of $f_w^r[x]$ by accessing the privacy-enhanced signal $f'_w[x]$. [15] shows that adding noise drawn from the Laplace$(4k/\epsilon)$ distribution with density function $\frac{\epsilon}{8k} e^{-\epsilon|x|/4k}$ guarantees ϵ-Pufferfish privacy for the aggregate as follows: An adversary cannot distinguish whether the value of a single channel is between $y-k$ and $y+k$ or one of the neighboring intervals $[y+k, y+3k)$ or $[y-3k, y-k)$. Let X be a random variable drawn from the above distribution and x be the coefficient to hide. We then generate the privacy-enhanced aggregate $f'_w[x]$ as follows:

$$f'_w[x] = f_w^r[x] + f_w^i[x] + \cdots + X$$

Note that adding noise does not require the disaggregation of the signal into several channels, i.e., $f'_w[x] = f_w[x] + X$. Adding noise already ensures Pufferfish privacy.

Since wavelet coefficients are time-located, it is possible to add noise for weekdays between 8:00 and 10:00, cf. Example 1. On the weekends, we add noise during the whole day on Levels 1 and 2. □

C. Step 3: Inverse Transformation

The last step transforms the abstracted and perturbed representation f'_w back to the time-based one f'. This is possible, since Definition 7 requires invertibility. Note that this approach ensures that only values relevant for the secrets specified, are perturbed, and the perturbation affects these values as little as possible.

D. Alternative Transformations

Using the Haar wavelet transform does not allow to express any secret conceivable. However, our approach can use any transformation fulfilling Definition 7. This includes well-known ones such as the decomposed wavelet transform, the wavelet packet transform, and the discrete Fourier transform. For example, the last one is suitable to represent secrets covering periodic events. Please see the technical report [13] for details.

IV. Evaluation

Our evaluation has two goals, generality and utility: First, an individual should be able to hide arbitrary information. Second, the disclosed data should still be useful while guaranteeing privacy to the extent specified.

Regarding the first issue, to evaluate objectively whether our approach is general enough to cover a broad range of privacy requirements we need a reliable source of such requirements. To our knowledge, such a source for smart meter data does not exist. However, there exist recent approaches extracting various kinds of information on individuals from smart meter data. The information these approaches try to extract can be perceived as information that is worth to be protected, i.e., as privacy requirements. We show that it is possible to define discriminative pairs of secrets suitable for these requirements. The approach explicitly considered in what follows is a re-identification approach (Section IV-A). We also have evaluated whether our approach can prevent non-intrusive appliance-load monitoring from extracting sensitive information. One result is that the appliance-detection rates drop. See [13] for details.

We now preview the second issue of quantifying utility. Abstract time-series-distance measures do not allow for meaningful conclusions regarding the utility of a modified time series for applications. To ensure realistic conditions, we evaluate the utility of a noisy, privacy-enhanced data set by means of a local electricity market (Section IV-B).

The approach presented hides user-defined preferences in a time series of smart-meter data. A comparison of our approach with another one regarding utility would only be conclusive if the reference point offered the same extent of privacy; but we are not aware of any such approach.

A. Generality: Re-Identification

Re-Identification means linking personal data which does not contain any direct identifiers (name, address, etc.) to individuals. Features of the consumption help to re-identify time series of power-consumption values [6]. To illustrate, we focus on the following four features: sum, maximum and minimum of the power consumption for a time interval and average bedtime hour, i.e., the first point of time in the evening when the consumption decreases significantly. Note that we also can hide all other features listed in [6]. Table I lists the necessary transformations and the relevant coefficients. Those four features have the same structure as almost half of the features in Table I.

We now review how re-identification works:

1) The adversary has feature values of households as external knowledge, e.g., a certain household usually goes to bed at $11pm$.
2) For each time series in question, the values for these features are computed. The adversary compares the results with the external knowledge.

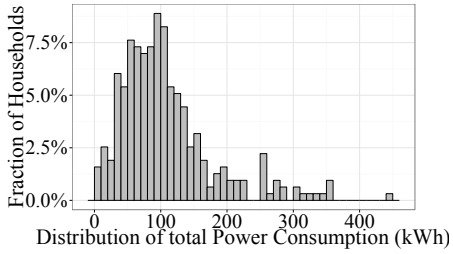

Figure 5. Total power consumption

Features	Transformation	Coefficients concerned
Sum	Haar-Wavelet	Scaling Coefficient
Maximum	Haar-Wavelet	Scaling Coefficient
Minimum	Haar-Wavelet	Scaling Coefficient
Evening Sum	Decomposed Wavelet	Relevant Scaling Coeff.
Morning Sum	Decomposed Wavelet	Relevant Scaling Coeff.
0.9 Quantile	Fourier	All
Standard Deviation	Fourier	All
Frequency of mode	Fourier	Significant Frequencies
Wakeup time	Haar-Wavelet	Level 1/2
Bedtime	Haar-Wavelet	Level 1/2

Table I
TRANSFORMATIONS FOR RE-IDENTIFICATION FEATURES

3) Assuming that households tend to have repeating behavior over time, features computed for a household for different time periods tend to have similar values. The system computes a score based on the difference between feature values that are part of the external knowledge and the values of the household in question. The smaller the score, the more likely the household is the sought one.
4) A household is deemed re-identified if its time series receives the n-th lowest score or lower. n is an external parameter and allows to overcome imprecision.

An earlier result is that up to 82.8% of the households can be re-identified [6] in an unmodified data set. To hinder re-identification, certain distinctive features need to be hidden. For the four secrets explicitly considered here, the wavelet transform with a Haar basis is suitable: The scaling coefficient (see Section II-D) represents the sum and also influences the maximum and minimum, see Section IV-A1. Levels 1 and 2 reflect the first significant decrease for the bedtime hour, like the heater starting or stopping.

1) Hiding Sum, Maximum and Minimum: Next, we say how the sum, the maximum and the minimum can be hidden. To do so, we take a closer look at re-identification. The total power consumption of a time period is the sum of all channels $i \in [1 \dots n]$:

$$\sum_{\forall t \in \mathcal{T}} f[t] = \sum_{\forall t \in \mathcal{T}} f^1[t] + \dots + \sum_{\forall t \in \mathcal{T}} f^n[t]$$

An adversary with external knowledge on the power consumption trying to re-identify a record has to take inaccuracies into account, i.e., he typically does not know the total consumption for sure, only within a certain range. Thus, we partition the channels into a known one, such as the relevant channel r, and the ones not known. The channels not known are responsible for the difference between the known channels and the total consumption at each point of time.

$$\sum_{\forall t \in \mathcal{T}} f[t] = \sum_{\forall t \in \mathcal{T}} f^1[t] + \dots + \sum_{\forall t \in \mathcal{T}} f^r[t] + \dots + \sum_{\forall t \in \mathcal{T}} f^n[t]$$

Based on the sum $\sum_{\forall t \in \mathcal{T}} f[t]$ the adversary has to decide whether the known channel is consistent with his knowledge. Adding Laplace noise in line with ϵ-Pufferfish privacy leads to uncertainty regarding $\sum_{\forall t \in \mathcal{T}} f^r[t]$. Re-identification is successful if an adversary is able to single out the true individual record. In particular, this is relatively easy if the feature values of individuals are spread over a wide range and are rather unique. Thus, individual privacy requirements depend on assumptions regarding other individuals in the data set. Describing a suitable secret is deciding which interval is sufficient to hide $\sum_{\forall t \in \mathcal{T}} f^r[t]$ amongst other channels. We use the following notation:

$s_k =$ 'Known power consumption is in interval [y-k, y+k]'

The discriminative pairs can be of the form $s_{pair} = (s_k, s_{3k})$. One way to determine k is to look at the distribution of a known data set. Figure 5 indicates that $k = 5kWh$ is sufficient to hide a single household amongst more than 10 others for a large number of households. These considerations also hold for the features 'Minimum' and 'Maximum'.

Applying noise to the scaling coefficient Applying noise to the scaling coefficient is special, compared to other coefficients. In particular, the scaling coefficient is normed. It represents the sum, minimum and maximum, and is calculated as follows: $\frac{\sum_{\forall t \in \mathcal{T}} f[t]}{\sqrt{\|\mathcal{T}\|}}$. Thus, the additive noise $Laplace(4k/\epsilon)$ is normed as well: $\frac{\sum_{\forall t \in \mathcal{T}} f[t]}{\sqrt{\|\mathcal{T}\|}} + \frac{Laplace(4k/\epsilon)}{\sqrt{\|\mathcal{T}\|}}$.

2) Hiding Bed-Time and Wakeup-Time Hours: According to [6], the bedtime hour is when a household switches off certain devices, e.g., the television, right before going to bed. This does not have to be the same devices for different households as long as they are usually switched off right before going to bed. We consider switch-off events only between $4pm$ and $2am$. Some appliances may still run, but only the change of consumption is of interest. An adversary trying to re-identify a household is interested in deciding whether the devices are switched off or not. Thus, an individual wants to hide the discriminative pair

s_{pair} consisting of the following secrets: $s_1 =$ 'Household switches off devices before bedtime' and $s_2 =$ 'Household does not switch off devices before bedtime'. The relevant channel r includes the devices mentioned for s_{pair}.

$$f_w[x] = f_w^r[x] + f_w^1[x] + \cdots + f_w^n[x]$$

The switch-off decreases the power consumption of $0.5kWh$ on $f_w^{s_{pair}}[x]$. Thus, we apply Laplace$((4 \times 0.5)/\epsilon)$ noise on Level 1 and Laplace$((4 \times \frac{0.5}{\sqrt{2}})/\epsilon)$ noise on Level 2 during $4pm$ and $2am$. Hiding wakeup times is similar.

3) Results: It is possible to hide all other features for re-identification [6]; Table I lists the necessary transformations.

To quantify effectiveness, we look at the relative decrease in accuracy, i.e., the number of households re-identified with and without applying noise. While re-identification makes use of a combination of features, to isolate the effects of hiding specific secrets we only look at features relevant for the secret. While this reduces the number of households re-identified, this is the case both with and without applying noise, so our evaluation is still conclusive. We deem a household re-identified if its time series receives the n-th lowest score at least. In total, we tested 158 household from the CER data set and set $\epsilon = 0.1$. This data set consists of roughly 5000 homes in Ireland with different numbers of inhabitants, measuring electricity consumption every 30 minutes over more than one year [12]. Table II contains our results. It contains the feature set used for re-identification and the accuracy decrease after applying the Pufferfish framework. First, independent of the feature set, there is a significant decrease in accuracy. Thus, hiding the features in the described way is effective. However, the algorithm still can re-identify a small number of households: In our evaluation, we have assumed the same discriminative pair for all households. However, for outliers in particular, e.g., a household consuming a lot of electricity and thus being easy to re-identify, discriminative pairs should differ. In particular, the k of the interval must be larger. If the feature value of a number of households is similar, then the re-identification algorithm starts to guess. Random 'correct' guesses become more with $n = 5$. Still, Pufferfish allows the definition of suitable secrets to hinder re-identification. Even with secrets designed in a straightforward way without considering outliers the accuracy decreases significantly.

B. Utility: Welfare of a Local Energy Market

A privacy method must protect sensitive information of individuals. However, it is also important that the data can still be used for certain purposes afterwards. In order to evaluate to which extent the proposed mechanism preserves utility, we integrate it into a local energy-market scenario and measure the effect on the welfare. Welfare is a well-known economic measure: It is the sum of consumer surplus (difference between willingness to pay and clearing price) and producer surplus (difference between clearing price and costs). In a local energy market, consumers and producers can trade electricity. In general, this leads to a more effective allocation of renewables, including a drop of CO_2 emissions. However, individuals have to reveal their prospective consumption to other market participants. Obviously, the prospective consumption tends to be similar or even identical to the actual one. With any reasonable market mechanism, if participants reveal their true demand they will receive the highest welfare. In turn, revealing a privacy-enhanced demand induces a loss of welfare. However, protecting privacy has a value for the individuals as well. Thus it is insightful to investigate this tradeoff. This method has already been tested in another similar context, cf. [7].

1) Results: We have evaluated a town with 300 persons living in households of up to five persons. The time interval examined is five days. The consumption data has come from the CER data source [12]. As renewable sources we have taken 150 photovoltaic sites as well as 150 combined heat and power plants. As privacy requirements, we have chosen to hide the bedtime and the total sum. Since Pufferfish as well as the selection of households include randomness, we repeat each experiment ten times. We measure the relative welfare, which is the welfare using the privacy method in relation to the welfare for the original data.

Hiding the bedtime results in a welfare loss of 26% on average, with a low spread, see Figure 6. Hiding requires applying noise to 10 hours a day. This includes the consumption after $4pm$, which contains a large fraction of the daily consumption due to evening activities of households. Hiding the sum respectively the minimum and maximum consumption leads to a smaller relative welfare loss compared to the bedtime requirement on average, but has a larger spread of values. In this case, applying noise shifts all the values of the time series up- or downwards, but it keeps the shape. This is because the actual development is not influenced. Thus, we see that hiding different secrets has different effects on the utility (Figure 6). Note that the welfare loss of 26% is relative to the theoretical maximum efficiency (cf. [7]), i.e., the loss of welfare is low.

C. Summary of Results

The evaluation has shown that Pufferfish privacy can indeed shield personal information from information-extraction approaches. The potential of an adversary to gain information from the disclosed data set has dropped significantly. On the other hand, we have shown by means of a local energy market that the utility of the resulting data set still is on an acceptable level. Again, we have used secrets that prevent state-of-the-art information-extraction methods from providing meaningful results.

V. Conclusions

Disclosure of data plays a significant role in the context of the smart grid. However, time series of smart meter data

Feature Set	Top n	w/o noise	noise	Accuracy Decr.
Sum	1	6	2	66%
Min	1	15	1	93%
Max	1	7	3	57%
Sum,Min,Max	1	30	8	73%
Bedtime	5	8	5	37.5%
Wakeup time	5	6	3	50%
Bed-, Wakeup time	5	13	6	53.8%

Table II
RESULTS RE-IDENTIFICATION

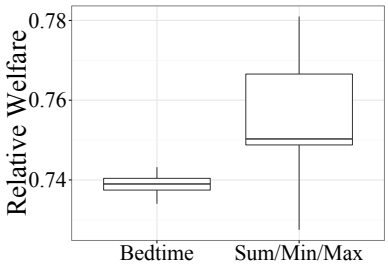

Figure 6. Relative Social Welfare

contain sensitive information, represented in many different ways. Individuals might not allow access to the data as long as sensitive information based on individual privacy preferences is not removed. Pufferfish is a state-of-the-art approach to hide specific information. However, application-specific work is required when applying it to smart meter data and carrying out an evaluation that is conclusive. This includes the definition of how sensitive information is represented, how data-evolution scenarios can be applied, and how the information can be perturbed to give Pufferfish guarantees. Next, it is challenging to evaluate the general coverage of secrets and the utility of the perturbed data. Our study has addressed these points.

Our study has featured a general way of describing secrets in smart-meter data. Transforming time series of such data is one possible way to facilitate the definition of arbitrary secrets. A certain set of transformations is sufficient to cover a broad variety of secrets, decreasing the impact of information-extraction methods on privacy significantly with a tolerable impact on the utility of the data.

REFERENCES

[1] O. Abul et al. Never Walk Alone: Uncertainty for Anonymity in Moving Objects Databases. *ICDE*, 2008.

[2] G. Acs and C. Castelluccia. I have a DREAM! (DIffeRentially PrivatE smart Metering). *Information Hiding*, 2011.

[3] A. Albert and R. Rajagopal. Smart Meter Driven Segmentation: What Your Consumption Says About You. *IEEE Trans. on Power Systems*, 28(4), 2013.

[4] N. Batra et al. INDiC: Improved Non-intrusive Load Monitoring Using Load Division and Calibration. *Conference on Machine Learning and Applications*, 2013.

[5] F. Bonchi and L. Lakshmanan. Trajectory Anonymity in Publishing Personal Mobility Data. *ACM SIGKDD Explorations*, 13(1), 2011.

[6] E. Buchmann et al. Re-identification of Smart Meter data. *Personal and Ubiquitous Computing*, 17(4), 2012.

[7] E. Buchmann et al. The Costs of Privacy in Local Energy Markets. In *IEEE CBI*, 2013.

[8] A. Das. *An introduction to continuous wave communication and signal processing*. 2012.

[9] C. Dwork. Differential privacy. *Automata, languages and programming*, 2006.

[10] H. Goncalves and A. Ocneanu. Unsupervised Disaggregation of Appliances using aggregated Consumption Data. In *The 1st KDD Workshop on Data Mining Applications in Sustainability*, 2011.

[11] G. Hart. Nonintrusive Appliance Load Monitoring. *Proceedings of the IEEE*, 80(12), 1992.

[12] Irish Social Science Data Archive. Electricity Customer Behaviour Trial. http://www.ucd.ie/issda/, 2012.

[13] S. Kessler et al. *Deploying and Evaluating Pufferfish Privacy for Smart Meter Data*. Karlsruhe Reports in Informatics; 01. 2015.

[14] D. Kifer and A. Machanavajjhala. No free Lunch in Data Privacy. *SIGMOD*, 2011.

[15] D. Kifer and A. Machanavajjhala. A Rigorous and Customizable Framework for Privacy. *PODS*, 2012.

[16] H. Kim et al. Unsupervised Disaggregation of low frequency Power Measurements. Number i. HP Labs Tech. Report, 2010.

[17] J. Kolter and M. Johnson. REDD: A public data set for energy disaggregation research. *Workshop on Data Mining Applications in Sustainability*, (1), 2011.

[18] A. Molina-Markham and P. Shenoy. Private Memoirs of a Smart Meter. *BuildSys*, 2010.

[19] M. Nergiz and M. Atzori. Towards Trajectory Anonymization: a Generalization-based Approach. *SIGSPATIAL*, 2(106), 2008.

[20] S. Papadimitriou et al. Time series compressibility and privacy. *VLDB*, 2007.

[21] D. B. Percival and A. Walden. *Wavelet Methods for Time Series Analysis*. 2006.

[22] V. Rastogi and S. Nath. Differentially private aggregation of distributed time-series with transformation and encryption. In *SIGMOD*, 2010.

[23] I. Richardson et al. Domestic electricity use: A high-resolution energy demand model. *Energy and Buildings*, 42(10), 2010.

[24] L. Shou et al. Supporting Pattern-Preserving Anonymization For Time-Series Data. *TKDE*, 2011.