Computer Security Symposium 2017
23 - 25 October 2017

Yet Another Experiment on Privacy-Utility Tradeoff for Power Usage Data

Mitsuhiro Hattori[1,a)] Takato Hirano[1] Fumio Omatsu[1] Rina Shimizu[1]

Abstract: Privacy-preserving data mining technologies have been studied extensively, and as a general approach, du Pin Calmon and Fawaz have proposed a data distortion mechanism based on a statistical inference attack framework. This theory has been extended by Erdogdu et al. to time-series data and been applied to energy disaggregation of smart-meter data. However, their theory assumes both smart-meter data and sensitive appliance state information are available when applying the privacy-preserving mechanism, which is impractical in typical smart-meter systems where only the total power usage is available. This issue was partially solved by the work we presented at ACISP 2017, but the experiment was done against a non-public dataset and this prevented interested researchers from conducting replication studies. In this paper, we conduct additional experiments against a publicly-available power usage dataset called the UK-DALE dataset. The results exhibited a similar tendency to those we obtained in ACISP 2017; namely, our privacy-utility mechanism works highly effectively when high-power appliances such as kettle are designated as sensitive.

Keywords: privacy-preserving data mining, statistical inference, convex optimization, non-intrusive appliance load monitoring

1. Introduction

The proliferation of personal devices capable of Internet connectivity has enabled new applications and services [3]. Examples include healthcare advice service based on the user's activity data captured by fitness tracking devices, navigation services based on the GPS data from the user's smart phone, and demand response services based on the power consumption data of household smart-meters. Such new services will definitely enrich our everyday life.

At the same time, however, these services will collect users' personal data intentionally or unintentionally, which may in some cases violate their privacy [29]. In a well-known case, a retail company identified a teenage girl as pregnant based on her shopping habits [6], which can be thought of as illegal acquisition of sensitive information. The primary target of the paper is smart-meter data, which has been shown to potentially reveal the behavior of individuals [24, 25].

These privacy concerns in the era of Internet of Things have triggered re-examination of privacy regulation around the world. For instance, the EU Parliament passed the General Data Protection Regulation (GDPR) in 2016 which will be enforced in 2018. Most of the new privacy regulations, including the GDPR, now require explicitly that "natural persons should have control of their own personal data."[*1] It is therefore required for any service providers to treat users' personal data solicitously according to the demands of each individual. This social trend motivates the rapid development of privacy-preserving data mining technologies.

A prominent line of privacy-preserving techniques is k-anonymity [28,31] and its derivatives such as ℓ-diversity [21], t-closeness [20] and m-invariance [32]. Their primary goal is to convert an aggregation of personal data into a non-personal (anonymous) dataset while preserving information as much as possible. Although their privacy metrics are intuitive and easy to evaluate, it is difficult or almost impossible to protect users' privacy according to the detailed demands of each individual. Indeed, their basic strategy is to anonymize individuals by bundling similar records into indistinguishable bunches via generalization and omission of data. However, by nature of these metrics, privacy on an individual basis cannot be addressed.

Differential privacy [7, 8] is in another line of research. Unlike k-anonymity and its derivatives, differential privacy defines the privacy metrics based on a rigorous mathematical framework. The privacy definition of differential privacy is such that an adversary querying the database, which contains personal data of many individuals, should face difficulty in determining whether the data record of any specific individual is even in the database. Anonymity is their primary concern and accommodating users' specific privacy demands is therefore almost outside of their scope.

The most relevant work to ours is the consideration of privacy within a statistical inference attack framework [4,5,10,11,27]. In this framework, privacy is modeled as the amount of information obtained about the sensitive data when observing the released data. It is therefore possible

[1] Mitsubishi Electric Corporation
[a)] Hattori.Mitsuhiro@eb.MitsubishiElectric.co.jp
[*1] In Recital 7 of the GDPR.

to evaluate privacy on an individual basis by modeling the system with an appropriate definition of the sensitive and useful data. The primary goal of this framework is to find an optimal balance between privacy of an individual and utility of the service, and the problem of finding an optimal balance is formalized as an optimization problem where the objective function and constraint functions represent the privacy and utility. A solution of the optimization problem gives an optimal privacy mapping which distorts the useful data to obtain privacy while still proving utility.

The theoretical aspect of this framework is proposed and analyzed by du Pin Calmon and Fawaz [5]. Salamatian et al. applied the theory to a Census dataset and TV rating dataset, and showed that it is indeed possible to reduce the revelation of political affiliation while enabling TV program recommendation services [27]. Erdogdu et al. extended the theory to time-series datasets and applied the extended theory to energy disaggregation of smart-meter data [10, 11]. They showed that it is possible to modify power data to conceal the usage of a sensitive appliance while still allowing detection of the usage of a useful appliance, where the useful and sensitive appliances in their experiments were the washer-dryer and microwave, respectively.

Although Erdogdu et al. [10,11] made a significant step towards applying the theory to real systems, there is still much room for improvement. For example, they considered only the case where both the smart-meter data and usage data of the sensitive appliances are directly observable. However, in actual use cases such as ordinary smart-meter systems, individual appliance usage data may not be directly observable. Therefore, it is desirable to achieve the optimal privacy mapping even in the case where usage of sensitive appliances is not available.

These issues were partially solved by the work we presented at ACISP 2017 [15], where we modified the optimization problem of Erdogdu et al. [10, 11] in such a way that individual appliance energy usage data is not required. We also conducted in [15] several experiments of applying the proposed mechanism to the power usage data which we collected at an actual house. However, since the data was not publicly available, it was difficult for such researchers that are interested in the work to conduct replecation studies.

In this paper, we conduct additional experiments against a publicly-available power usage dataset called the UK-DALE dataset [18]. The procedure of the experiments is almost same as that we conducted in [15], except for the insertion of an additional step which converts the power usage data of individual appliances into binary operation data (ON and OFF). This additional step is required to fill a gap between the UK-DALE dataset and the ACISP dataset. The experimental results exhibited a similar tendency to those we obtained in [15]; namely, our privacy-utility mechanism works highly effectively when high-power appliances such as kettle are designated as sensitive. As with [15], we elaborate in this paper the steps we conducted, the parameters we computed and the inference results we obtained in detail, so that interested researchers can follow our work using the same or a similar dataset.

The rest of the paper is organized as follows. For the sake of self-containment and understandability, Sect. 2 reviews our target application and defines a system model and an adversary model, and Sect. 3 gives our theoretical analysis and proposition, both of which have already been given in [15]. Our experimental results against the UK-DALE dataset and discussions are described in Sect. 4. Section 5 concludes the paper with future directions.

2. Target Application: System and Adversarial Models

In this section, we first elaborate our target application and its privacy issue. Then we define a system model of the application and an adversary model of an "honest-but-curious" service provider.

2.1 Target Application and Privacy Issue

The target application we consider in this paper is an anomaly detection service of elderly residents living alone. More concretely, we consider an application where smart meter data, which is the aggregated power usage of all the appliances in a household, is collected from the house and disaggregated on a remote monitoring site, and appliance states are inferred whereby anomalies of the residents are detected. This service is proposed by Alcalá et al. [1, 2] and implemented by Song et al. [30].

The use of smart-meter for an anomaly detection is preferable in that unlike anomaly detection using additional sensors such as wearable medical devices, we need no extra devices since smart-meters have already been installed in many countries and are ready for use. The rapid development of energy disaggregation technologies, also known as non-intrusive appliance load monitoring (NILM) [14, 17, 19, 22, 23], also motivates the use of smart-meter as a sensor device for anomaly detection.

A straightforward way of implementing this service will be to disaggregate and detect the anomaly state on the user side and notify it to the service provider. However, energy disaggregation and anomaly detection could be too computationally intensive to be performed efficiently in a typical smart-meter with limited processing and memory capabilities. Besides, the correctness of anomaly detection can be improved by comparing the smart-meter data of a user with that of other users, which is easily conducted on the provider side but difficult on the user side.

The privacy issue we need to resolve in this application is that the service provider may infer states of the appliances that the user think of as sensitive, as well as those required for anomaly detection. For example, the kettle is ideal for anomaly detection because many people, especially those in the UK, use it regularly and also they often think of it as a non-sensitive appliance. The hairdryer, on the other hand, is useful but many people (especially women) would think of it as sensitive because usage of the dryer implies that the

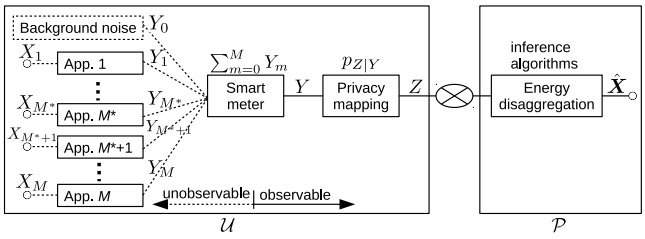

Fig. 1 Our system model. We assume App. 1 through App. M^* are the appliances that a user $\mathcal{U}$ designated as sensitive, and App. M^*+1 through App. M are those that $\mathcal{U}$ designated as non-sensitive. Solid lines represent observable data and dotted lines represent unobservable data.

user must have taken a bath. The difficulty of this issue lies in the fact that appliances in a household differ from person to person and the sensitivity to each appliance also differ. It is therefore required to develop a privacy technology that can prevent the service provider from inferring states of the appliances that the user thinks of as sensitive while allowing inference of states of the non-sensitive appliances, based on the preference of each user.

From the cryptography perspective, the service provider can be thought of as so-called "honest-but-curious" adversary, because he basically obeys the protocol (providing anomaly detection service to the user) but at the same time he tries to extract as much sensitive information as possible (inferring states of the appliances that the user think of as sensitive). We capture this adversarial situation with our adversary model in Sect. 2.2.

2.2 System and Adversary Models

2.2.1 System Model

Our system model is depicted in Fig. 1.

Suppose there are M appliances in the house of a user $\mathcal{U}$. Each appliance App. m ($m = 1, \ldots, M$) has several operating states denoted by $1, \ldots, K_m$, where $X_m \in \{1, \ldots, K_m\}$ denotes the realization of its operating state, and Y_m denotes its energy consumption. Note however that we cannot directly measure either X_m or Y_m, and can only measure the aggregated power usage $Y = \sum_{m=0}^{M} Y_m$ at the smart-meter, where Y_0 is the background noise. The smart-meter data Y is then passed to the privacy mapping module which takes as input Y and maps it into the distorted data Z. Here the mapping from Y to Z is according to the conditional probability distribution $p_{Z|Y}$ which is computed beforehand by solving the privacy-utility tradeoff problem proposed in Sect. 3. The distorted data Z is then sent to a service provider $\mathcal{P}$, and $\mathcal{P}$ will conduct energy disaggregation with Z and infer the appliance states $\hat{\boldsymbol{X}} = (\hat{X}_1, \ldots, \hat{X}_M)$ using some inference algorithms.

The smart-meter measures the power usage Y regularly (typically every one minute), and the distorted data Z is sent to $\mathcal{P}$ successively. $\mathcal{P}$ may store all the time-series $Z^{(1)}, \ldots, Z^{(T)}$ for some time period T (typically one day; we used nine days for our experiment in Sect. 4) and use them for inference of $\hat{\boldsymbol{X}}^{(1)}, \ldots, \hat{\boldsymbol{X}}^{(T)}$.

We should note here that although we modeled in Fig. 1 that the privacy mapping module is on the outside of the smart-meter, this is only for clarity and in practice it can be integrated into the smart-meter. Indeed, the privacy mapping operation is lightweight and can be executed with limited processing power and memory.

2.2.2 Adversary Model

The goal of an adversarial service provider $\mathcal{P}$ is to infer states of the appliances that $\mathcal{U}$ thinks of as sensitive. Suppose that $\mathcal{U}$ designated appliances App. 1 through App. M^* as sensitive and App. M^*+1 through App. M as non-sensitive. In this case, the adversarial goal of $\mathcal{P}$ is to infer $X_1, \ldots, X_{M^*}$ from Z.

We assume that $\mathcal{P}$ knows all the appliances in $\mathcal{U}$'s house. Also, we assume $\mathcal{P}$ knows the statistical distribution of each appliance.

The most typical probabilistic model used in energy disaggregation is the factorial hidden Markov model (FHMM) [12]. In FHMM, the emission distribution, transition probabilities and initial probabilities of all the appliances are used for inference of the hidden states. Therefore, concretely we make the following assumptions. First, we assume that $\mathcal{P}$ knows App. 1, ..., App. M, including the fact that $\mathcal{U}$ designated App. 1 through App. M^* as sensitive and App. M^*+1 through App. M as non-sensitive. $\mathcal{P}$ also knows the emission distribution $p_{Y_m^{(t)}|X_m^{(t)}}(y_m|x_{m,k})$ for all $m = 1, \ldots, M$ and $k = 1, \ldots, K_m$. I.e., we assume that $\mathcal{P}$ knows the probability distribution of the power usage of App. m at the state $x_{m,k}$, for all m and k. $\mathcal{P}$ additionally knows the transition probabilities $P_{X_m^{(t+1)}|X_m^{(t)}}(x_{m,k'}|x_{m,k})$ and the initial probabilities $P_{X_m^{(1)}}(x_{m,k})$ for all $m = 1, \ldots, M$ and $k, k' = 1, \ldots, K_m$, i.e., the probability with which App. m transits the state from $x_{m,k}$ to $x_{m,k'}$ when the time steps from t to $t+1$.

We now elaborate the justification of these assumptions. In actual use cases, $\mathcal{P}$ does not necessarily need to know the parameters for the sensitive appliances App. 1, ..., App. M^*. Namely, $\mathcal{P}$ does not need to know $p_{Y_m^{(t)}|X_m^{(t)}}(y_m|x_{m,k})$ and $P_{X_m^{(t+1)}|X_m^{(t)}}(x_{m,k'}|x_{m,k})$ for $m = 1, \ldots, M^*$. However, we make this assumption to consider a more adversarial $\mathcal{P}$.

3. A Proposed Privacy-Utility Tradeoff Mechanism

In this section, we modify the optimization problem of Erdogdu et al. [10, 11] in such a way that appliance usage data is not required. We formalize the optimization problem with definitions of privacy and utility in Sect. 3.1. Then in Sect. 3.2 we modify the problem by applying the linear Gaussian model assumption.

3.1 Formalization of the Problem

Here we formalize the privacy-utility tradeoff problem in a rigorous way.

3.1.1 Notation

Suppose $X \in \mathcal{X}$ is a discrete random variable and $Y \in \mathcal{Y}$ is a continuous random variable, where $\mathcal{X}$ and $\mathcal{Y}$ are some

(possibly infinite) sets. We use capital $P_X(x)$ for the probability mass function of X and small $p_Y(y)$ for the probability density function of Y. $E_Y[f(Y)]$ denotes the expected value of function $f(Y)$, i.e., $E_Y[f(Y)] = \int_{\mathcal{Y}} p_Y(y)f(y)dy$. We use $\mathcal{N}(\mu, \sigma^2)$ to denote the Gaussian distribution with mean μ and variance σ^2, and $p_{Y|X=x} \sim \mathcal{N}(\mu, \sigma^2)$ denotes that given that $X = x$, Y is conditionally distributed according to the Gaussian distribution with mean μ and variance σ^2.

Let $\boldsymbol{X} = (\boldsymbol{X}^*, \bar{\boldsymbol{X}})$ be a vector of discrete random variables representing the appliance states, where $\boldsymbol{X}^* = (X_1, X_2, \ldots, X_{M^*})$ are discrete random variables of the sensitive appliance states and $\bar{\boldsymbol{X}} = (X_{M^*+1}, X_{M^*+2}, \ldots, X_M)$ are those of the non-sensitive appliance states, both of which are designated by $\mathcal{U}$.

3.1.2 Definitions of Privacy and Utility

The privacy metric we consider in this paper is as follows.

Definition 1 (Privacy metric)**.** The privacy metric is the mutual information of sensitive appliance states $\boldsymbol{X}^*$ and distorted smart-meter data Z; i.e.,

$$I(\boldsymbol{X}^*; Z) = \sum_{\boldsymbol{x}^* \in \boldsymbol{\mathcal{X}}^*} P_{\boldsymbol{X}^*}(\boldsymbol{x}^*) \int_{\mathcal{Z}} p_{Z|\boldsymbol{X}^*}(z|\boldsymbol{x}^*) \log \frac{p_{Z|\boldsymbol{X}^*}(z|\boldsymbol{x}^*)}{p_Z(z)} dz \ . \tag{1}$$

The mutual information $I(\boldsymbol{X}^*; Z)$ represents the quantity of information one can obtain about $\boldsymbol{X}^*$ from the observed Z. It is therefore used extensively in the literature as a privacy metric [5, 10, 11, 26, 33]. Note however that $\boldsymbol{X}^*$ is a vector of discrete random variables while Z is a continuous random variable, which is different from the situation considered in the literature where all the random variables were discrete. We therefore extended the theory.

Utility is measured by the following distortion metric.

Definition 2 (Distortion metric)**.** Let $d : \mathcal{Y} \times \mathcal{Z} \to \mathbb{R}^+$ be some distortion function.[*2] The distortion metric is the expectation of $d(Y, Z)$; i.e.,

$$E_{Y,Z}[d(Y, Z)] = \iint_{\mathcal{Y}\times\mathcal{Z}} p_{Z|Y}(z|y)p_Y(y)d(y, z)dydz \ . \tag{2}$$

Lower distortion intuitively corresponds to better utility.

However, the distortion metric in Definition 2 may appear slightly different from what we should deal with in this paper. Indeed, the ideal distortion metric would be the one that directly captures the degradation of the results of appliance usage analysis. However, the outcome of the appliance usage analysis depends heavily on the algorithms used for the analysis and therefore it is infeasible to estimate the degradation in general. Also, empirically the distortion metric in Definition 2 is effective, as shown in Sect. 4.

3.1.3 The Privacy-Utility Tradeoff Problem

Suppose for now that the joint distribution $p_{\boldsymbol{X}^*,Y}$ is already known. Then given $p_{\boldsymbol{X}^*,Y}$, a distortion function d and a distortion constraint δ, the privacy mapping $p_{Z|Y}$ that minimizes the privacy information leakage can be found by solving the following optimization problem:

$$\inf_{p_{Z|Y}} I(\boldsymbol{X}^*; Z) \quad \text{subject to } E_{Y,Z}[d(Y, Z)] \le \delta \ . \tag{3}$$

[*2] Examples of distortion function include the L_1 norm, L_2 norm and more generally L_p norm.

3.2 Gaussian Model Assumption

We assumed in Sect. 3.1 that $p_{\boldsymbol{X}^*,Y}$ is already known. In practical smart-meter systems, however, this assumption does not hold and we need to substitute $p_{\boldsymbol{X}^*,Y}$ with other known parameters. We propose here the substitution method.

First, observe that from the law of total probability,

$$\begin{aligned} p_{\boldsymbol{X}^*,Y}(\boldsymbol{x}^*, y) &= \sum_{\bar{\boldsymbol{x}} \in \bar{\boldsymbol{\mathcal{X}}}} p_{\boldsymbol{X}^*,\bar{\boldsymbol{X}},Y}(\boldsymbol{x}^*, \bar{\boldsymbol{x}}, y) = \sum_{\bar{\boldsymbol{x}} \in \bar{\boldsymbol{\mathcal{X}}}} p_{\boldsymbol{X},Y}(\boldsymbol{x}, y) \\ &= \sum_{\bar{\boldsymbol{x}} \in \bar{\boldsymbol{\mathcal{X}}}} P_{\boldsymbol{X}}(\boldsymbol{x}) p_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) \ . \end{aligned} \tag{4}$$

Now, computing $p_{\boldsymbol{X}^*,Y}(\boldsymbol{x}^*, y)$ boils down to computing $P_{\boldsymbol{X}}(\boldsymbol{x})$ and $p_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$.

In order to compute $p_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$, we apply a linear Gaussian model. This model has been used extensively to simulate the emission of home appliances in the energy disaggregation literature [13, 19, 22].

Let Y_0 be a random variable of the background noise and Y_m be that of the emission of appliance m. Then,

$$Y = Y_0 + \sum_{m=1}^{M} Y_m \ , \tag{5}$$

$$p_{Y_0} \sim \mathcal{N}(\mu_0, \sigma_0^2) \ , \tag{6}$$

$$p_{Y_m|X_m=x_{m,k}} \sim \mathcal{N}(\mu_{m,k}, \sigma_{m,k}^2) \ , \tag{7}$$

where μ_0 and σ_0^2 are the mean and variance of the Gaussian distribution of the background noise, and $\mu_{m,k}$ and $\sigma_{m,k}^2$ are those of appliance m in state k. Then, according to the standard probability theory [9],

$$p_{Y|\boldsymbol{X}=\boldsymbol{x}} \sim \mathcal{N}\left(\mu_0 + \sum_{m=1}^{M} \mu_{m,k},\ \sigma_0^2 + \sum_{m=1}^{M} \sigma_{m,k}^2\right). \tag{8}$$

Equation (8) implies that computing $p_{Y|\boldsymbol{X}}$ is now reduced to obtaining the parameters $\Theta = \{\mu_0, \sigma_0^2, \{\mu_{m,k}, \sigma_{m,k}^2\}\}$. These parameters can be obtained either from the specification documents or reference models of the appliances, or by doing preliminary training activities.

Assuming that the variance of the total power data Y is independent of states of the appliances, (8) can further be simplified as

$$p_{Y|\boldsymbol{X}=\boldsymbol{x}} \sim \mathcal{N}\left(\mu_0 + \sum_{m=1}^{M} \mu_{m,k},\ \sigma^2\right), \tag{9}$$

where σ^2 is the variance of Y. In this case, computing $p_{Y|\boldsymbol{X}}$ can be reduced to obtaining the parameters $\Theta' = \{\mu_0, \{\mu_{m,k}\}, \sigma^2\}$. We use this simplified model in Sect. 4.

$P_{\boldsymbol{X}}(\boldsymbol{x})$ can also be obtained from the reference models of the appliances or by doing preliminary training activities.

Now, it is easy to see that $p_{\boldsymbol{X}^*,Y}$ can be obtained from Θ' and $P_{\boldsymbol{X}}$ and therefore the optimization problem is solvable.

4. Experiments on Household Power Usage Data

This section exhibits our experimental results of applying the proposed mechanism to the power usage data of an actual household. We give an overview of our experiments in Sect. 4.1, and we discuss in Sect. 4.2 the power usage dataset that we used for the experiments. Section 4.3 shows the datasets and parameters that we obtained in the experiments. The optimization problem is solved and the privacy mapping is applied in Sect. 4.4. Section 4.5 evaluates the privacy and utility aspects of our mechanism quantitatively, and the implications of the results are discussed in Sect. 4.6.

4.1 Overview

We used as an actual household dataset the UK-DALE dataset. This dataset contains both the whole-house power usage data and the individual power usage data of each appliance. In order to fill a gap between the UK-DALE dataset and the dataset we used in the ACISP 2017 work [15] (*the ACISP dataset*, hereafter), we converted the individual power usage data into binary operation data (ON and OFF) using the threshold power usage value described in the metadata of the UK-DALE dataset.

Then, we considered four use cases: 1) tv is designated as sensitive; 2) kettle is designated as sensitive; 3) toaster is designated as sensitive; and 4) hair dryer is designated as sensitive. For each case, we chose an appropriate distortion constraint δ by trial-and-error, and with Θ', $P_{\boldsymbol{X}}$ and δ, we solved the convex optimization problem (3) and obtained a privacy mapping $p_{Z|Y}$. We then distorted the power usage data according to $p_{Z|Y}$, and obtained distorted power usage data. In order to evaluate the privacy and utility of our mechanism, we applied an inference algorithm to the distorted data to infer the appliance usage of the sensitive and non-sensitive appliances, and compared the performance with that of the original data.

Since the UK-DALE dataset contains discrete values only, in the experiments we regard Y and Z as discrete random variables $\tilde{Y}$ and $\tilde{Z}$ respectively, and compute and apply a conditional probability mass function $P_{\tilde{Z}|\tilde{Y}}$.

4.2 Data Source

We use as a publicly available dataset the UK-DALE dataset [18] for our experiment. Among several publicly available datasets, this is one of the most desirable dataset in that it contains fine-grained and long-period power usage data as well as detailed metadata and it can be readily available from the website.[*3] Details of the dataset including the environment of the data acquisition, types of the data obtained, and the statistics of the data, can be available at [16, 18].

[*3] http://jack-kelly.com/data/

4.3 Datasets and Parameters

4.3.1 Power Usage and Appliance Usage Datasets

The UK-DALE dataset contains five household data. For our experiment, we use House 1 data because it has minimum data loss among the five.[*4] The House 1 data consists of three different forms of data: the 6 second data, the 1 second data and the 16 kHz data, each of which is captured according to the specified frequency (6 second means power usage is captured every 6 seconds). We use 6 second data, and downsample it to 1 minute resolution in order to obtain a dataset whose time resolution is identical to those of the ACISP dataset.

The difference between the UK-DALE dataset and the ACISP dataset is not only the time resolution; the format of the appliance usage data is also different. In the ACISP dataset, the appliance usage is collected manually in binary form (ON and OFF) based on actual operation of the appliance. In the UK-DALE dataset, however, only the actual power usage of each appliance is available.[*5] We therefore convert the power usage data into binary operation data using the threshold power usage value given in the metadata of the UK-DALE dataset.

The House 1 data starts on November 9, 2012, 22:28:15 GMT and ends on April 26, 2017, 18:35:53 BST. In our experiment, we use the one-month data starting on August 1, 2016, 0:00 GMT and ending on August 31, 2016, 23:59 for both the supervised learning of the model parameters described in Sect. 4.3.2 and the evaluation of privacy and utility described in Sect. 4.5.

4.3.2 Model Parameters

In order to obtain the model parameters Θ' and $P_{\boldsymbol{X}}$ from the power usage data and the ground truth, we used a supervised learning algorithm in the same way as we did against the ACISP dataset.

For simplicity, we employed a couple of simplification techniques. First, we modeled the hidden states of the appliances with the factorial hidden Markov model (FHMM) [12]. The FHMM assumes that the hidden states between appliances are independent, which reduces the computational complexity of learning and inference. This assumption is reasonable in our situation and therefore we used this model to simplify the computation of $\Theta' = \{\mu_0, \{\mu_{m,k}\}, \sigma^2\}$.

Second, we assumed each appliance has only two possible states: $\mathcal{X}_m = \{\mathsf{ON}, \mathsf{OFF}\}$ for all $m \in \{1, 2, \ldots, M = 53\}$. This two-state assumption simplifies the computation of $P_{\boldsymbol{X}}$. Note here that since we have assumed all the appliances behave independently from each other, we can compute $P_{\boldsymbol{X}}(\boldsymbol{x})$ as the product of probability of each appliance; i.e., $P_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{m=1}^{M} P_{X_m}(x_m)$. We also assume that the appliance state Markov chains have already converged to the steady-state, that is, the initial state distributions are equal to the steady-state distributions implied by the transi-

[*4] Refer to Figure 3 in [18] for details.

[*5] To be precise, several appliances are accompanied by a binary operation data. However, they are not necessarily perfect and reliable.

Table 1 Appliances used in the House 1 of the UK-DALE dataset where the estimated mean power is greater than 100 Watts, and their parameters obtained from the supervised learning. $\mu_{m,\mathsf{ON}}$ is the estimated mean power of appliance m. a_m and b_m are the estimated transition probabilities of transiting from OFF to ON and from ON to OFF, respectively.

m	Appliance	$\mu_{m,\mathsf{ON}}$	a_m	b_m
0	background noise	239.68		
5	washing machine	425.42	0.001394	0.037783
6	dishwasher	1013.42	0.000762	0.024756
7	tv	231.73	0.001455	0.009122
8	kitchen lights	123.16	0.00847	0.050728
10	kettle	2300.00	0.002338	0.675325
11	toaster	1601.85	0.000898	0.377358
13	microwave	1381.56	0.000381	0.253731
22	hoover	1688.85	0.000472	0.164063
36	coffee machine	179.18	0.000135	0.022814
39	hair dryer	793.82	0.000202	0.230769
40	straighteners	226.56	0.00008963	0.266667
44	child's table lamp	313.50	0.000134	1.000
49	office lamp2	130.49	0.000113	0.013736
51	office pc	265.53	0.000293	0.058559
53	LED printer	144.12	0.0000672	0.103448
				$\sigma^2 = 68618.95$

tion distributions. Thus, each $P_{X_m}(x_m)$ is stationary across time and can be computed from the transition probabilities of the appliance states.

Let a_m be the transition probability of appliance m from OFF to ON and b_m be that of the opposite direction (ON to OFF). Then,

$$P_{X_m}(\mathsf{ON}) = \frac{a_m}{a_m + b_m}, \quad P_{X_m}(\mathsf{OFF}) = \frac{b_m}{a_m + b_m}. \quad (10)$$

Hence, $P_{\boldsymbol{X}}$ can be computed by $\{a_m, b_m\}$. In addition, we assumed that $\mu_{m,\mathsf{OFF}} = 0$ for all m.

We used all of the one-month data of power usage and appliance usage for the supervised learning, and obtained $\Theta' = \{\mu_0, \{\mu_{m,\mathsf{ON}}\}, \sigma^2\}$ and $\{a_m, b_m\}$. Especially, the mean values are computed by solving the normal equation of the linear regression. For the sake of simplicity, though, we discarded the appliances whose estimated mean power is below 100 Watts and regarded them as a part of the background noise. We repeated the computation of the normal equation several times while excluding the low-power-consuming appliances. The results are shown in Table 1.

4.4 Optimization and Distortion

As we explained in Sect. 4.1, we considered the following four use cases:

Case 1 tv ($m = 7$) is designated as sensitive,
Case 2 kettle ($m = 10$) is designated as sensitive,
Case 3 toaster ($m = 11$) is designated as sensitive,
Case 4 hair dryer ($m = 39$) is designated as sensitive.

For each case, we solved the convex optimization problem and obtained a discrete privacy mapping $P_{\tilde{Z}|\tilde{Y}}$. For ease of computation, though, we quantized $\tilde{y}$ and $\tilde{z}$ into 20-Watt-resolution data $\breve{y}$ and $\breve{z}$, respectively, and computed $P_{\breve{Z}|\breve{Y}}$ as an alternative of $P_{\tilde{Z}|\tilde{Y}}$. We used as a distortion metric the L_1 distance $d(\breve{y}, \breve{z}) = |\breve{y} - \breve{z}|$. The optimization problem was solved by the convex optimization software CVX,[*6] where we used $\delta = 750$ for Case 1, $\delta = 10$ for Case 2, $\delta = 75$ for Case 3 and $\delta = 100$ for Case 4. Then we distorted the power usage data according to $P_{\breve{Z}|\breve{Y}}(\breve{z}|\breve{y})$.

[*6] http://cvxr.com/cvx/

4.5 Evaluation of Privacy and Utility

We now evaluate both the privacy and utility aspects of the distorted power usage data.

Since our goal of the privacy-utility tradeoff is to retain the inference of the non-sensitive appliance states while preventing that of the sensitive appliance states, we evaluate them by measuring the degradation of the appliance state inference. We therefore apply an inference algorithm to the raw data and the distorted data (for all the four cases) to infer the hidden states of the appliances, and evaluate the detection rates.

We again model the hidden states with the FHMM accompanied by the parameters we obtained in the supervised learning, and infer the hidden states for the same one month using an approximate inference algorithm called the completely factorized variational approximation (CFVA) [12]. CFVA is used to avoid the computational complexity of exact inference algorithms. For this binary (ON and OFF) classification, the CFVA algorithm provides marginal posterior likelihoods which we can threshold at custom values to obtain a receiver operating characteristic (ROC) curve in order to evaluate the inference performance across different tradeoffs between true positive and false positive rates. We can also compute the area under the curve (AUC) which quantifies the inference performance across this tradeoff in a single number. We perform and compare this evaluation between the raw data and the distorted data.

Figure 2 shows the ROC curve of the inference results of several appliances, where the analysis was performed on the raw dataset. The AUC values are evaluated and shown in Table 2. As the AUC values tell, the states of the kettle and oven toaster are inferred almost correctly, the states of the tv and hair dryer are inferred with high accuracy, and the states of the office pc are inferred with marginal accuracy.

Figure 3 gives ROC curves of the inference results with the distorted data for Case 1. The inference performance for the target appliance is degraded as required, but at the same time other appliances are also degraded severely.

Figure 4 gives ROC curves of the inference results with the distorted data for Case 2. The inference performance for the target appliance is degraded as required, and the other appliances are not damaged at all.

Figure 5 gives ROC curves of the inference results with the distorted data for Case 3. The inference performance for the toaster is degraded severely. The inference performance for the kettle and hair dryer is degraded to some extent. The inference performance for the tv and office pc is preserved almost identically.

Figure 6 gives ROC curves of the inference results with the distorted data for Case 4. The inference performance for the hair dryer is degraded severely. The inference performance for the kettle and toaster is degraded to some extent. The

Table 2 AUC values of the ROC curves (Figures 2, 3, 4, 5 and 6)

		AUC				
m	Appliance	raw (Fig. 2)	tv (Fig. 3)	kettle (Fig. 4)	toaster (Fig. 5)	hair dryer (Fig. 6)
7	tv	0.927	*0.542*	0.928	0.923	0.916
10	kettle	0.988	0.783	*0.523*	0.821	0.846
11	toaster	0.982	0.822	0.931	*0.558*	0.692
39	hair dryer	0.951	0.659	0.964	0.664	*0.611*
51	office pc	0.887	0.499	0.895	0.877	0.874

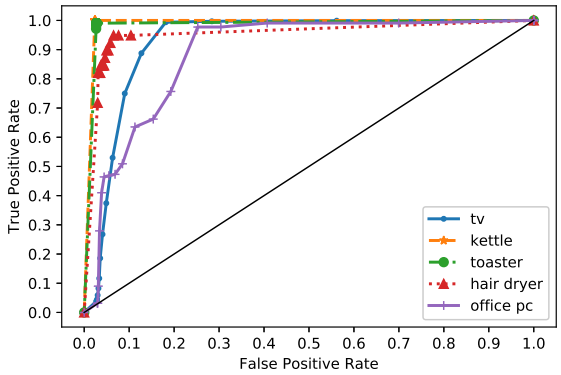

Fig. 2 ROC curves of the results of inference with raw data

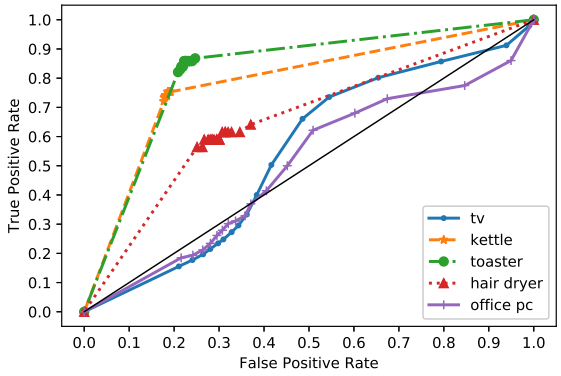

Fig. 3 ROC curves of the results of inference with distorted data (sensitive appliance is tv and $\delta = 750$)

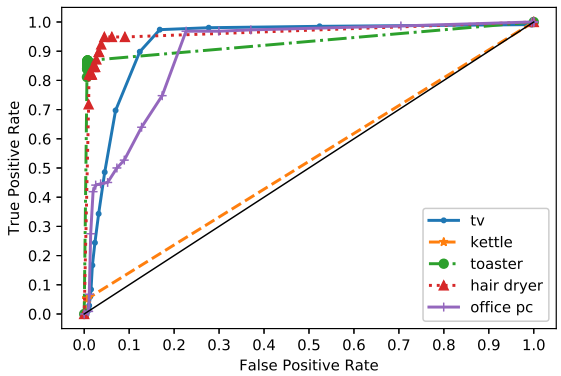

Fig. 4 ROC curves of the results of inference with distorted data (sensitive appliance is kettle and $\delta = 10$)

inference performance for the tv and office pc is preserved almost identically.

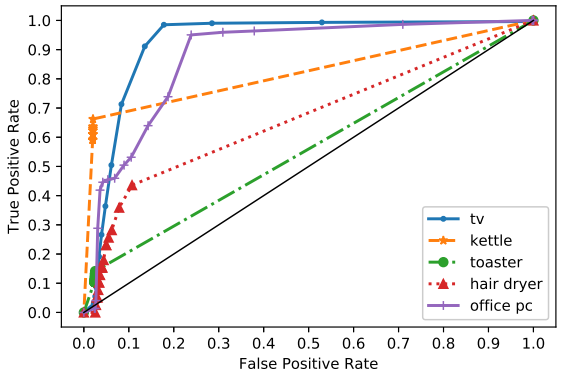

Fig. 5 ROC curves of the results of inference with distorted data (sensitive appliance is toaster and $\delta = 75$)

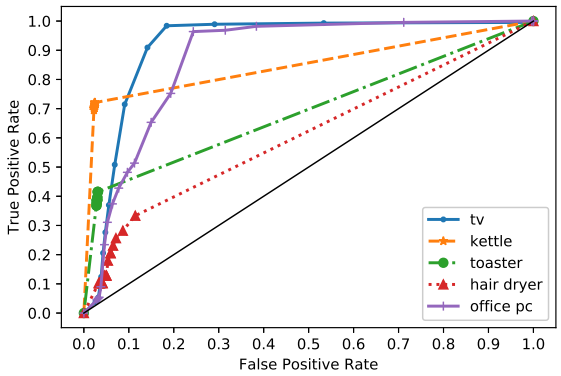

Fig. 6 ROC curves of the results of inference with distorted data (sensitive appliance is hair dryer and $\delta = 100$)

4.6 Discussion

As we have shown in Sect. 4.5, the distortion works highly effectively for the case where the sensitive appliance is the kettle. This may be due to the fact that the kettle is realistically modeled with only two states: {ON, OFF}, and therefore our simplified model fits well. Moreover, the consumed power is as high as 2,300 Watts, which enables us to compute an optimal privacy mapping $P_{\tilde{Z}|\tilde{Y}}$ that attains both small mutual information and small distortion such as $\delta = 10$.

On the other hand, for the case where the sensitive appliance is the television, the distortion renders inference of the sensitive appliance almost impossible but at the same time makes inference of the other appliances degraded severely. This may stem from the fact that the television consumes a relatively low power of 232 Watts and thus distortion of middle power values would affect other middle-power appli-

ances.

We should discuss the impact of the assumptions and approximations we made in the evaluation. We modeled the hidden states of the appliances with FHMM. FHMM is used typically in the energy disaggregation literature [19] and therefore this can be thought of as a reasonable modeling, but other inference algorithms such as neural networks [17] may give greater advantage to an adversarial $\mathcal{P}$. We used a binary-state (ON and OFF) assumption for all the appliances. This may fit to some appliances (e.g. kettle) but not to others (e.g. tv). Multiple-state model will definitely give higher performance to both benign and adversarial $\mathcal{P}$. Use of an exact inference algorithm will make the performance better at the price of computational complexity.

5. Conclusion

In this paper, we extended the work we presented at ACISP 2017 by conducting additional experiments against a publicly-available power usage dataset called the UK-DALE dataset. We followed almost the same procedure as that of the ACISP 2017 work, and obtained the results exhibiting a similar tendency; namely, our privacy-utility mechanism works highly effectively when high-power appliances such as kettle are designated as sensitive. Since we elaborated the steps we conducted, the parameters we computed and the inference results we obtained in detail, interested researchers can follow our work using the same or a similar dataset.

Future work will be to extend this theory to the case where the service provider uses other inference algorithms such as neural networks.

Acknowledgments The authors would like to thank Jack Kelly, William Knottenbelt and their anonymous colleagues for releasing their own power usage data.

References

[1] Alcalá, J., Parson, O. and Rogers, A.: Detecting Anomalies in Activities of Daily Living of Elderly Residents via Energy Disaggregation and Cox Processes, *BuildSys 2015*, ACM, (online), DOI: https://doi.org/10.1145/2821650.2821654 (2015).

[2] Alcalá, J., Ureña, J. and Hernández, A.: Activity supervision tool using Non-Intrusive Load Monitoring Systems, *ETFA 2015*, IEEE, (online), DOI: https://doi.org/10.1109/ETFA.2015.7301622 (2015).

[3] Atzori, L., Iera, A. and Morabito, G.: The Internet of Things: A survey, *Computer Networks*, Vol. 54, No. 15, pp. 2787–2805 (online), DOI: https://doi.org/10.1016/j.comnet.2010.05.010 (2010).

[4] Basciftci, Y. O., Wang, Y. and Ishwar, P.: On privacy-utility tradeoffs for constrained data release mechanisms, *ITA 2016*, IEEE, (online), DOI: https://doi.org/10.1109/ITA.2016.7888175 (2016).

[5] du Pin Calmon, F. and Fawaz, N.: Privacy against statistical inference, *Allerton 2012*, IEEE (2012).

[6] Duhigg, C.: How companies learn your secrets, *The New York Times Magazine*, Vol. 16, p. 2012 (2012).

[7] Dwork, C.: Differential privacy, *ICALP 2006*, Springer Berlin Heidelberg, pp. 1–12 (2006).

[8] Dwork, C.: Differential privacy: A survey of results, *TAMC 2008* (Agrawal, M., Du, D., Duan, Z. and Li, A., eds.), Lecture Notes in Computer Science, Vol. 4978, Springer Berlin Heidelberg, pp. 1–19 (2008).

[9] Eisenberg, B. and Sullivan, R.: Why is the sum of independent normal random variables normal?, *Mathematics Magazine*, Vol. 81, No. 5, pp. 362–366 (2008).

[10] Erdogdu, M. A. and Fawaz, N.: Privacy-utility trade-off under continual observation, *ISIT 2015*, IEEE, pp. 1801–1805 (online), DOI: https://doi.org/10.1109/ISIT.2015.7282766 (2015).

[11] Erdogdu, M. A., Fawaz, N. and Montanari, A.: Privacy-utility trade-off for time-series with application to smart-meter data, *AAAI 2015 Workshop on Computational Sustainability* (2015).

[12] Ghahramani, Z. and Jordan, M. I.: Factorial hidden Markov models, *Machine Learning*, Vol. 29, No. 2/3, pp. 245–273 (1997).

[13] Guo, Z., Wang, Z. J. and Kashani, A.: Home appliance load modeling from aggregated smart meter data, *IEEE Transactions on Power Systems*, Vol. 30, No. 1, pp. 254–262 (online), DOI: https://doi.org/10.1109/TPWRS.2014.2327041 (2015).

[14] Hart, G.: Nonintrusive appliance load monitoring, *Proceedings of the IEEE*, Vol. 80, No. 12, pp. 1870–1891 (1992).

[15] Hattori, M., Hirano, T., Matsuda, N., Shimizu, R. and Wang, Y.: Privacy-Utility Tradeoff for Applications Using Energy Disaggregation of Smart-Meter Data, *ACISP 2017, Part II* (Pieprzyk, J. and Suriadi, S., eds.), LNCS, Vol. 10343, pp. 214–234 (online), DOI: https://doi.org/10.1007/978-3-319-59870-3_12 (2017).

[16] Kelly, J.: Disaggregation of domestic smart meter energy data, PhD Thesis, University of London (2017).

[17] Kelly, J. and Knottenbelt, W.: Neural NILM: Deep neural networks applied to energy disaggregation, *BuildSys 2015*, ACM (2015).

[18] Kelly, J. and Knottenbelt, W.: The UK-DALE dataset, domestic appliance-level electricity demand and whole-house demand from five UK homes, *Scientific Data*, Vol. 2, No. 150007, p. 150007 (2015).

[19] Kim, H., Marwah, M., Arlitt, M., Lyon, G. and Han, J.: Unsupervised disaggregation of low frequency power measurements, *SDM11*, SIAM, pp. 747–758 (2011).

[20] Li, N., Li, T. and Venkatasubramanian, S.: t-Closeness: Privacy beyond k-anonymity and l-diversity, *ICDE 2007*, IEEE, pp. 106–115 (2007).

[21] Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkitasubramaniam, M.: l-diversity: Privacy beyond k-anonymity, *ACM Transactions on Knowledge Discovery from Data*, Vol. 1, No. 1, p. 3 (2007).

[22] Parson, O., Ghosh, S., Weal, M. and Rogers, A.: Non-intrusive load monitoring using prior models of general appliance types, *AAAI 2012*, pp. 356–362 (2012).

[23] Parson, O., Ghosh, S., Weal, M. and Rogers, A.: An unsupervised training method for non-intrusive appliance load monitoring, *Artificial Intelligence*, Vol. 217, pp. 1–19 (2014).

[24] Pillitteri, V. Y. and Brewer, T. L.: Guidelines for smart grid cybersecurity, Internal Report NISTIR 7628 Revision 1, National Institute of Standards and Technology (2014).

[25] Quinn, E. L.: Smart metering and privacy: Existing laws and competing policies, *SSRN Electronic Journal* (2009).

[26] Rajagopalan, S. R., Sankar, L., Mohajer, S. and Poor, H. V.: Smart meter privacy: A utility-privacy framework, *SmartGridComm 2011*, IEEE (2011).

[27] Salamatian, S., Zhang, A., du Pin Calmon, F., Bhamidipati, S., Fawaz, N., Kveton, B., Oliveira, P. and Taft, N.: Managing your private and public data: Bringing down inference attacks against your privacy, *IEEE Journal of Selected Topics in Signal Processing*, Vol. 9, No. 7, pp. 1240–1255 (2015).

[28] Samarati, P.: Protecting respondents identities in microdata release, *IEEE Transactions on Knowledge and Data Engineering*, Vol. 13, No. 6, pp. 1010–1027 (2001).

[29] Schneier, B.: *Data and goliath: The hidden battles to collect your data and control your world*, W. W. Norton & Company (2015).

[30] Song, H., Kalogridis, G. and Fan, Z.: Short paper: Time-dependent power load disaggregation with applications to daily activity monitoring, *WF-IoT 2014*, IEEE (2014).

[31] Sweeney, L.: k-anonymity: A model for protecting privacy, *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, Vol. 10, No. 05, pp. 557–570 (2002).

[32] Xiao, X. and Tao, Y.: m-invariance: Towards privacy preserving re-publication of dynamic datasets, *SIGMOD 2007*, ACM, pp. 689–700 (2007).

[33] Yang, W., Li, N., Qi, Y., Qardaji, W., McLaughlin, S. and McDaniel, P.: Minimizing private data disclosures in the smart grid, *ACM CCS 2012*, ACM, pp. 415–427 (2012).