Privacy-Utility Tradeoff for the Appliance Usage Analysis of Smart-Meter Data

Mitsuhiro Hattori[1,a)] Takato Hirano[1] Nori Matsuda[1] Rina Shimizu[1]
Ye Wang[2]

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 a reference power usage dataset. However, their theory assumes both power usage data and sensitive appliance state information are available when computing the privacy-preserving mechanism, which is impractical in typical smart-meter systems. We propose in this paper a privacy-utility tradeoff mechanism for the smart-meter systems in which sensitive information is not directly observable. We apply a linear Gaussian model to the system and thereby reduce the problem of obtaining unobservable information to that of learning the system parameters. Experimental results show that the proposed mechanism works effectively; i.e. it prevents usage analysis of sensitive appliances while at the same time preserving that of non-sensitive appliances.

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

1. Introduction

1.1 Background

The proliferation of personal devices capable of Internet connectivity has been promoting brand-new applications and services. 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 [16]. In a well-known case, a retail company identified a teenage girl as pregnant based on her shopping habits [5], which can be thought of as illegal acquisition of sensitive information. As for smart-meter data, which is the primary target of the paper, it is noted that analyzing them may lead to behavioral inferences of individuals [13].

These privacy concerns in the era of Internet of Things have triggered re-examination of privacy regulation around the world. The Japanese Diet passed amendments to the Personal Information Protection Act in 2015. The White House released the Consumer Privacy Bill of Rights Act of 2015. The EU Parliament passed the General Data Protection Regulation (GDPR) in 2016. In the GDPR especially, it is noted explicitly that "natural persons should have control of their own personal data." This requires service providers to treat users' personal data solicitously according to the demands of each individual.

Balancing between the utility of services and the privacy of individuals is therefore important for the success of personalized services, and various kinds of privacy-preserving data mining technologies have been proposed accordingly.

1.2 Related Work

The most prominent technology is k-anonymity [15], [17] and its derivatives such as ℓ-diversity [12] and t-closeness [11]. 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. Demands of individuals therefore are not taken into account.

Differential privacy [6] 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 not be able

[1] Mitsubishi Electric Corporation
[2] Mitsubishi Electric Research Laboratories
[a)] Hattori.Mitsuhiro@eb.MitsubishiElectric.co.jp

to determine 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], [7], [8], [14]. 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 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 [4]. 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 [14]. Erdogdu et al. extended the theory to time-series datasets and applied the extended theory to smart-meter data [7], [8]. 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. reveal the results of experiments, they do not exhibit the details of the experiments. Moreover, 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.

1.3 Contribution

We propose in this paper a privacy-utility tradeoff mechanism for the smart-meter systems in which appliance usage is unobservable. In order to complement the lack of information needed for the tradeoff computation, we apply a linear Gaussian model to the system and thereby reduce the problem of determining an unknown system model to that of learning the model parameters such as the mean power consumption of each appliance and the transition probabilities of appliance states. These system parameters are, in some cases, available without conducting supervised learning on each household, because the mean power is often listed on a specification document of the appliance and the transition probabilities can be simulated. Therefore, our mechanism is considered to be practical. Additionally, we extend the theory to the case where the privacy metric is defined by a mixture of continuous and discrete random variables.

In order to show the practicality of our theory, we conduct several experiments of applying the proposed mechanism to the power usage data of an actual household. We collected power usage data for nine days, and we also manually collected the ground truth appliance usage for the same period to compute the system parameters. Optimal privacy-utility tradeoffs are computed for two use cases, and the raw power data is distorted according to the optimized mechanism. We evaluate the privacy and utility aspects by examining degradation of appliance usage inference performance. It is shown quantitatively that our mechanism is reasonable and effective, especially when high-power appliances such as the oven toaster are designated as sensitive. 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.

1.4 Organization of the Paper

The rest of the paper is organized as follows. Section 2 introduces our notations and several useful facts. Our theoretical analysis and proposition is given in Section 3, and experimental results and discussions are described in Section 4. Section 5 concludes the paper with future directions.

2. Preliminaries

In this section, we give notations used throughout the paper and present an information-theoretic definition that we utilize.

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. $E_Y[f(Y)] = \int_{\mathcal{Y}} p_Y(y) f(y) dy$. $\mathcal{N}(\mu, \sigma^2)$ denotes the Gaussian distribution with mean μ and variance σ^2. $p_{Y|X=x} \sim \mathcal{N}(\mu, \sigma^2)$ denotes that given the condition that $X = x$, Y distributes according to the Gaussian distribution with mean μ and variance σ^2.

As in [1], the mutual information of a discrete random variable X and continuous random variable Y is defined as

$$I(X;Y) = \sum_{x \in \mathcal{X}} P_X(x) \int_{\mathcal{Y}} p_{Y|X}(y|x) \log \frac{p_{Y|X}(y|x)}{p_Y(y)} dy.$$

We use this type of mutual information for our privacy metric in Section 3.

3. A Proposed Privacy-Utility Tradeoff Mechanism

This section describes the theoretical aspects of our privacy-utility tradeoff mechanism. We first define in Section 3.1 the goal of our tradeoff mechanism in an abstract way. In Section 3.2, we turn the problem of pursuing the abstract goal into a formal optimization problem by model-

ing the tradeoff situation with the rigorous privacy and distortion metrics. The optimization problem contains an unknown probability distribution for the system model, which we assume has linear Gaussian form in Section 3.3. Section 3.4 modifies the optimization problem a bit to accommodate discrete power usage data.

3.1 The Goal of the Privacy-Utility Tradeoff

We consider a situation where a privacy-conscious user $\mathcal{U}$ releases the (possibly distorted) smart-meter data to a service provider $\mathcal{P}$ in the prospect of receiving some useful service(s). The primary goal of $\mathcal{P}$ is to recover the appliance usage from the released smart-meter data, so that it can provide some utility to $\mathcal{U}$. The primary concern of $\mathcal{U}$ is that $\mathcal{P}$ may recover the usage of the appliances that $\mathcal{U}$ considers sensitive. The privacy-utility tradeoff we consider in this paper is therefore to retain the non-sensitive appliance usage analysis as much as possible while at the same time preventing the usage analysis of the sensitive appliances that $\mathcal{U}$ designates.

This goal is exemplified by the following application. Suppose $\mathcal{U}$ is an elderly person living alone and $\mathcal{P}$ is a security company providing a remote monitoring service. $\mathcal{P}$ monitors $\mathcal{U}$'s state by extracting the appliance usage data from the smart-meter data of $\mathcal{U}$'s household. For example, if $\mathcal{P}$ extracts information that a microwave is used in the morning, then $\mathcal{P}$ will understand that $\mathcal{U}$ is in the normal state. Here, $\mathcal{U}$ may think that a dryer should not be detected because it tells when $\mathcal{U}$ took a bath. In this case, the dryer is considered as a sensitive appliance while other appliances including a microwave are non-sensitive.

3.2 Formalization of the Problem

In this section, we formalize the privacy-utility tradeoff problem given in Section 3.1 in a rigorous way.

3.2.1 Notation

Suppose there are M appliances named $1, 2, \ldots, M$ in $\mathcal{U}$'s household, and out of them $\mathcal{U}$ designates $M^*(< M)$ appliances as sensitive. For instance, $\mathcal{U}$ may have 17 appliances as shown in **Table 1**, and among them $\mathcal{U}$ may designate television as sensitive. Without loss of generality, we assume appliances $1, 2, \ldots, M^*$ are designated as sensitive and appliances $M^*+1, M^*+2, \ldots, M$ are non-sensitive.

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.

Appliance m takes K_m states; i.e. $X_m \in \mathcal{X}_m = \{x_{m,1}, x_{m,2}, \ldots, x_{m,K_m}\}$. For example, a dryer may take one of three states: $\mathcal{X} = \{\mathsf{STRONG}, \mathsf{WEAK}, \mathsf{OFF}\}$. $\boldsymbol{X}$ takes one of $\boldsymbol{\mathcal{X}} = \mathcal{X}_1 \times \mathcal{X}_2 \times \cdots \times \mathcal{X}_M$ states.

Let $Y \in \mathcal{Y} \subseteq \mathbb{R}$ be a continuous random variable representing the observed smart-meter data and $Z \in \mathcal{Z} \subseteq \mathbb{R}$ represent the distorted data. Our goal is to find the distortion distribution $p_{Z|Y}$ that attains the optimal privacy-utility tradeoff.

3.2.2 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 [4], [7], [8], [14]. 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[*1]. The distortion metric is the expectation of $d(Y, Z)$; i.e.,

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

The lower the distortion is, the better the utility should be, intuitively.

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 Section 4.

3.2.3 The Privacy-Utility Tradeoff Problem

Suppose for now that the joint distribution $p_{\boldsymbol{X}^*,Y}$ is already known. Then, it is easy to see that 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:

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

Indeed, as we assume that $p_{\boldsymbol{X}^*,Y}$ is already known, we can actually compute both the objective function (Equation 1) and constraint function (Equation 2), and therefore we can solve the optimization problem in theory.

We additionally note here that Equation 3 is a *convex*

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

optimization problem. This is because, as with [4], the objective function and the constraint function are convex functions of the optimization variables $p_{Z|\boldsymbol{X}^*}$, p_Z and $p_{Z|Y}$. Convex optimization has several desirable properties. From an analytical viewpoint, it is assured that any local minimum is a global minimum and finding a global minimum is therefore reduced to finding a local minimum [2]. From a practical viewpoint, efficient algorithms such as interior-point methods have been proposed, and software libraries are available [3].

The issue that needs to be resolved is that $p_{\boldsymbol{X}^*,Y}$ must be known. However, with data collected by smart-meter systems, we may not directly know $p_{\boldsymbol{X}^*,Y}$ and therefore we cannot compute Equation 1 and Equation 2. In the following section, we present an assumed Gaussian model.

3.3 Gaussian Model Assumption

In this section, we present the assumed model for the distribution $p_{\boldsymbol{X}^*,Y}$ and show that the optimization problem (Equation 3) is solvable.

First, observe that from the law of total probability,

$$\begin{aligned} p_{\boldsymbol{X}^*,Y}(\boldsymbol{x}^*, y) &= \sum_{\bar{\boldsymbol{x}} \in \bar{\boldsymbol{\mathcal{X}}}} p_{\boldsymbol{X}^*,\bar{\boldsymbol{X}},Y}(\boldsymbol{x}^*, \bar{\boldsymbol{x}}, y) && (4)\\ &= \sum_{\bar{\boldsymbol{x}} \in \bar{\boldsymbol{\mathcal{X}}}} p_{\boldsymbol{X},Y}(\boldsymbol{x}, y) && (5)\\ &= \sum_{\bar{\boldsymbol{x}} \in \bar{\boldsymbol{\mathcal{X}}}} P_{\boldsymbol{X}}(\boldsymbol{x}) p_{Y|\boldsymbol{X}}(y|\boldsymbol{x}). && (6) \end{aligned}$$

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

3.3.1 Modeling and Parameters

In order to compute $p_{Y|\boldsymbol{X}}(y|\boldsymbol{x})$, we apply a linear Gaussian model. Let Y_0 be a random variable of the background noise and Y_m be that of the emission of appliance m. Then

$$\begin{aligned} Y &= Y_0 + \sum_{m=1}^{M} Y_m, && (7)\\ p_{Y_0} &\sim \mathcal{N}(\mu_0, \sigma_0^2), && (8)\\ p_{Y_m|X_m=x_{m,k}} &\sim \mathcal{N}(\mu_{m,k}, \sigma_{m,k}^2), && (9) \end{aligned}$$

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,

$$p_{Y|\boldsymbol{X}=\boldsymbol{x}} \sim \mathcal{N}\left(\mu_0 + \sum_{m=1}^{M} \mu_{m,k},\ \sigma_0^2 + \sum_{m=1}^{M} \sigma_{m,k}^2\right). \quad (10)$$

Equation 10 implies 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, Equation 10 can further be simplified as

$$p_{Y|\boldsymbol{X}=\boldsymbol{x}} \sim \mathcal{N}\left(\mu_0 + \sum_{m=1}^{M} \mu_{m,k},\ \sigma^2\right), \quad (11)$$

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 Section 4.

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

3.3.2 Solvability

Here we demonstrate that the optimization problem (Equation 3) is solvable, by showing that both objective function (Equation 1) and constraint function (Equation 2) are functions of the optimization variable $p_{Z|Y}$ and known distributions, given Θ' and $P_{\boldsymbol{X}}$.

We start with Equation 1. As $\boldsymbol{X} \to Y \to Z$ forms the Markov chain, $p_{Z|\boldsymbol{X}^*}(z|\boldsymbol{x}^*)$ in Equation 1 can be written as

$$p_{Z|\boldsymbol{X}^*}(z|\boldsymbol{x}^*) = \int_{\mathcal{Y}} p_{Z|Y}(z|y) p_{Y|\boldsymbol{X}^*}(y|\boldsymbol{x}^*) dy. \quad (12)$$

Here,

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

and assuming $p_{\bar{\boldsymbol{X}}|\boldsymbol{X}^*} = p_{\bar{\boldsymbol{X}}}$, i.e., the non-sensitive appliances behave independently from the sensitive appliances (or more broadly all the appliances behave independently from each other),

$$p_{Y|\boldsymbol{X}^*}(y|\boldsymbol{x}^*) = \sum_{\bar{\boldsymbol{x}} \in \bar{\boldsymbol{\mathcal{X}}}} p_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) p_{\bar{\boldsymbol{X}}}(\bar{\boldsymbol{x}}). \quad (13)$$

Substituting Equation 13 into Equation 12, we obtain

$$\begin{aligned} p_{Z|\boldsymbol{X}^*}(z|\boldsymbol{x}^*) &= \int_{\mathcal{Y}} p_{Z|Y}(z|y) \sum_{\bar{\boldsymbol{x}} \in \bar{\boldsymbol{\mathcal{X}}}} p_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) P_{\bar{\boldsymbol{X}}}(\bar{\boldsymbol{x}}) dy\\ &= \sum_{\bar{\boldsymbol{x}} \in \bar{\boldsymbol{\mathcal{X}}}} P_{\bar{\boldsymbol{X}}}(\bar{\boldsymbol{x}}) \int_{\mathcal{Y}} p_{Z|Y}(z|y) p_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) dy. \end{aligned} \quad (14)$$

Also, $p_Z(z)$ in Equation 1 can be written as

$$\begin{aligned} p_Z(z) &= \int_{\mathcal{Y}} p_{Z|Y}(z|y) p_Y(y) dy\\ &= \sum_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}}} \int_{\mathcal{Y}} p_{Z|Y}(z|y) p_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) P_{\boldsymbol{X}}(\boldsymbol{x}) dy\\ &= \sum_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}}} P_{\boldsymbol{X}}(\boldsymbol{x}) \int_{\mathcal{Y}} p_{Z|Y}(z|y) p_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) dy. \end{aligned} \quad (15)$$

Combining Equation 14 and Equation 15 with Equation 1, we can confirm that the objective function (Equation 1) is computable.

The constraint function (Equation 2) is also computable because $p_Y(y)$ in Equation 2 can be written as

$$p_Y(y) = \sum_{\boldsymbol{x} \in \boldsymbol{\mathcal{X}}} p_{Y|\boldsymbol{X}}(y|\boldsymbol{x}) P_{\boldsymbol{X}}(\boldsymbol{x}).$$

3.4 Modification to Discrete Power Data

Until now we have considered the case where the smart-meter data and distorted data are continuous. In practical situations, however, it is possible that the smart-meter data is quantized to discrete levels. Indeed, as we describe in detail in Section 4, we use discrete power data in our experiment that has been quantized to a resolution of 7 Watts. It is therefore required to modify the optimization problem (Equation 3) to accommodate such cases. We describe in this section the discretized version of the optimization problem.

Let $\tilde{Y} \in \tilde{\mathcal{Y}}$ be a discrete random variable representing the quantized smart-meter data and $\tilde{Z} \in \tilde{\mathcal{Z}}$ represent the distorted data, where $\tilde{\mathcal{Y}}$ and $\tilde{\mathcal{Z}}$ are finite sets. Let $d : \tilde{Y} \times \tilde{Z} \to \mathbb{R}^+$ be some distortion function. Then the optimization problem in Equation 3 becomes

$$\begin{aligned} &\min_{p_{\tilde{Z}|\tilde{Y}}} I(\boldsymbol{X}^*; \tilde{Z}) \\ &\text{subject to } E_{\tilde{Y},\tilde{Z}}[d(\tilde{Y}, \tilde{Z})] \leq \delta, \end{aligned} \tag{16}$$

where

$$\begin{aligned} &I(\boldsymbol{X}^*; \tilde{Z}) \\ &= \sum_{\boldsymbol{x}^* \in \boldsymbol{\mathcal{X}}^*} \sum_{\tilde{z} \in \tilde{\mathcal{Z}}} P_{\boldsymbol{\mathcal{X}}}(\boldsymbol{x}^*) P_{\tilde{Z}|\boldsymbol{X}^*}(\tilde{z}|\boldsymbol{x}^*) \log \frac{P_{\tilde{Z}|\boldsymbol{X}^*}(\tilde{z}|\boldsymbol{x}^*)}{P_{\tilde{Z}}(\tilde{z})} \end{aligned} \tag{17}$$

and

$$E_{\tilde{Y},\tilde{Z}}[d(\tilde{Y}, \tilde{Z})] = \sum_{\tilde{y} \in \tilde{\mathcal{Y}}} \sum_{\tilde{z} \in \tilde{\mathcal{Z}}} P_{\tilde{Z}|\tilde{Y}}(\tilde{z}|\tilde{y}) P_{\tilde{Y}}(\tilde{y}) d(\tilde{y}, \tilde{z}). \tag{18}$$

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 Section 4.1, and we show in Section 4.2 the electric power meter and the home appliances 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 Section 4.4. Section 4.5 evaluates the privacy and utility aspects of our mechanism quantitatively, and the implications of the results are discussed in Section 4.6.

4.1 Overview

Our goal is to examine whether the theory we propose in Section 3 is effective in an actual situation (i.e. in natural daily life), not in an artificial environment or in a special circumstance. To this end, we collected the power usage data of an actual household using a commercially available power meter device for nine days. As our proposed theory requires estimation of Θ' and $P_{\boldsymbol{X}}$ for the assumed model distribution, we also manually collected the ground truth of the appliance usage in the household for the same nine days, and then applied a supervised learning algorithm to estimate those parameters.

Table 1 Appliances used in the target household and the 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 + refrigerator	103.44		
1	bathroom light	12.73	0.000404	0.0219
2	dryer	380.02	0.000159	0.667
3	electric heater	350.37	0.000161	0.0148
4	entrance light	90.90	0.000318	0.222
5	kitchen light	175.63	0.00257	0.0321
6	kotatsu	168.77	0.000652	0.0246
7	laundry machine	57.00	0.00408	0.0142
8	lavatory light	51.95	0.00408	0.505
9	living room light	84.69	0.00194	0.00544
10	microwave	1115.80	0.000159	0.333
11	oven toaster	1133.40	0.000957	0.245
12	personal computer	111.85	0.000663	0.0152
13	reading room light	72.24	0.00219	0.0366
14	rice steamer	323.62	0.000322	0.0230
15	television	123.16	0.00366	0.00335
16	vacuum cleaner	1057.90	0.000159	0.182
17	washstand light	34.89	0.000637	0.216
				$\sigma^2 = 5436.5$

Then, we considered two use cases: 1) oven toaster is designated as sensitive; and 2) television 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 (Equation 16) and obtained a privacy mapping $P_{\tilde{Z}|\tilde{Y}}$. We then distorted the power usage data according to $P_{\tilde{Z}|\tilde{Y}}$, and obtained distorted power usage data. In order to evaluate the privacy and utility of our mechanism in line with the goal given in Section 3.1, 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.

4.2 Devices

The electric power meter we used is the *OWL +USB*[*2] which records the electric power used in a household every minute. This power meter is attached to the circuit-breaker of the target household, and the total power usage of the household is recorded. Due to limitation of the A/D converter used in the power meter, the resolution of the power recorded is 7 Watts.

The appliances in the target household are listed in Table 1. As Table 1 shows, a total of 17 appliances are present[*3].

4.3 Datasets and Parameters

4.3.1 Power Usage and Appliance Usage Datasets

We collected the power usage data for nine days. Samples

[*2] `http://www.theowl.com/index.php/energy-monitors/standalone-monitors/owl-usb/`

[*3] Strictly speaking, the number of appliances used in the household is 18 because a refrigerator is also used. However, it was always ON throughout the data collection and therefore we regarded it as a part of the background noise.

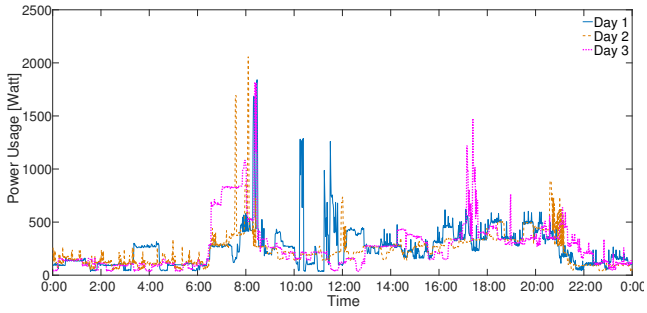

Fig. 1 Samples of the power usage data

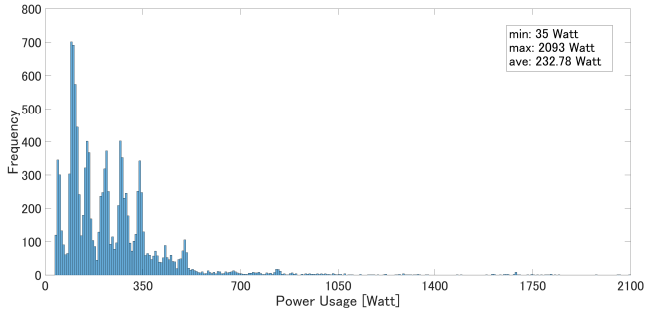

Fig. 2 Histogram of the power usage data

Table 2 An excerpt from the ground truth of the appliance usage

Day	Time	m	Appliance	Operation
1	3:20	12	personal computer	ON
	3:20	13	reading room light	ON
	4:16	8	lavatory light	ON
	4:17	8	lavatory light	OFF
	⋮	⋮	⋮	⋮
9	4:33	13	reading room light	ON
	⋮	⋮	⋮	⋮
	23:18	9	living room light	OFF
	23:20	15	television	OFF

are shown in **Fig. 1** and the histogram is shown in **Fig. 2**. The minimum power is 35 Watt, the maximum power is 2093 Watt and the average power is 232.78 Watt.

We also collected manually the ground truth of the appliance usage in the target household. **Table 2** shows an excerpt from the ground truth.

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.

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) [9]. FHMM is novel in that it assumes the hidden states are factorized into multiple independent states and thereby 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 $m \in \{1, 2, \ldots, M = 17\}$. 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 transition 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,

$$P_{X_m}(\mathsf{ON}) = \frac{a_m}{a_m + b_m}, \quad P_{X_m}(\mathsf{OFF}) = \frac{b_m}{a_m + b_m}. \quad (19)$$

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 nine day data of power usage and appliance usage for the supervised learning, and obtained $\Theta' = \{\mu_0, \{\mu_{m,\mathsf{ON}}\}, \sigma^2\}$ and $\{a_m, b_m\}$. The results are shown in Table 1.

4.4 Optimization and Distortion

As we explained in Section 4.1, we considered the following two use cases:

Case 1 Oven toaster ($m = 11$) is designated as sensitive,

Case 2 Television ($m = 15$) is designated as sensitive.

For each case, we solved the convex optimization problem and obtained a privacy mapping $P_{\tilde{Z}|\tilde{Y}}$. Then we distorted the raw power data according to each $P_{\tilde{Z}|\tilde{Y}}$.

Let $\tilde{Y}$ and $\tilde{Z}$ be the discrete random variables defined in Section 3.4, and let $\tilde{\mathcal{Y}}$ and $\tilde{\mathcal{Z}}$ be a finite set of possible values of $\tilde{Y}$ and $\tilde{Z}$, respectively. The size of these sets affects computational complexity of the optimization problem; concretely, the computational complexity is $O(|\tilde{\mathcal{Y}}||\tilde{\mathcal{Z}}|)$. We defined $\tilde{\mathcal{Y}} = \tilde{\mathcal{Z}} = \{0, 7, 14, 21, \ldots, 2093\}$. The distortion metric we used is the L_1 distance $d(\tilde{y}, \tilde{z}) = |\tilde{y} - \tilde{z}|$. Then, for each of the two use cases, we solved the optimization problem using the convex optimization software CVX[*4] and obtained a privacy mapping $P_{\tilde{Z}|\tilde{Y}}$. We used $\delta = 6$ for Case 1 and $\delta = 72$ for Case 2. A graphical representation of $P_{\tilde{Z}|\tilde{Y}}$ is shown in **Fig. 3** and **Fig. 4**.

Then we distorted the power usage data according to $P_{\tilde{Z}|\tilde{Y}}(\tilde{z}|\tilde{y})$. A sample of the raw and distorted power usage data is shown in **Fig. 5** and **Fig. 6**.

4.5 Evaluation of Privacy and Utility

This section evaluates both the privacy and utility aspects of the distorted power usage data. Since our goal of the privacy-utility tradeoff is to retain the usage analysis of the non-sensitive appliance while preventing that of the sensitive appliances, we evaluate them by measuring how the appliance usage analysis degrades. We therefore apply an inference algorithm to the raw data and the distorted data (both Case 1 and 2) to infer the hidden states of the appliances, and evaluate the detection rates.

[*4] `http://cvxr.com/cvx/`

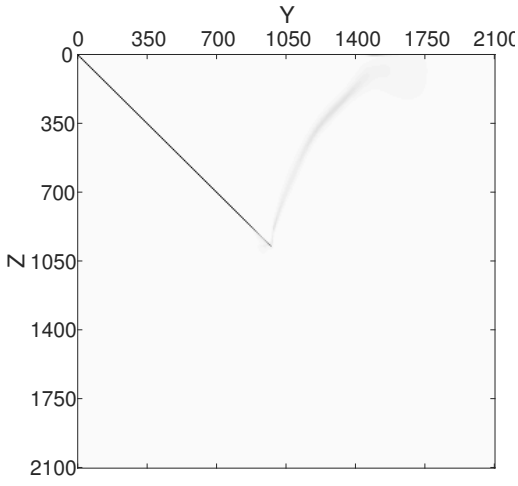

Fig. 3 A graphic representation of the privacy mapping $P_{\tilde{Z}|\tilde{Y}}$ (sensitive appliance is oven toaster and $\delta = 6$). Pure white represents 0, pure black represents 1, and values in between 0 and 1 are represented by shades of grey. This picture therefore tells that $P_{\tilde{Z}|\tilde{Y}}$ is almost identical to an identity matrix when $0 \leq Y \leq 910$ but becomes weird when $Y > 910$.

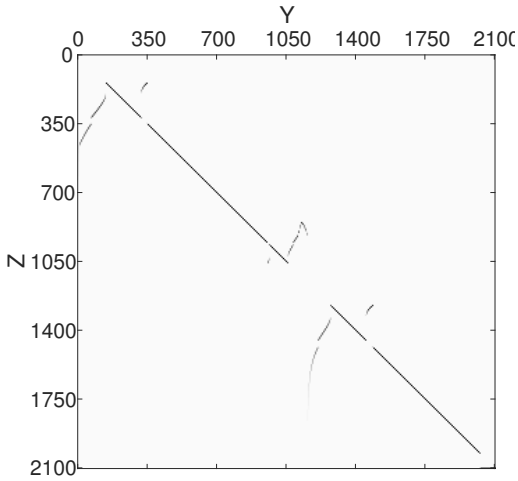

Fig. 4 A graphic representation of the privacy mapping $P_{\tilde{Z}|\tilde{Y}}$ (sensitive appliance is television and $\delta = 72$)

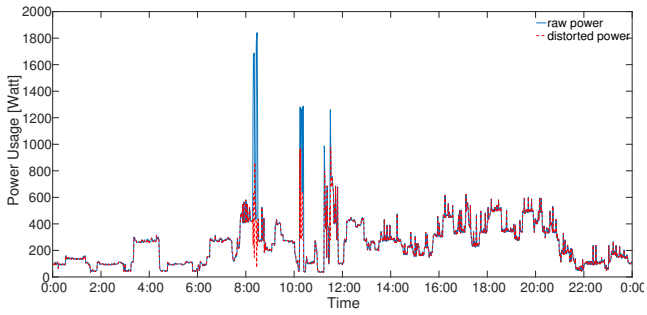

Fig. 5 A sample of the raw and distorted power usage data (sensitive appliance is oven toaster and $\delta = 6$)

We again model the hidden states with FHMM accompanied by the parameters we obtained in the supervised learning, and infer the hidden states using an approximate inference algorithm called the completely factorized variational approximation (CFVA) [9]. 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.

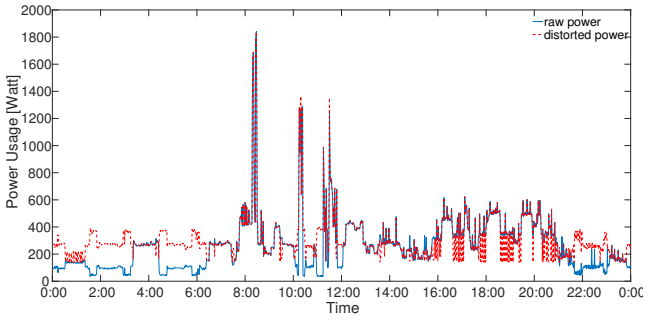

Fig. 6 A sample of the raw and distorted power usage data (sensitive appliance is television and $\delta = 72$)

Table 3 AUC values of the ROC curves (Fig. 7, Fig. 8 and Fig. 9)

m	Appliance	AUC		
		raw (Fig. 7)	oven toaster (Fig. 8)	television (Fig. 9)
3	electric heater	0.904	0.902	0.889
11	oven toaster	0.969	*0.551*	0.969
12	personal computer	0.787	0.783	0.648
15	television	0.914	0.913	*0.459*

Fig. 7 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 3**. As the AUC values tell, the states of the oven toaster are inferred almost correctly, the states of the electric heater and television are inferred with high accuracy, and the states of the personal computer are inferred with marginal accuracy.

Fig. 8 gives ROC curves of the inference results with the distorted data for Case 1. The inference performance for the oven toaster is degraded severely while the inference performance for the other appliances are preserved, as desired.

Fig. 9 gives ROC curves of the inference results with the distorted data for Case 2. The inference performance for the television is degraded severely. The inference performance for the oven toaster and electric heater are preserved almost completely. The inference performance for the personal computer is degraded to some extent, but still enables meaningful inference.

4.6 Discussion

As we have shown in Section 4.5, the distortion works highly effectively for the case where the sensitive appliance is oven toaster. This may be due to the fact that the oven toaster is realistically modeled with only two states: {ON, OFF}, and therefore our simplified model fit well. Moreover, the consumed power is as high as 1kW, which enables us to compute an optimal privacy mapping $P_{\tilde{Z}|\tilde{Y}}$ that at-

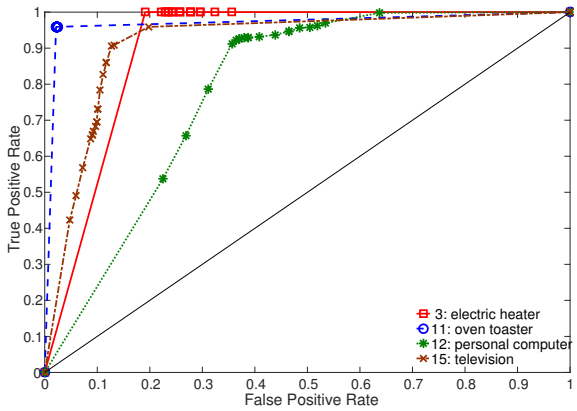

Fig. 7 ROC curves of the results of inference with raw data

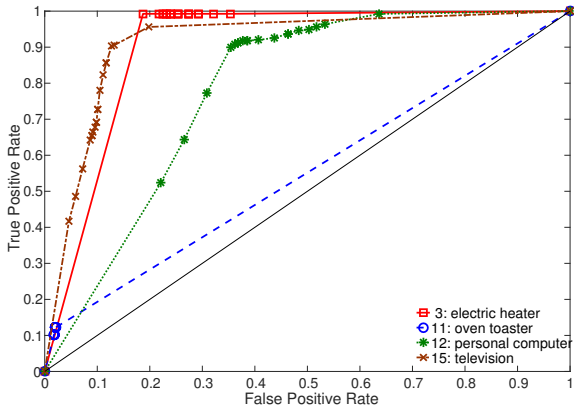

Fig. 8 ROC curves of the results of inference with distorted data (sensitive appliance is oven toaster and $\delta = 6$)

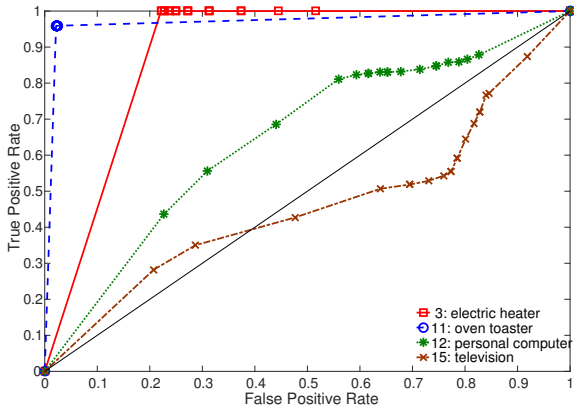

Fig. 9 ROC curves of the results of inference with distorted data (sensitive appliance is television and $\delta = 72$)

tains both small mutual information and small distortion such as $\delta = 6$. Note that 1kW or higher power consumption occurs rarely, as Fig. 2 shows, and thus the distortion of higher power values does not affect other low-power appliances.

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 personal computer degraded to some extent. This stems from the fact that the television consumes a relatively low power of 123W and thus distortion of middle power values would affect other middle-power appliances including the personal computer.

We should note that we assumed a Gaussian FHMM and used the CFVA approximation for inference. Other inference algorithms such as neural networks may bring other conclusion [10].

5. Conclusion

We proposed in this paper a privacy-utility tradeoff mechanism which accommodates the situation where sensitive appliance usage is not observable. We formalized the tradeoff as a convex optimization problem that we show can be solved. We then exhibited experimental results on smart-meter data and showed that the proposed mechanism is practical.

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

Acknowledgments We would like to thank all the anonymous volunteers including Chiyo, Kaori and Nagisa for their dedicated support for our experiment.

References

[1] Batina, L., Gierlichs, B., Prouff, E., Rivain, M., Standaert, F.-X. and Veyrat-Charvillon, N.: Mutual information analysis: A comprehensive study, *Journal of Cryptology*, Vol. 24, No. 2, pp. 269–291 (2010).

[2] Boyd, S. and Vandenberghe, L.: *Convex optimization*, Cambridge University Press (2004).

[3] Boyd, S., Vandenberghe, L. and Grant, M.: Advances in convex optimization, *CCC 2006*, IEEE (2006).

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

[5] Duhigg, C.: How companies learn your secrets, *The New York Times Magazine*, (online), available from ⟨http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html⟩ (2012).

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

[7] Erdogdu, M. A. and Fawaz, N.: Privacy-utility trade-off under continual observation, *ISIT 2015*, IEEE, pp. 1801–1805 (2015).

[8] 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).

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

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

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

[12] Machanavajjhala, A., Kifer, D., Gehrke, J. and Venkitasubramaniam, M.: L-diversity: Privacy beyond k-anonymity, *ACM TKDD*, Vol. 1, No. 1 (2007).

[13] 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).

[14] 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 JSTSP*, Vol. 9, No. 7, pp. 1240–1255 (2015).

[15] Samarati, P.: Protecting respondents identities in microdata release, *IEEE TKDE*, Vol. 13, No. 6, pp. 1010–1027 (2001).

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

[17] Sweeney, L.: k-anonymity: A model for protecting privacy, *IJUFKS*, Vol. 10, No. 05, pp. 557–570 (2002).