

Private Disclosure of Information in Health Tele-monitoring

Daniel Aranki and Ruzena Bajcsy

Electrical Engineering and Computer Science Department,
University of California, Berkeley,
Berkeley, CA 94720 USA
{daranki, bajcsy}@eecs.berkeley.edu

Abstract

We present a novel framework, called Private Disclosure of Information (PDI), which is aimed to prevent an adversary from inferring certain sensitive information about subjects using the data that they disclosed during communication with an intended recipient. We show cases where it is possible to achieve perfect privacy regardless of the adversary's auxiliary knowledge while preserving full utility of the information to the intended recipient and provide sufficient conditions for such cases. We also demonstrate the applicability of PDI on a real-world data set that simulates a health tele-monitoring scenario.

1 Introduction

Data collection and sharing is growing to unprecedented volumes. Some of the reasons for this phenomenon include the decrease in storage cost, the rise of social networks, the ubiquity of smartphones and law regulations. For example, in many states in the US, medical institutions are obliged to make demographics data public about their patients (NAHDO, 1996; Sweeney, 2002; OSHPD, 2014).

Warner (1965) argues that the lack of privacy guarantees can cause subjects to be reluctant to share their data with data collectors (such as doctors, government agencies, researchers, etc.) or even result in subjects providing false information. Therefore, subjects need to be assured that their privacy will be preserved throughout the whole process of data collection and use.

One of the emerging areas with growing interest to collect sensitive personal and private data is health tele-monitoring. In this setting, a technology is used to collect health-related data about patients, which are later submitted to a medical staff for monitoring. The data are then used to assess the health status of patients and provide them with feedback and/or intervention. Research indicates that such technologies can improve readmission rates and lower overall costs (Clark et al., 2007; Chaudhry et al., 2010; Inglis, 2010; Giamouzis et al., 2012;

Aranki et al., 2014). In such scenarios, the collected data are usually of sensitive nature from a privacy point of view and therefore privacy preserving technologies are needed in order to protect patients' privacy and increase compliance.

There are multiple stages in the life-cycle of data, including *i)* the disclosure (or submission) of the data by the subjects to the data collector; *ii)* the processing of the data; *iii)* the analysis; and/or *iv)* the publishing of (often a privatized version of) the data or some findings based on them. In this paper we focus on the phase of disclosure of privacy-sensitive data by the data owners. Our framework for Private Disclosure of Information (PDI) is thus aimed to prevent an adversary from inferring certain sensitive information about the subject using the data that were disclosed during communication with an intended recipient. This is analogous to the problem of attribute linkage in statistical database privacy.

In traditional encryption approaches to maintaining privacy, it is often implicitly assumed that the data themselves *are* the private information. However, in more general scenarios, the data *can be used to infer some private information* about the subjects for which the data apply. For example, respiration rate by itself might not be considered private information. However, if the data from the collected respiration rate are used to infer whether the individual is a smoker or not, they become sensitive information. One can argue that because the information about whether someone smokes is private, the respiration rate data become private *by implication*.

Under such circumstances, one should attempt to privatize the transmitted data in a way that reveals as little as possible about the private information to an adversary. In summary, our objective is to encode the transmitted data in order to hide another private piece of information. In the words of Sweeney (2002): "Computer security is not privacy protection." The converse is also true, privacy does not replace security. Our approach is therefore to be viewed as complementary to classical security approaches. For example, data can be privatized then encrypted.

The rest of this paper is organized as follows. In Section 2 we provide a survey of the literature for related work. In Sec-

tion [3] we provide the motivation to the problem and formulate it, followed by further analysis in Section [4]. We then discuss implementation details of the learning problem in Section [5] followed by experimental results in Section [6]. Finally, we close by discussing our conclusions and future research directions in Section [7].

2 Related Work

The study of privacy-preserving techniques and technologies in the fields of statistics, computer security and databases, and their intersections, dates back to at least [1965] when [Warner] proposed a randomization technique for conducting surveys and collecting responses for the purpose of statistical and population analysis. Since then, extensive privacy research in the fields above was conducted. Therefore, in the interest of brevity, we provide a brief overview of the areas of study related to our work and refer the reader to more comprehensive surveys in each area.

Recently, attention to privacy has been rising in the health-care domain with the spread of electronic health-records usage and the growing data sharing between medical institutions. It has been reported that consumers are expressing increasing concerns regarding their health privacy [Bishop et al. 2005, Hsiao and Hing 2012]. Most of the research in privacy from the health community focuses on medical data publishing and is therefore database-centric. For a survey of results in this domain, we refer the reader to [Gkoulalas-Divanis et al. 2014].

In more general-purpose scenarios, the privacy of statistical databases and data publishing has been extensively studied. [Denning and Schlorer 1983] presented some of the early threats related to inference in statistical databases and reviewed controls that are based on the lattice model [Denning 1976]. [Duncan and Lambert 1989, 1986] studied methods for limiting disclosure and linkage risks in data publishing. [Sandhu 1993] provided a tutorial on lattice-based access controls for information flow security and privacy. Later, [Farkas and Jajodia 2002] provided a survey of more results in the field of access controls to the inference problem in database security. For rigorous surveys in the fields of data publishing privacy and statistical databases privacy, we refer the reader to [Adam and Worthmann 1989, Fung et al. 2010].

Two semantic models of database privacy of growing interest in the privacy literature are k -anonymity [Sweeney 2002] and differential privacy [Dwork 2006, 2008]. In k -anonymity, given a set of quasi-identifiers that can be used to re-identify subjects, a table is called k -anonymous if every combination of quasi-identifiers in the table appears in at least k records. If a table is k -anonymous, assuming each individual has a single record in the table, then the probability of linking a record to an individual is at most $1/k$. Other extensions and refinements of k -anonymity have been proposed including l -diversity [Machanavajjhala et al. 2007], t -closeness [Li et al.

[2007] and others.

In differential privacy, the requirement is that the output of a statistical query should not be too sensitive to any single record in the database. Formally, given a statistical query M , then M is ϵ -differentially private if $\mathbb{P}(M(D_1) \in S) \leq e^\epsilon \times \mathbb{P}(M(D_2) \in S)$ for any two realizations D_1 and D_2 of the database such that $|D_1 \Delta D_2| = 1$ and all $S \subset \text{Range}(M)$, where $D_1 \Delta D_2$ is the symmetric difference between D_1 and D_2 [Dwork 2006, 2008]. [Cormode 2011] showed that sensitive attribute inference can be done on databases that are differentially private and l -diverse with similar accuracy.

As can be seen from the review above, most of the research in data-privacy is focused on privacy-preserving data publishing and privacy-preserving statistical databases. In contrast, in this work we focus on preventing adversarial statistical inference of a piece of private information based on the disclosed messages in an individual's information exchange scenario during communication.

3 Problem Formulation

3.1 Notation

We use the following shorthand notation for probability density (mass) functions. We always use a pair of a capital and a small symbols of the same letter for a random variable and a realization of it, respectively. For notation simplicity and conciseness, given random variables X and Y , instead of writing $p_X(x)$ for the marginal density (mass) function of X we simply write $p(x)$, and instead of writing $p_{X|Y}(x|y)$ for the conditional density (mass) function of X given Y , we simply write $p(x|y)$.

3.2 Motivation and Threat Model → focus

We are primarily motivated by the tele-monitoring setting. In this setting, a doctor wishes to monitor her patients remotely using a technology that can collect and transmit health-related data. The shared data are of sensitive nature because they can be used to infer private pieces of information like a health-condition or a disease. For example, updates about a patient's weight can lead to disclosure of obesity as it will be demonstrated in Section [6].

More generally, an information provider Bob wants to disclose a piece of information x to some recipient Alice. Furthermore, the information x can be used to infer some private information c about Bob. However, there is no guarantee that the transmitted information will not be intercepted and potentially used for inference of the private information c about Bob by an untrusted but passive eavesdropper Eve. Finally, in this setting, we assume that Alice is more certain about c than Eve is. The problem at hand is delivering the information x under these circumstances such that Alice can make full use of the

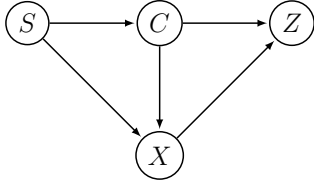


Figure 1: The Graphical Model of PDI

information but that Eve's ability to infer c about Bob, using the transmitted message, is minimized.

As a concrete example, consider the following scenario in health tele-monitoring. A patient Bob is trying to update his physician Alice about his weight and body mass index (BMI).¹ Since Alice is Bob's physician, she already knows the weight status category of Bob which he considers to be private information.² Eve, however, does not know Bob's weight status category *a priori* but would like to learn it from the messages he sends to Alice. If Eve succeeds to listen in on the communication between Bob and Alice, Eve can, with some accuracy, infer the weight status category of Bob. Alice, being a considerate physician, wants to ensure the privacy of her patients. Alice decides to create an encoding scheme (that can be made public) for the communication such that the encoding is different per weight status group. Her objective is to make this encoding scheme "as privacy-preserving as possible" in the sense of keeping her patients' weight status category information as private as possible to someone who does not know it *a priori*.

It is important to compare this scenario with the classical security approach. In classical security, the objective is to protect the transmitted message itself without taking into consideration an adversarial effort to statistically infer private information using the cipher-text. It has been demonstrated that statistical inference can still be performed on encrypted data (For example [White et al., 2011, Miller et al., 2014]). We complement this by capturing the notion of statistical inference of the private information c from the transmitted data, and aim to find a way to minimize the ability of an adversary to infer c using the transmitted data.

3.3 Problem Definition

Towards a more formal representation of the problem, we consider scenarios where *i)* Bob's identity, s , is attached to any message that is sent by him; *ii)* there is no guarantee that the sent information will not be intercepted by an untrusted but passive eavesdropper Eve; *iii)* the information x can be used to infer some private information c about Bob; and *iv)* Alice knows the private information c about Bob but Eve does not.

¹BMI is a measure of relative weight based on an individual's mass and height. Defined as $BMI \triangleq \frac{mass(kg)}{height(m)^2}$.

²Weight status category indicates if an individual is underweight, overweight, obese or has a healthy weight.

$$z = (s, x)_{\text{package}}$$

Under these assumptions, Bob would like to exploit the fact that Alice knows c but Eve does not in order to send a message z that is more useful to Alice than Eve. The utility value of the message follows the following decoding and "hiding class" (HC) premises:

DECODING Alice can make full use of the sent information z , i.e. obtain the original message x from the transmitted message z ; and

HC Eve's ability to make inference about c given s , based on the sent information z is minimized.

Formally, we use \mathcal{S} for the set of identifiers of information providers, \mathcal{I} for the information space and Σ for the set of private classes (the private information about the information providers). Similarly, we define the random variables S for the identifier of the information provider, X for the piece of information that the provider would like to disclose, C for the class that the provider belongs to and Z for the encoded message that will be sent (called *privatized information*), which is a function of the original information and the class. We call this function a *privacy mapping function* and define it as $R : \Sigma \rightarrow \mathcal{I}^{\mathcal{I}}$ where $\mathcal{I}^{\mathcal{I}}$ is the set of injective functions $\mathcal{I} \rightarrow \mathcal{I}$. A simple way to think about R is as an encoding scheme. That is, for every class $c \in \Sigma$, it outputs an encoding function for the input information x . Given $c \in \Sigma$, since $R(c)$ is injective, then there exists a left inverse $R^l(c)$ which will be used to decode the messages z sent from subjects in class c .³ From that, Z is simply equal to $[R(C)](X)$. The statistical model that relates these random variables is described in Figure 1.

For conciseness, in this paper we treat the case of continuous information spaces. Note that in the case of a discrete information space, the reader is instructed to follow the discussion by substituting probability density functions with probability mass functions for the distributions of X and Z . Note that our treatment also covers the case of information spaces of mixed nature (that are discrete in some attributes and continuous in others) by using the appropriate probability distribution functions.

For the model in Figure 1 one needs to supply the following probability distributions. $p(s)$, the prior of subjects transmitting messages in the system. $p(c|s)$, the adversary's prior of class membership for the different subjects (based on auxiliary knowledge). $p(x|c, s)$, the generative model of data given a class and a subject. Finally, $p(z|x, c)$ is simple and can be modeled as $\mathbb{P}(Z = z|X = x, C = c) = 1$ if and only if $z = [R(c)](x)$ and 0 otherwise, for all $z, x \in \mathcal{I}$ and $c \in \Sigma$.

Recall that the identity s of the information provider is attached with the transmitted message. Moreover, the intended recipient knows the class c of the information provider. Therefore, because of the injectivity requirement of the privacy mapping function, the intended recipient can decode the sent information.

³We say that $g : D_2 \rightarrow D_1$ is a left inverse of a function $f : D_1 \rightarrow D_2$ if for all $x \in D_1$ we have $g(f(x)) = x$.

$$z = (s, x)_{\text{package}}$$

$$x \rightarrow c_{\text{infer}}$$

$$z \rightarrow x \rightarrow c_{\text{injectivity}}$$

mation z back to the original message x . Hence the requirement **(DECODING)** is satisfied.

Finally, in order to satisfy the second requirement **(HC)** we would like to find a privacy mapping function R that minimizes the amount of information that the privatized information Z carries for the sake of inferring the private class C , given the subject identifier S , to an adversary. We adopt the measure of (conditional) mutual information to model this quantity. We present the definition of conditional mutual information for continuous random variables, and refer the reader to [Cover and Thomas, 2006, Definitions 2.61 and 8.54] for the corresponding definitions concerning discrete random variables and random variables that can be mixtures of discrete and continuous, respectively.

Definition 1 [Cover and Thomas, 2006, c.f. Definition 8.49]. Let X, Y and Z be random variables. The conditional mutual information of X and Y given Z , $I(X, Y|Z)$, is defined as

$$I(X, Y|Z) \triangleq E_{p(x, y, z)} \left[\log \frac{p(x, y|z)}{p(x|z)p(y|z)} \right]$$

Intuitively, $I(Z, C|S; R)$ measures in bits, the expected amount of mutual information that the random variables $Z = [R(C)](X)$ and C have, given the information in S .⁴ Mutual information also provides a sufficient and necessary condition for conditional independence as follows.

Lemma 1 [Cover and Thomas, 2006, c.f. Corollary 2.92; c.f. Theorem 8.6.1]. $I(Z, C|S; R) \geq 0$ for any privacy mapping function R . Furthermore, $I(Z, C|S; R) = 0$ if and only if Z and C are conditionally independent given S using the privacy mapping function R .

From the intuition above, and the fact in Lemma 1 we set our objective to find a privacy mapping function R that minimizes the conditional mutual information of the privatized information Z and the private class C given the identity of the information provider S such that the model in Figure 1 holds. In short,

$$R^* = \arg \min_R I(Z, C|S; R) \quad (1)$$

subject to R is a privacy mapping function
and Model in Figure 1

Once a privacy mapping function R is chosen, the communication process can be carried as follows.

Sending The transaction of disclosing a piece of information $x \in \mathcal{I}$ by an information provider belonging to class $c \in \Sigma$ is performed by applying the following transformation $z \leftarrow [R(c)](x)$ and sending z (or some encrypted version of it).

Receiving The transaction of receiving a piece of information $z \in \mathcal{I}$ sent by an information provider belonging to class

$c \in \Sigma$ is performed by applying $x \leftarrow [R^l(c)](z)$. Where $R^l(c)$ is a left inverse of $R(c)$.

Note that the problem in Equation (1) is not a convex problem. Furthermore, it is of interest to study how to learn the model in Figure 1 and find an optimal privacy mapping function R from data. We will address this question in Section 5 but first we further study the properties of the formulated framework in the following section.

4 Further Analysis

First, we relate the value of the objective function in Equation (1) to Bayesian inference in the following lemma.

Lemma 2. If a privacy mapping function R yields $I(Z, C|S; R) = 0$ then Bayesian inference of C based on Z is prevented for the adversary.

Proof. From Lemma 1 we know that Z is conditionally independent of C given S which means $p(c|z, s) = p(c|s)$ which is the prior of the class membership that the adversary already possesses. Therefore, the disclosure of Z does not change the adversary's belief regarding the private information C given the subject identifier S . \square

The next question that we need to ask is whether a privacy mapping function R satisfying $I(Z, C|S; R) = 0$ is ever attainable. There are three reasons for this question. First, if such a privacy mapping function R exists, then it means that by knowing S (which is always attached to the message), Z provides no extra information to inferring C to an adversary, which sounds surprising. Second, there is generally a trade-off between information utility and privacy where optimal privacy is usually only attained at the cost of no utility [Dwork, 2006]. In our case, the utility of the information Z to the intended recipient is always fully preserved, unrelated of the choice of R , since $R(c)$ is injective for all $c \in \Sigma$. From this it follows that the scenario of perfect privacy seems to be unattainable.⁵ Finally, if such a privacy mapping function R exists, it would assure optimality of Equation (1). Fortunately (and somewhat unintuitively), such a mapping function can be attained as shown in the following sequence of results.

Lemma 3. If there exists a function $f(z, s)$ such that $p(z|c, s) = f(z, s)$ for all $c \in \Sigma, z \in \mathcal{I}$ and $s \in \mathcal{S}$ then $p(z|s) \equiv f(z, s)$

Proof. $p(z|s) = \sum_{c \in \Sigma} p(z, c|s) = \sum_{c \in \Sigma} p(z|s, c) \cdot p(c|s) = \sum_{c \in \Sigma} f(z, s) \cdot p(c|s) = f(z, s) \cdot \sum_{c \in \Sigma} p(c|s) = f(z, s)$ \square

Using Lemma 3 we prove the following theorem, which is a sufficient condition for optimality of Equation (1).

⁵We consider "perfect privacy" to be that the adversary's belief about C given S doesn't change after observing Z .

⁴The units are bits assuming the log base in Definition 1 is 2.

Theorem 1. If there exists a function $f(z, s)$ such that $p(z|c, s) = f(z, s)$ for all $c \in \Sigma, z \in \mathcal{I}$ and $s \in \mathcal{S}$ then $D_{KL}(p(c|z, s)||p(c|s)) = 0$ for all $z \in \mathcal{I}$ and $s \in \mathcal{S}$ ⁶.

Proof. Since $p(z|c, s) = f(z, s)$ then using Lemma 3 we know that $p(z|s) \equiv f(z, s)$. Therefore, for any $z \in \mathcal{I}$ and $s \in \mathcal{S}$ such that $f(z, s) = p(z|c, s) = p(z|s) \neq 0$ we get $\frac{p(c|z, s)}{p(c|s)} = \frac{p(z|c, s) \cdot p(c|s)}{p(c|s) \cdot p(z|s)} = \frac{p(z|c, s)}{p(z|s)} = 1$. This implies $D_{KL}(p(c|z, s)||p(c|s)) = 0$. \square

Corollary 1. If a privacy mapping function R achieves $p(z|c, s) = f(z, s)$ for some function $f(z, s)$, for all $c \in \Sigma, z \in \mathcal{I}$ and $s \in \mathcal{S}$ then R is the optimal solution to Equation (1).

Proof. The result follows from Theorem 1 and the fact that $I(Z, C|S; R) = E_{p(z, s)}[D_{KL}(p(c|z, s; R)||p(c|s; R))]$. \square

Note that Theorem 1 is independent of the model of $p(c|s)$ (and $p(s)$). This is a very important observation since it means that in cases where a privacy mapping function R satisfies the condition of the theorem, modeling the adversary's prior knowledge about information providers' class memberships is not needed. Furthermore, such privacy mapping function achieves perfect privacy against any adversary, regardless of her auxiliary knowledge $p(c|s)$ (or $p(s)$). In the following theorems we provide examples of using Theorem 1 that also serve as cases where such privacy mapping functions are attainable.

Theorem 2. If $X|C = c, S = s \sim N(\mu_c, \Sigma_c)$ (Normal distribution) for every $c \in \Sigma$ and $s \in \mathcal{S}$, then $[R(c)](x) = \Sigma_c^{-\frac{1}{2}} \cdot (x - \mu_c)$ is an optimal solution to Equation (1).

Proof. It is easy to verify that $Z|C = c, S = s \sim N(\bar{0}, I)$ for every $s \in \mathcal{S}$ and $c \in \Sigma$, where $\bar{0}$ is the origin in the information space (vector of zeros) and I is the identity matrix (of the appropriate dimensions). This means that $p(z|c, s) \equiv f(z, s)$ (not a function of c). By using Theorem 1 we therefore know that R is the optimal solution to Equation (1). \square

The proofs of the following theorems are similar to this of Theorem 2 and were thus omitted for conciseness.

Theorem 3. If $X|C = c, S = s \sim \text{Exp}(\lambda_c)$ (Exponential distribution) for every $c \in \Sigma$ and $s \in \mathcal{S}$, then $[R(c)](x) = \lambda_c x$ is an optimal solution to Equation (1).

Theorem 4. If $X|C = c, S = s \sim \text{Gamma}(k, \theta_c)$ (Gamma distribution with shape and scale parameters) for every $c \in \Sigma$ and $s \in \mathcal{S}$, then $[R(c)](x) = \frac{x}{\theta_c}$ is an optimal solution to Equation (1).

Theorem 5. If $X|C = c, S = s \sim U(a_c, b_c)$ (Continuous Uniform distribution) for every $c \in \Sigma$ and $s \in \mathcal{S}$, then $[R(c)](x) = \frac{x - a_c}{b_c - a_c}$ is an optimal solution to Equation (1).

⁶(Cover and Thomas [2006] Definition 8.46): The Kullback-Leibler divergence is defined as $D_{KL}(p||q) = E_p \left[\log \frac{p}{q} \right]$.

5 Implementation

In this section, we briefly describe an implementation of the learning problem that is publicly available in the form of a MATLAB⁷ toolbox (Aranki and Bajcsy [2015]). In this implementation, we investigate the question of learning a privacy mapping function R from a labeled data set $\mathcal{D} = \{(x_i, c_i)_i\}$. This implies a simplifying assumption of ignoring the modeling of the random variable S corresponding to the identity of the information providers. This assumption has the following implications on the model in Figure 1. First, it implies that the adversary views information providers as uniformly distributed, that is $p(s) = \frac{1}{|\mathcal{S}|}$ for all $s \in \mathcal{S}$. Second, the assumption implies that the subject-class membership belief function of the adversary is equal for all subjects, that is $p(c|s) = p(c)$ for all $s \in \mathcal{S}$ and $c \in \Sigma$. As discussed in Section 4 in the cases where perfect privacy is achievable, the solutions are independent of these models and therefore these implications are not limiting. Further study is necessary to assess the level of privacy-degradation incurred by this assumption in cases of imperfect privacy. Third, this assumption implies that the generative model of data per class is independent of the subjects, that is $p(x|c, s) = p(x|c)$ for all $x \in \mathcal{I}, c \in \Sigma$ and $s \in \mathcal{S}$. Finally, $I(Z, C|S; R)$ simplifies to $I(Z, C; R)$.

In order to make the problem in Equation (1) computationally tractable, a parametrized space for the privacy mapping functions can be introduced, allowing for the optimization to be performed on the parameter space. For example, consider the following parameter space

$$\Theta(n, \Sigma) = \{(A_c, b_c)_{c \in \Sigma} | \forall c \in \Sigma : A_c \in \mathbb{R}^{n \times n}, b_n \in \mathbb{R}^n, \det(A_c) \neq 0\}$$

Then a parametrized space for affine privacy mapping functions on the classes set Σ and information space \mathcal{I} of dimension n can be defined as

$$I_D(n, \Sigma, \mathcal{I}) = \{R(\cdot; \theta) | \theta \in \Theta(n, \Sigma), R(\cdot; \theta) \in (\Sigma \rightarrow \mathcal{I}^{\mathcal{I}}), \forall c \in \Sigma : [R(c; \theta)](x) = A_c \cdot (x - b_c)\}$$

Provided a parameter search space Θ , the optimization problem in Equation (1) can be re-written as

$$\theta^* = \arg \min_{\theta \in \Theta} I(Z, C; R(\cdot; \theta)) \quad (2)$$

The straightforward way to modeling the required distributions $p(c)$ and $p(x|c)$, from data, is non-parametrically by using high-dimensional histograms. This approach, while simple to implement, suffers from the curse of dimensionality as its complexity grows exponentially with the dimension of the information space. Once the models for $p(c)$ and $p(x|c)$ are constructed, the model for $p(z|c)$ can be computed for any choice of $\theta \in \Theta$ allowing the computation of the objective function

⁷<https://www.mathworks.com/products/matlab/>

Table 1: BMI-for-age weight status categories and The corresponding BMI percentiles.

Weight Category	BMI Percentile Range
Underweight	$BMI < 5\%$
Healthy Weight	$5\% \leq BMI < 85\%$
Overweight	$85\% \leq BMI < 95\%$
Obese	$95\% \leq BMI$

in Equation (2). Since the problem is non-convex, in order to optimize the objective function, we employ the genetic algorithm with the fitness function equal to the objective function in Equation (2). The chosen selection policy is fitness-proportional while the chosen transformations (evolution/genetic) operators are both mutations and crossovers (Banzhaf et al. 1998).

6 Experimentation

In this section we walk the reader through an example that aims to motivate and demonstrate PDI. In this example we use data that are published by the Center for Disease Control and Prevention (CDC) as part of the National Health and Nutrition Examination Survey of 2012⁸. Specifically, we use the Body Measures (BMX_G) portion of the data⁹.

6.1 Setting

In our setting, we consider the disclosed information to be both Body Mass Index (BMI) and weight. Our information providers are assumed to be individuals of both genders that are 19 years of age or less. We consider the private information to be the weight status category of the subject. The CDC considers the following four standard weight status categories for the aforementioned age group *i*) underweight; *ii*) healthy weight; *iii*) overweight; and *iv*) obese. There are 3355 data points in the data set with subjects of 19 years of age or less.

According to the definitions of the CDC, the BMI category of a child or a teen is classified based on the individual’s BMI percentile among the same age and gender group as described in Table 1. Since the age of the information provider is not part of the information space, the inference of the weight status category of the information provider based on BMI and weight is not perfect. The data for the different classes are depicted in Figure 2.

⁸https://wwwn.cdc.gov/nchs/nhanes/search/nhanes11_12.aspx

⁹https://wwwn.cdc.gov/nchs/nhanes/2011-2012/BMX_G.htm

6.1.1 Inference Based on Original Data

Using the data, we trained 3 SVM classifiers with Gaussian kernels. The classifiers are aggregate in terms of the “positive” class in the following sense. The first classifier treats the “positive” class as the Underweight category (and so the “negative” class is the rest of the categories). The second classifier treats the “positive” class as either the underweight or healthy weight categories. Finally, the third classifier treats the “positive” class as any category except the obese category. We used a 40 – 60 split for training-testing. In numbers, we used 1371 data points for training and 1984 data points for testing.

The training for all SVMs was done using 10-fold cross-validation among the data in the training set to pick the best σ of the Gaussian kernels and the best box boundaries of the classifiers. The classification phase is done by taking a majority vote from the 3 classifiers and the output is the class which most classifiers agree on. The results of the classifier are described in Table 2 in terms of the confusion matrix of the different categories. The total accuracy of the classifier is 88.31%¹⁰.

6.2 Privatizing Information

We would like to privatize the information at hand (BMI and weight) in order to maintain the weight status category as private as possible (based on the training set only). This scenario simulates a tele-monitoring scenario and fits the assumptions and motivation introduced in Section 3. Therefore, we aim to utilize PDI in order to privatize the data as discussed earlier. In order to learn the privacy mapping function from the training data, we use the MATLAB toolbox mentioned in Section 5 (Aranki and Bajcsy, 2015). We used the affine privacy mapping functions for the parameterized search space as shown in the example in Section 5. Note that there are extra degrees of freedom in the problem, since any privacy mapping functions R_1 and R_2 related by $\forall c \in \Sigma : R_2(c) = A \cdot (R_1(c) - b)$ yield the same objective value in Equation (1) for any $A \in \mathbb{R}^{n \times n}$, $\det(A) \neq 0$ and $b \in \mathbb{R}^n$. That is, applying the same injective affine transformation to all encoding functions in R does not change the value of $I(Z, C|S; R)$. Therefore, in our problem we fix the encoding function of the “underweight” class to the identity function, i.e. $[R(\text{“underweight”})](x) = x$.

The resultant privatized information is depicted in Figure 3. It is clear that it should be much harder to do inference of the weight category based on this privatized data, given the decreased distinguishability between classes. Note that calculating the privatized information is simple and efficient since now we know the parameters for the privacy mapping functions for the different classes.

¹⁰The adopted total accuracy measure is $\text{trace}(M)/N$ where M is the confusion matrix and N is the cardinality of the test set. This is the percentage of true classifications over the test set.

Table 2: Confusion matrix before privatizing. UW = Underweight, HW = Healthy Weight, OW = Overweight, OB = Obese

Predicted Category		Ground Truth Category			
		UW	HW	OW	OB
UW		47	20	0	0
HW		14	1203	66	1
OW		0	45	194	47
OB		0	2	37	308

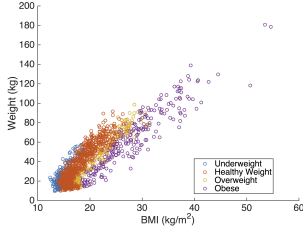


Figure 2: BMI and weight for the different weight status groups.

6.2.1 Inference Based on Privatized Data

In order to evaluate the quality of the privatization, we now train new 3 SVM classifiers with the same training procedure as in Section 6.1.1 but this time using the privatized data (and of course, encoding the test set too for evaluation). Same as before, we then use a majority vote from the 3 classifiers to predict the class of any data point. The resultant confusion matrix is described in Table 3.

It is clear that the classification results are degraded after privatizing the information. The total accuracy dropped to 66.03% (from 88.31%). Given that the data from different classes are highly indistinguishable, the classifier now classifies most data points as “healthy weight”. This is to be expected since most of the data points are in the “healthy weight” category. In informal words, if a classifier would have to make a “bet”, it would bet on the class with the most amount of data points. Formally, a lower bound on the total accuracy can be achieved by considering the trivial classifier that always predicts “healthy weight” (deterministic), which has total accuracy of $1270/1948 = 64.01\%$. This shows that our result of 66.03% is not much further from a lower-bound guaranteed accuracy.

Note that the data set is biased in size against the “underweight” category. There are only 126 data points with weight category “underweight” out of the 3355 total data points (3.76%). This makes privatizing that class particularly hard, especially because the modeling is based on n -dimensional histograms and is not parametric. For this reason the classification results before and after privatization for the

Table 3: Confusion matrix after privatizing. UW = Underweight, HW = Healthy Weight, OW = Overweight, OB = Obese

Predicted Category		Ground Truth Category			
		UW	HW	OW	OB
UW		48	14	0	5
HW		13	1217	276	290
OW		0	25	13	29
OB		0	14	0	32

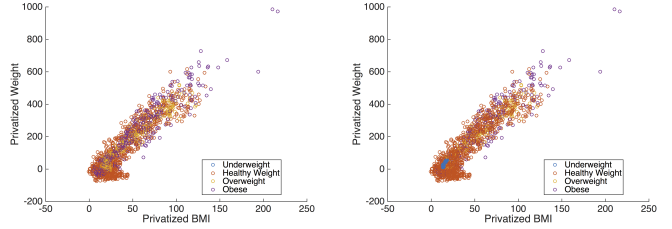


Figure 3: BMI and weight for the different weight status groups after privatization. The difference between the two plots is the order of plotting the different classes (for visual clarity).

“underweight” category are comparable.

To intuitively demonstrate how privacy is preserved, we take a piece of privatized information at random from our data set, $z = [77.17, 296.45]^T$, without looking at its ground truth weight category. If we decode this data point using the decoding function of “healthy weight”, we get $x = [21, 53.8]^T$, which is a legitimate “healthy weight” BMI and weight data point. If we use the decoding function of “overweight”, we get $x = [25.12, 62.4]^T$, which is also a legitimate “overweight” BMI and weight data point. Similarly, if we use the decoding function of “obese”, we get $x = [30.42, 69.08]^T$, which is also a legitimate “obese” BMI and weight data point.

7 Discussion and Future Work

In this paper, we presented a view on privacy in which the data themselves *need not* be the private object, but rather can *be used* to infer private information. From this point of view, we derived a framework that preserves the privacy of the private information from being inferred from the communicated messages. We provided theoretical analysis and properties of the devised framework. An important result (Theorem 1) provided conditions that ensure perfect privacy while preserving full data utility. We showed that such conditions are achievable by providing closed-form solutions to some cases of data generative models. Theorem 1 further showed that perfect privacy is not a function of the modeling of the adversary’s auxiliary

knowledge about the private information per subject, $p(c|s)$ (or $p(s)$). This observation is important because modeling adversary’s auxiliary knowledge is generally a hard problem, and because it showed that perfect privacy can be achieved regardless of the adversary’s auxiliary knowledge. That is, the same privatization protects information providers from all adversaries, regardless of their auxiliary knowledge.

Subsequently, we discussed an implementation of the learning problem resulting from the framework and demonstrated its use with a data set published by the Center for Disease Control and Prevention using data about individuals’ Body Mass Indices, weights and their weight status categories. The experimentation shows that after privatizing the data set, the classification accuracy drops significantly, near a lower bound of guaranteed classification accuracy, thus achieving our set goal.

We make two important remarks about the approach presented in this paper. First, the described approach is philosophically different from the classical cryptography as it provides a model where the objective is maintaining the secrecy of the private information that is not the data themselves but the information that can be inferred based on the data. Second, even though the proposed approach is privacy-centric, it is not meant to serve as an alternative to cryptography but as a complement to it. That said, any message can be “privatized” then encrypted. If the encryption is in that case compromised by an adversary getting access to the clear text message, the privacy is still preserved.

The current implementation of the devised learning problem suffers from the curse of dimensionality. The cost of learning grows exponentially with the number of dimensions of the information space. This is a result of our choice to model $p(z|c)$ as a multi-dimensional histogram. To make this framework practical, there is a need to study other ways of estimating the mutual information measure between the disclosed information and the private class. One appealing option is leveraging parametric learning and modeling each distribution $p(z|c)$ as a mixture model which could result in more computationally efficient estimation of the mutual information measure.

The presented framework has the potential of being extended to scenarios where the data recipient is not completely certain about the private class but is still more certain than the adversary. Such scenarios are clearly more general and may result in wider applicability of the framework to other scenarios than presented here. Indeed, in such scenarios, communicated messages can only be interpreted in a statistical sense and the implications of such assumptions must be studied as well.

Furthermore, the current implementation of the learning problem assumes that adversaries have equal belief about all the information providers so that the adversary’s belief about C is independent of S and that the generative model of data X per private class C is independent of S . This is a simplifying assumption and its implications need to be further studied and remedied.

Given the non-convexity and the complexity of the problem at hand, areas for future research include studying heuristic techniques to learn the privacy mapping functions from sufficient and/or necessary conditions for local improvements in the mutual information as a function of local changes in the privacy mapping functions. This approach, as opposed to finding global optimal solutions to Equation (1), is analogous to finding minimal anonymization as opposed to optimal anonymization in privacy preserving data publishing (Fung et al. 2010).

Acknowledgments

We would like to thank Katherine Driggs Campbell for the initial conversation that spurred this idea. We are also greatly indebted to Gregorij Kurillo, Yusuf Erol and Arash Nourian for their fruitful discussions and feedback that significantly improved the quality of this paper. This work was supported in part by TRUST, Team for Research in Ubiquitous Secure Technology, which receives funding support for the National Science Foundation (NSF award number CCF-0424422).

References

- Nabil R Adam and John C Worthmann. Security-control methods for statistical databases: A comparative study. *ACM Computing Surveys (CSUR)*, 21(4):515–556, 1989.
- Daniel Aranki and Ruzena Bajcsy. Private disclosure of information matlab toolbox, 2015. URL <https://www.eecs.berkeley.edu/~daranki/PDI/>
- Daniel Aranki, Gregorij Kurillo, Posu Yan, David Liebovitz, and Ruzena Bajcsy. Continuous, real-time, tele-monitoring of patients with chronic heart-failure - lessons learned from a pilot study. *ICST*, 11 2014. doi: 10.4108/icst.bodynets.2014.257036.
- Wolfgang Banzhaf, Peter Nordin, Robert E Keller, and Frank D Francone. *Genetic programming: An introduction*, volume 1. Morgan Kaufmann Publishers, Inc., 1998.
- Lynne Bishop, Bradford J Holmes, and Christopher M Kelley. National consumer health privacy survey 2005. *California HealthCare Foundation, Oakland, CA*, 2005.
- Sarwat I Chaudhry, Jennifer A Mattera, Jephtha P Curtis, John A Spertus, Jeph Herrin, Zhenqiu Lin, Christopher O Phillips, Beth V Hodshon, Lawton S Cooper, and Harlan M Krumholz. Telemonitoring in patients with heart failure. *New England Journal of Medicine*, 363(24):2301–2309, 2010.
- Robyn A Clark, Sally C Inglis, Finlay A McAlister, John GF Cleland, and Simon Stewart. Telemonitoring or structured

- telephone support programmes for patients with chronic heart failure: Systematic review and meta-analysis. *BMJ*, 334(7600):942, 2007.
- Graham Cormode. Personal privacy vs population privacy: Learning to attack anonymization. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1253–1261. ACM, 2011.
- Thomas M Cover and Joy A Thomas. *Elements of information theory*. John Wiley & Sons, 2 edition, 2006.
- Dorothy E. Denning. A lattice model of secure information flow. *Commun. ACM*, 19(5):236–243, May 1976. ISSN 0001-0782. doi: 10.1145/360051.360056. URL <http://doi.acm.org/10.1145/360051.360056>
- Dorothy E. Denning and Jan Schlorer. Inference controls for statistical databases. *Computer*, 16(7):69–82, 1983.
- George Duncan and Diane Lambert. The risk of disclosure for microdata. *Journal of Business & Economic Statistics*, 7(2):207–217, 1989. doi: 10.1080/07350015.1989.10509729. URL <http://www.tandfonline.com/doi/abs/10.1080/07350015.1989.10509729>
- George T Duncan and Diane Lambert. Disclosure-limited data dissemination. *Journal of the American statistical association*, 81(393):10–18, 1986.
- Cynthia Dwork. Differential privacy. In *Automata, languages and programming*, pages 1–12. Springer, 2006.
- Cynthia Dwork. Differential privacy: A survey of results. In *Theory and Applications of Models of Computation*, pages 1–19. Springer, 2008.
- Csilla Farkas and Sushil Jajodia. The inference problem: A survey. *SIGKDD Explor. Newsl.*, 4(2):6–11, December 2002. ISSN 1931-0145. doi: 10.1145/772862.772864. URL <http://doi.acm.org/10.1145/772862.772864>
- Benjamin Fung, Ke Wang, Rui Chen, and Philip S Yu. Privacy-preserving data publishing: A survey of recent developments. *ACM Computing Surveys (CSUR)*, 42(4):14, 2010.
- Gregory Giamouzis, Dimos Mastrogiannis, Konstantinos Koutrakis, George Karayannis, Charalambos Parisi, Chris Rountas, Elias Adreanides, George E Dafoulas, Panagiotis C Stafylas, John Skoularigis, et al. Telemonitoring in chronic heart failure: A systematic review. *Cardiology Research and Practice*, 2012, 2012.
- Aris Gkoulalas-Divanis, Grigorios Loukides, and Jimeng Sun. Publishing data from electronic health records while preserving privacy: A survey of algorithms. *Journal of biomedical informatics*, 50:4–19, 2014.
- Chun-Ju Hsiao and Esther Hing. *Use and characteristics of electronic health record systems among office-based physician practices, United States, 2001-2012*. US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics, 2012.
- Sally Inglis. Structured telephone support or telemonitoring programmes for patients with chronic heart failure. *Journal of Evidence-Based Medicine*, 3(4):228–228, 2010.
- Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *IEEE International Conference on Data Engineering*, volume 7, pages 106–115, 2007.
- Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data*, 1(1), March 2007. ISSN 1556-4681. doi: 10.1145/1217299.1217302. URL <http://doi.acm.org/10.1145/1217299.1217302>
- Brad Miller, Ling Huang, Anthony D Joseph, and J Doug Tygar. I know why you went to the clinic: Risks and realization of https traffic analysis. *arXiv preprint arXiv:1403.0297*, 2014.
- National Association of Health Data Organization. A guide to state-level ambulatory care data collection activities, October 1996.
- Ravi S Sandhu. Lattice-based access control models. *Computer*, 26(11):9–19, Nov 1993. ISSN 0018-9162. doi: 10.1109/2.241422.
- State of California Office of Statewide Health Planning and Development. *California inpatient data reporting manual, medical information reporting for California*, 7th edition, September 2014.
- Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.
- Stanley L Warner. Randomized response: A survey technique for eliminating evasive answer bias. *Journal of the American Statistical Association*, 60(309):63–69, 1965.
- Andrew M White, Austin R Matthews, Kevin Z Snow, and Fabian Monrose. Phonotactic reconstruction of encrypted voip conversations: Hookt on fon-iks. In *Security and Privacy (SP), 2011 IEEE Symposium on*, pages 3–18. IEEE, 2011.