Linking Differential Identifiability with Differential Privacy

Anis Bkakria(✉), Nora Cuppens-Boulahia, and Frédéric Cuppens

IMT Atlantique, 2 Rue de la Châtaigneraie, 35576 Cesson Sévigné, France
{anis.bkakria,nora.cuppens,frederic.cuppens}@imt-atlantique.fr

Abstract. The problem of preserving privacy while mining data has been studied extensively in recent years because of its importance for enabling sharing data sets. Differential Identifiability, parameterized by the probability of individual identification ρ, was proposed to provide a solution to this problem. Our study of the proposed Differential Identifiability model shows that: First, its usability is based on a very strong requirement. That is, the prior probability of an individual being present in a database is the same for all individuals. Second, there is no formal link between the proposed model and well known privacy models such as Differential Privacy. This paper presents a new differential identifiability model for preventing the disclosure of the presence of an individual in a database while considering an adversary with arbitrary prior knowledge about each individual. We show that the general Laplace noise addition mechanism can be used to satisfy our new differential identifiability definition and that there is a direct link between differential privacy and our proposed model. The evaluation of our model shows that it provides a good privacy/utility trade-off for most aggregate queries.

1 Introduction

Many privacy models have been proposed for protecting individuals' privacy in published data, e.g., k-anonymity [14], l-diversity [13], t-closeness [10], etc. These models suffer from a key limitation: They cannot guarantee that the relationship between individuals' identities and their sensitive information are protected in case in which the adversary has additional knowledge. A privacy notion that is progressively gaining acceptance for overcoming the previously mentioned privacy problem is differential privacy (DP). Informally, DP requires that the impact of the presence of any individual entity in a dataset on the output of the queries to be limited. More specifically, DP ensures that any two databases that differ only in one record will induce output distributions that are close in the manner that the probabilities of each possible query's outputs differ by a bounded multiplicative factor ϵ.

The authors thank the Région Bretagne, CNRS, IMT, FEDER, CD 35 and Rennes Métropole for their support through the CPER Cyber SSI.

D. Naccache et al. (Eds.): ICICS 2018, LNCS 11149, pp. 232–247, 2018.
https://doi.org/10.1007/978-3-030-01950-1_14

Several research have investigated whether DP can provide sufficient protection and how to choose the right value for the parameter ϵ. Lee and Clifton showed in [9] that the DP's parameter ϵ can only limit how much one individual can change the output of a query. It does not limit the amount of information that are revealed about an individual. This limitation makes DP not fully matching the legal definition of privacy that requires the protection of individually identifiable data. Attempting to meet the previous privacy definition, Lee and Clifton proposed in [9] a new privacy model called *differential identifiability* (DI). They assume that a database record can be linked to the identity of an individual, and they provide a model to quantify the leakage of the information on whether an individual participates in the database or not. Informally, if we denote by *possible worlds* the set of all possible databases resulting from removing an (any) individual from the initial database, DI ensures that the identifiability risk of any individual in the universe is less than or equal to a parameter ρ. This parameter can be interpreted as the degree of indistinguishability between possible worlds, where the possible worlds differ by (any) one individual. Unfortunately, the proposed model is based on the assumption that the prior probability of an individual being in the database is the same for all individuals. We believe that this is a very strong requirement since it requires an adversary to know exactly the same amount of information about each individual in the database. Clearly, this assumption is seldom satisfied in the real environments. Moreover, There is no direct translation from the DI parameter ρ to the DP parameter ϵ, and thus, the data utility may become unable to estimate.

In this paper, we try to remedy the previously mentioned drawbacks by proposing a new model called (α, β)-*DI*. The model aims to limit the leakage of information on whether an individual participates in a database or not when considering an adversary with arbitrary prior knowledge about each individual in the database (the same strong guarantees as DP). We show that the general Laplacian noise addition mechanism for differential privacy can be adapted to provide (α, β)-differentially identifiable outputs and that there is a direct translation between DP and our (α, β)-DI model. In a thorough experimental evaluation on real datasets, we studied the utility that can be provided by our model for several kinds of statistical queries.

The rest of the paper is organized as follows. Section 2 introduces the notations and preliminaries that we are going to use. Section 3 presents the problem we address and the adversary model we consider in our work. In Sect. 4, we first show how to model the belief of an adversary about the individuals present in the database, and second, how the belief of the adversary will change when he/she interacts with the database. Section 5 defines our new Differential Identifiability model. In Sect. 6, we studied whether is it possible to provide privacy and utility without making assumptions about the prior knowledge of the adversary. Then we propose a general Laplacian noise addition mechanism to satisfy (α, β)-DI. Section 7 presents a translation from the two parameters α and β we are using in our model to the DP parameter ϵ. Section 8 evaluates the utility/privacy tradeoff provided by our model for different kinds of aggregate queries. We discuss related work in Sect. 9 and conclude the paper in Sect. 10.

Table 1. List of symbols

D	Database to be queried
D^s	Database containing records having the sensitive property s
M	A privacy preserving mechanism
$\mathcal{U}$	The universe of individuals
ι	An entity in the universe $\mathcal{U}$

2 Notations and Preliminaries

In our model, we used the set of notations given in Table 1. A dataset D is generated from the data associated with a subset of entities in $\mathcal{U}$. For all $D', D^s \in \mathcal{U}$, the prior belief that some database D' is equal to D^s is given by $\mathcal{B}_\emptyset(D' = D^s)$. The posterior belief that some database D' is equal to D^s after observing the response τ of a query q is given by $\mathcal{B}_{q,\tau}(D' = D^s)$.

Definition 1 (Adjacent Databases). *Two databases D_1 and D_2 are adjacent ($D_1 \overset{\iota}{\sim} D_2$) if they differ on the data of a single individual ι.*

For sake of simplicity, we will suppose that each individual has only one record in the database. That is, two adjacent databases differ only in one record.

Definition 2 (Global Sensitivity). *Given a query function $q : \mathcal{U} \rightarrow \mathbb{R}$. The global sensitivity of q is defined as following:*

$$\Delta_q = \max_{\forall D, D' \in \mathcal{U}} |q(D) - q(D')|$$

where D and D' are adjacent database and $q(D)$ denotes the result of the execution of the q over the database D.

3 Problem Statement and Adversary Model

We consider a database D containing a set of information about a set of individuals in $\mathcal{U}$, and D^s the database containing the set of individuals in D having a sensitive property s (e.g., the set of individuals having VIH). As in the DP model, we consider a very strong adversary who knows every single information in D. That is, we suppose that the adversary knows every attribute value in D. In addition, we suppose that the adversary knows that D^s is composed of individuals who have s and that he/she don't know which individuals in D are in D^s. Considering that a privacy-preserving data analysis aims to release analysis results without revealing the identities of the individuals, a privacy breach is then to allow an adversary to figure out individual's presence/absence in D^s.

In our model, we suppose that the adversary has an infinite computational power which will be used to identify the set of individuals in D^s by combining

the knowledge of D and the results of the queries to be executed over D^s. This is identical to finding out the set of missing individuals in D^s from D. In our work, we will consider the worst case in which D and D^s are adjacent databases. That is, the adversary has to find out only the missing individual in D^s to know all individuals in D^s. In the remaining of this paper, we will use D^s to represent a D's adjacent database where all individual in D^s have a sensitive property s.

4 Adversary Knowledge Modeling

The key to a good privacy model is to correctly quantify how much information an adversary can deduce about the presence of an individual in the published data. This heavily depends on the knowledge the adversary possesses about the individuals in the database. Adversary belief changes each time a result of a query performed over D^s is observed by the adversary. We use the Bayesian inference to model an adversary belief change as defined in the following definition.

Definition 3 (Query observation impact on adversary belief). *For all two pairs of adjacent databases $D{\sim}D'$ and $D{\sim}D^s$ where $D, D', D^s \in \mathcal{U}$, given a query function $q : \mathcal{U} \to \mathbb{R}$, a mechanism M, and $\tau = M(q(D^s))$ the result of the execution of q using M. The adversary belief on $D' = D^s$ after observing $\tau = M(q(D^s))$ is defined as:*

$$\begin{aligned} \mathcal{B}_{q,\tau}(D' = D^s) &= Pr\big[D' = D^s | M(q(D^s)) = \tau\big] \\ &= \frac{Pr\big[M(q(D')) = \tau\big]}{Pr\big[(q,\tau)\big]} \quad \times \quad Pr\big[D' = D^s\big] \end{aligned} \tag{1}$$

where $Pr\big[D' = D^s\big]$ denotes the prior belief ($\mathcal{B}_\emptyset(D' = D^s)$) of the adversary on $D = D^s$ (before observing $\tau = M(q(D^s))$) and $Pr\big[(q,\tau)\big]$ denotes the probability of observing the result τ when the query q is performed.

5 Differential Identifiability: The New Model

Definition 4 ((α,β)-DI). *Given a query function $q : \mathcal{U} \to \mathbb{R}$, a randomized mechanism M is said to be (α,β)-differentially identifiable if for all two pairs of adjacent databases $D{\sim}D'$ and $D{\sim}D^s$ where $D, D', D^s \in \mathcal{U}$:*

$$(1-\alpha) \times \mathcal{B}_\emptyset(D' = D^s) \leq \mathcal{B}_{q,\tau}(D' = D^s) \leq (1+\beta) \times \mathcal{B}_\emptyset(D' = D^s) \tag{2}$$

where $0 < \alpha < 1$, $0 < \beta$, and τ denotes the result observed by the adversary for $M(q(D^s))$.

Informally, the randomized mechanism M is (α,β)-differentially identifiable means that the ratio of the adversary belief on $D' = D^s$ before and after observing $M(q(D^s)) = \tau$ is lower and upper bounded respectively by $1-\alpha$ and $1+\beta$. The identification risks represented by the lower bound $1-\alpha$ and the upper

bound $1+\beta$ are not the same. In the left side of Inequality (2), the value of α bounds the maximum attacker belief change on identifying the presence of the individual ι in the database D^s, where $D' \overset{\iota}{\sim} D$. More α is bigger, more the adversary belief in $D' = D^s$ will be smaller, and more the adversary belief in $D \overset{\iota}{\sim} D^s$ will be also smaller. In the right side of Inequality (2), the value of β bounds the maximum belief of an attacker on identifying all the individuals present in the database D^s. More β is bigger more the adversary belief in $D \overset{\iota}{\sim} D^s$ will be bigger.

Most existing privacy frameworks bound only the adversary's belief on the presence of one individual in the database. We believe that bounding the adversary's belief on identifying all individuals in the database is very useful. To illustrate, let us suppose that the database D contains information about 10 individuals and that the prior adversary's belief that each individual in $D \cap D^s$ is 10^{-1}. Now if we suppose that the data publisher wants to bound the probability of identifying the presence of an individual in D^s to 1/5, the adversary can end up with the following belief: for 9 individual, the probability that each one of them is in D^s is equal to $(10^{-1} - 10^{-6})/9$. For the last individual, the probability that he/she is in D^s is equal to 10^{-6}. Since D and D^s are neighboring database, the adversary might know all individual in D^s with a probability of $1 - 10^{-6}$.

We studied how our definition of DI composes. Given a data consumer (adversary) who access a database multiple times via differentially identifiable mechanisms each of which having its own DI guarantees, what level of DI is still guaranteed on the union of those outputs? In order to formally define composition, we consider a similar composition scenario as the one proposed in [6]. A composition experiment considers an adversary $\mathcal{A}$ who is trying to break privacy and figure out whether or not a particular individual is in the database by analyzing the hypotheses on the output of a sequential and adaptively chosen queries executed via differentially identifiable mechanisms. That is, we permit the adversary to have full control over which query to ask, and which differentially identifiable mechanism to be used for each query. In addition, the adversary is free to make these choices adaptively based on previous queries outcomes.

Theorem 1. *Given a set of queries functions* $\mathcal{Q} = \{q_1, \cdots, q_n\}$ $(\forall i \in [1,n], q_i : \mathcal{U} \rightarrow \mathbb{R})$ *and a set of* n *mechanisms* $M_1, \cdots, M_n$. *Each* M_i, $i \in [1,n]$, *is* (α_i, β_i)*-differentially identifiable. Then for all databases* D, D^s, *where* $D \sim D^s$, *the combination* $\mathcal{M} = (M_1(q_1(D^s)), M_2(q_2(D^s)), \cdots, M_n(q_n(D^s)))$ *is* (α_c, β_c)*-differentially identifiable where:*

$$\alpha_c = \sum_{k=1}^{n} (-1)^{k+1} \sigma_k(\alpha_1, \cdots, \alpha_n) \quad \text{and} \quad \beta_c = \sum_{k=0}^{n} \big(\sigma_k(\beta_1, \cdots, \beta_n)\big) - 1$$

with σ_k *denotes the elementary symmetric polynomials.*

Proof. Let us suppose that $\forall i \in [1,n] : M_i(q_i(D^s)) = \tau_i$ and that $\mathcal{R} = \{\tau_i | i \in [1,n]\}$. First, Let us prove by induction that, for all two pairs of adjacent databases $D \sim D'$ and $D \sim D^s$ where $D, D', D^s \in \mathcal{U}$, the belief of the adversary

on D' equals to D^s ($\mathcal{B}_{Q,K}(D' = D^s)$) after the observation of the results of the set of n arbitrary and adaptively chosen queries $\mathcal{Q}$ is bounded as following:

$$\prod_{i=1}^{n}(1-\alpha_i) \times \mathcal{B}_{\emptyset}(D' = D^s) \leq \mathcal{B}_{\mathcal{Q},\mathcal{R}}(D' = D^s) \leq \prod_{i=1}^{n}(1+\beta_i) \times \mathcal{B}_{\emptyset}(D' = D^s) \tag{3}$$

By definition (Definition 4), Inequality (3) holds for $n = 1$. That is, when using an (α_1, β_1)-differentially identifiable mechanism M_1 to perform q_1, based on Definition 4 we get:

$$(1-\alpha_1) \times \mathcal{B}_{\emptyset}(D' = D^s) \leq \mathcal{B}_{q_1,r_1}(D' = D^s) \leq (1+\beta_1) \times \mathcal{B}_{\emptyset}(D' = D^s) \tag{4}$$

Suppose now that Inequality (3) holds for $n = k$. Then, by denoting $\mathcal{Q}^k = \{q_1, q_2, \cdots, q_k\}$ and $\mathcal{R}^k = \{r_1, r_2, \cdots, r_k\}$, the following inequality holds:

$$\prod_{i=1}^{k}(1-\alpha_i) \times \mathcal{B}_{\emptyset}(D' = D^s) \leq \mathcal{B}_{\mathcal{Q}^k,\mathcal{R}^k}(D' = D^s) \leq \prod_{i=1}^{k}(1+\beta_i) \times \mathcal{B}_{\emptyset}(D' = D^s) \tag{5}$$

Let us now prove that the Inequality (3) holds for $n = k+1$. Since the adversary will observe the result of the query q_{k+1} after observing the results of the previous k queries $q_1, \cdots, q_k$. The adversary belief on D' equals to D^s before observing the output of q_{k+1} is $\mathcal{B}_{\mathcal{Q}^k,\mathcal{R}^k}(D' = D^s)$. By considering the fact that q_{k+1} is performed using the $(\alpha_{k+1}, \beta_{k+1})$-differential identifiable mechanism M_{k+1}, based on Definition 4, we get:

$$\begin{aligned}(1-\alpha_{k+1}) \times \mathcal{B}_{\mathcal{Q}^k,\mathcal{R}^k}(D' = D^s) \leq \mathcal{B}_{\mathcal{Q}^{k+1},\mathcal{R}^{k+1}}(D' = D^s) \leq (1+\beta_{k+1}) \\ \times \mathcal{B}_{\mathcal{Q}^k,\mathcal{R}^k}(D' = D^s)\end{aligned} \tag{6}$$

Since, $0 < \alpha < 1$ and that we supposed that Inequality (5) holds, we can use its left side to show that:

$$\prod_{i=1}^{k+1}(1-\alpha_i) \times \mathcal{B}_{\emptyset}(D' = D^s) \leq (1-\alpha_{k+1}) \times \mathcal{B}_{\mathcal{Q}^k,\mathcal{R}^k}(D' = D^s) \tag{7}$$

Then using the fact that $\beta > 0$ together with the right side of Inequality (5), we get:

$$(1+\beta_{k+1}) \times \mathcal{B}_{\mathcal{Q}^k,\mathcal{R}^k}(D' = D^s) \leq \prod_{i=1}^{k+1}(1+\beta_i) \times \mathcal{B}_{\emptyset}(D' = D^s) \tag{8}$$

Then based on Inequalities (6), (7), and (8) we get:

$$\begin{aligned}\prod_{i=1}^{k+1}(1-\alpha_i) \times \mathcal{B}_{\emptyset}(D' = D^s) \leq \mathcal{B}_{\mathcal{Q}^{k+1},\mathcal{R}^{k+1}}(D' = D^s) \leq \prod_{i=1}^{k+1}(1+\beta_i) \\ \times \mathcal{B}_{\emptyset}(D' = D^s)\end{aligned} \tag{9}$$

which prove that Inequality (3) holds for $n = k+1$, and by induction it holds for all $n \in \mathbb{N}^*$. Now, based on the fundamental theorem of symmetric polynomials we have:

$$\begin{aligned}\prod_{i=1}^{n}(1-\alpha_i) &= 1 + \sum_{k=1}^{n}(-1)^k \sigma_k(\alpha_1, \cdots, \alpha_n) \\ &= 1 - \underbrace{\sum_{k=1}^{n}(-1)^{k+1}\sigma_k(\alpha_1, \cdots, \alpha_n)}_{\alpha_c}\end{aligned}$$

and

$$\begin{aligned}\prod_{i=1}^{n}(1+\beta_i) &= \sum_{k=0}^{n}\sigma_k(\beta_1, \cdots, \beta_n) \\ &= 1 + \underbrace{\sum_{k=1}^{n}\big(\sigma_k(\beta_1, \cdots, \beta_n)\big) - 1}_{\beta_c}\end{aligned}$$

6 Satisfying Differential Identifiability

Given the above, in this section, we show how to achieve (α, β)-DI. For this, we first define the *identifiability sensitivity* of a query as following.

Definition 5 (Query Identifiability Sensitivity). *For a given query function $q : \mathcal{U} \rightarrow \mathbb{R}$, the query identifiability sensitivity of q is*

$$\Theta_q = \max_{D, D_1, D_2 \in \mathcal{U}} \big|q(D_1) - q(D_2)\big|$$

where D_1 and D_2 are adjacent to D.

Note that the Identifiability Sensitivity of a query is different than its Global Sensitivity (Definition 2) used in DP. The Identifiability Sensitivity of a query q represents, for all two pairs of adjacent databases ($D \sim D_1$) and ($D \sim D_2$) in $\mathcal{U}$, the maximum difference between the outputs that q return when executed over D_1 and D_2.

Motivated by the difficulty for a data publisher to know the prior knowledge of an adversary about each individual in the database, we firstly investigate the achievement of the (α, β)-DI model without taking into consideration the prior knowledge of the adversary. The following theorem defines a prior-free Laplace distribution-based mechanism that achieves (α, β)-DI.

Theorem 2 (Prior-free mechanism). *Let $Lap(\lambda)$ be the Laplace distribution having a density function $h(x) = \frac{1}{2\lambda}\exp(-\frac{|X-\mu|}{\lambda})$ where $\lambda(>0)$ is a scale factor and μ is a mean. For a given query function q, a randomized mechanism M_L*

that returns $q(X)+Y$ as an answer where Y is drawn i.i.d from $Lap(\lambda)$ satisfies (α,β)-DI for any λ such that:

$$\lambda \geq \max\left(\frac{\Theta_q}{\log(1+\beta)}, \frac{-\Theta_q}{\log(1-\alpha)}\right)$$

Proof. Since $M_L = q(X) + Y$ where Y is drawn i.i.d from $Lap(\lambda)$, then, for all two pairs of adjacent databases $(D{\sim}D_1)$ and $(D{\sim}D_2)$ in $\mathcal{U}$, we have:

$$\begin{aligned}\frac{Pr\big[M_L(q(D_1')) = \tau\big]}{Pr\big[M_L(q(D_2')) = \tau\big]} &= \frac{\exp(-\frac{|\tau - q(D_1')|}{\lambda})}{\exp(-\frac{|\tau - q(D_2')|}{\lambda})} \\ &= \exp(\frac{|r - q(D_2')| - |r - q(D_1')|}{\lambda})\end{aligned}$$

we deduce then the following inequality:

$$\exp\left(-\frac{|q(D_1') - q(D_2')|}{\lambda}\right) \leq \frac{Pr\big[M_L(q(D_1')) = \tau\big]}{Pr\big[M_L(q(D_2')) = \tau\big]} \leq \exp\left(\frac{|q(D_1') - q(D_2')|}{\lambda}\right) \tag{10}$$

Then using Definition 5, we get:

$$\exp\left(-\frac{\Theta_q}{\lambda}\right) \leq \frac{Pr\big[M_L(q(D_1')) = \tau\big]}{Pr\big[M_L(q(D_2')) = \tau\big]} \leq \exp\left(\frac{\Theta_q}{\lambda}\right) \tag{11}$$

In other hand, using Definition 3, and for all two pairs of adjacent databases $(D{\sim}D_i')$ and $(D{\sim}D^s)$ in $\mathcal{U}$, we have

$$\begin{aligned}\mathcal{B}_{q,\tau}(D_i' = D^s) &= \frac{Pr\big[M_L(q(D_i')) = \tau\big] \times Pr\big[D_i' = D^s\big]}{\sum\limits_{D_j' \in \mathcal{D}'} Pr[D_j' = D^s] \times Pr[M_L(q(D_j')) = \tau]} \\ &= \frac{Pr\big[D_i' = D^s\big]}{Pr\big[D_i' = D^s\big] + \sum\limits_{D_j' \in \mathcal{D}', D_j' \neq D_i'} Pr[D_j' = D^s] \times \frac{Pr[M_L(q(D_j')) = \tau]}{Pr[M_L(q(D_i')) = \tau]}}\end{aligned}$$

Then using Inequality 11 we deduce

$$\frac{Pr\big[D_i' = D^s\big]}{Pr\big[D_i' = D^s\big] + \exp\left(\frac{\Theta_q}{\lambda}\right) \sum\limits_{D_j' \in \mathcal{D}', D_j' \neq D_i'} Pr[D_j' = D^s]} \leq \mathcal{B}_{q,\tau}(D_i' = D^s) \tag{12}$$

And

$$\mathcal{B}_{q,\tau}(D_i' = D^s) \leq \frac{Pr\big[D_i' = D^s\big]}{Pr\big[D_i' = D^s\big] + \exp\left(-\frac{\Theta_q}{\lambda}\right) \sum\limits_{D_j' \in \mathcal{D}', D_j' \neq D_i'} Pr[D_j' = D^s]} \tag{13}$$

Now, based on the fact that $\sum_{D'_j \in \mathcal{D}', D'_j \neq D'_i} Pr[D'_j = D^s] = 1 - Pr[D'_i = D^s]$, Inequality (12) can be transformed as

$$\frac{Pr\left[D'_i = D^s\right]}{Pr\left[D'_i = D^s\right]\left(1 - \exp\left(\frac{\Theta_q}{\lambda}\right) + \frac{\exp\left(\frac{\Theta_q}{\lambda}\right)}{Pr\left[D'_i = D^s\right]}\right)} \leq \mathcal{B}_{q,\tau}(D'_i = D^s)$$
$$\frac{1}{1 - \exp\left(\frac{\Theta_q}{\lambda}\right) + \frac{\exp\left(\frac{\Theta_q}{\lambda}\right)}{Pr\left[D'_i = D^s\right]}} \leq \tag{14}$$

Since $1 - \exp\left(\frac{\Theta_q}{\lambda}\right) \leq 0$ and by considering $Pr[D'_i = D^s] = \mathcal{B}_{\emptyset}(D'_i = D^s)$ (Definition 3), we obtain

$$\exp\left(-\frac{\Theta_q}{\lambda}\right) \leq \frac{\mathcal{B}_{q,\tau}(D'_i = D^s)}{\mathcal{B}_{\emptyset}(D'_i = D^s)} \tag{15}$$

We apply the same transformations to Inequality (13) to get

$$\frac{\mathcal{B}_{q,\tau}(D'_i = D^s)}{\mathcal{B}_{\emptyset}(D'_i = D^s)} \leq \exp\left(\frac{\Theta_q}{\lambda}\right) \tag{16}$$

Using Inequalities (15) and (16) together with Definition 4, $M_L = q(X) + Y$ where Y is drawn i.i.d from $Lap(\lambda)$ satisfies (α, β)-DI if:

$$1 - \alpha \leq \exp\left(-\frac{\Theta_q}{\lambda}\right) \quad and \quad \exp\left(\frac{\Theta_q}{\lambda}\right) \leq 1 + \beta$$

Rearranging yields

$$\lambda \geq \frac{\Theta_q}{\log(1+\beta)} \quad and \quad \lambda \geq \frac{-\Theta_q}{\log(1-\alpha)}$$

Finally, we obtain the following

$$\lambda \geq \max\left(\frac{\Theta_q}{\log(1+\beta)}, \frac{-\Theta_q}{\log(1-\alpha)}\right)$$

The previous theorem uses Laplace distribution to satisfy (α, β)-DI without taking into consideration the prior knowledge of the adversary about the presence of each individual in the database D^s. The proposed construction seems to be useful to satisfy (α, β)-DI in case in which the prior knowledge of the adversary could not be known in advance. Unfortunately, in practice, it is not possible to properly instantiate our previous construction, i.e., to find the right values of α and β that make the model useful for an adversary having arbitrary prior belief. That is, in one hand, the values of α and β should be non-negligible so that the model provides an acceptable utility level for the queries that will be performed

by the adversary over the database. In the other hand, the value of α and β should not be bigger enough to allow the adversary to be sure about the presence of any individual in the database. Let us suppose that the adversary is not fully sure that an individual ι is in the database D^s (i.e., $Pr[\iota \in D^s] < 1$). If we consider only the left-hand side of Inequality (2), for any value of $\alpha \in]0, 1[$, our model will still ensure that the adversary cannot be 100% sure that ι is in D^s. Nevertheless, the previous construction may allow the adversary to be pretty much sure that ι is in the database D^s (i.e., for $D' \overset{\iota}{\sim} D : \mathcal{B}_{q,\tau}(D' = D^s)$ is very close to zero). Things are much more difficult for choosing the right value of β. According to the definition of our model (Definition 4), to prevent the adversary from knowing with certainty all individuals in the database D^s, the data publisher should choose a β value such that: $\mathcal{B}_{\emptyset}(D' = D^s) \times (1 + \beta) < 1$. Unfortunately, satisfying the previous condition becomes not possible if the adversary prior on $D' = D^s$ is not taken into consideration.

Seeking to overcome the previous limitation, we define a prior-dependent Laplace distribution based mechanism for achieving (α, β)-DI. The following theorem gives a lower bound for the quantity of Laplace noise to be added to the response of a query q to achieve (α, β)-DI for a given adversary's prior distribution $\mathbb{P}$.

Theorem 3 (Prior-dependent mechanism). *For all database $D \in \mathcal{U}$ of size $n(> 1)$, let $\mathcal{D}'$ be the set of D's adjacent databases. For a given prior distribution $\mathbb{P}$, For a given query function $q : \mathcal{U} \to \mathbb{R}$, a randomized mechanism M_L that returns $q(X) + Y$ as an answer where Y is drawn i.i.d from $Lap(\lambda)$ satisfies (α, β)-DI for any λ such that:*

$$\lambda \geq \Theta_q \times \max\left(\log\left(\frac{1 + P_{min}(\alpha - 1)}{(1 - \alpha)(1 - P_{min})}\right)^{-1}, \log\left(\frac{(1 + \beta)(1 - P_{min})}{1 - P_{min}(1 + \beta)}\right)^{-1}\right)$$

were $0 < \alpha < 1, 0 < \beta < (1/P_{max}) - 1$, $P_{min} = \min_{D_j, D^s \in \mathcal{D}'} Pr[D_j = D^s]$, *and* $P_{max} = \max_{D_j, D^s \in \mathcal{D}'} Pr[D_j = D^s]$.

We note that in the previous theorem, condition $\beta < (1/P_{max}) - 1$ is used to be sure that for any possible values of α and β, the usage of M_L will effectively prevent the adversary from knowing with certainty the content of the database $D^s \in \mathcal{D}'$. Obviously, P_{max}'s value should be lesser than 1. Otherwise, there are no possible values for α and β that can prevent the adversary from knowing with certainty the content of the database D^s, since he/she already does.

Proof. To prove the previous theorem, we start by following the same steps as in the proof of Theorem 2 to get Inequalities (12) and (13). By considering the fact that $\sum_{D'_j \in \mathcal{D}', D'_j \neq D'_i} Pr[D'_j = D^s] = 1 - Pr[D'_i = D^s]$, we transform Inequality (13) to get Inequality (14) which will be transformed as following:

$$\frac{1}{Pr\left[D'_i = D^s\right] + \exp\left(\frac{\Theta_q}{\lambda}\right)\left(1 - Pr\left[D'_i = D^s\right]\right)} \leq \frac{\mathcal{B}_{q,\tau}(D'_i = D^s)}{\mathcal{B}_{\emptyset}(D'_i = D^s)} \tag{17}$$

Since $P_{min} \leq Pr\left[D'_i = D^s\right] \leq P_{max}$, we have:

$$\frac{1}{P_{min}\left(1-\exp\left(\frac{\Theta_q}{\lambda}\right)\right)+\exp\left(\frac{\Theta_q}{\lambda}\right)} \leq \frac{1}{Pr\left[D'_i = D^s\right]+\exp\left(\frac{\Theta_q}{\lambda}\right)\left(1-Pr\left[D'_i = D^s\right]\right)} \tag{18}$$

Using Inequalities (17) and (18) together with Definition 4, M_L satisfies (α, β)-DI if:

$$1-\alpha \leq \frac{1}{P_{min}\left(1-\exp\left(\frac{\Theta_q}{\lambda}\right)\right)+\exp\left(\frac{\Theta_q}{\lambda}\right)}$$

Since $P_{min} \leq 1/n$, we have: $1 + P_{min}(\alpha - 1) > 0$. Then, rearranging yields

$$\lambda \geq \Theta_q \log\left(\frac{1+P_{min}(\alpha-1)}{(1-\alpha)(1-P_{min})}\right)^{-1} \tag{19}$$

On the other hand, by considering the fact that $\sum_{D'_j \in \mathcal{D}', D'_j \neq D'_i} Pr[D'_j = D^s] = 1 - Pr[D'_i = D^s]$, we transform Inequality (13) to get

$$\frac{\mathcal{B}_{q,\tau}(D'_i = D^s)}{\mathcal{B}_{\emptyset}(D'_i = D^s)} \leq \frac{1}{Pr\left[D'_i = D^s\right]+\exp\left(\frac{-\Theta_q}{\lambda}\right)\left(1-Pr\left[D'_i = D^s\right]\right)} \tag{20}$$

Then, considering the fact that $P_{min} \leq Pr\left[D'_i = D^s\right] \leq P_{min}$, we have:

$$\frac{1}{Pr\left[d_i = d^s\right]+\exp\left(\frac{-\Theta_q}{\lambda}\right)\left(1-Pr\left[d_i = d^s\right]\right)} \leq \frac{1}{P_{min}\left(1-\exp\left(\frac{-\Theta_q}{\lambda}\right)\right)+\exp\left(\frac{-\Theta_q}{\lambda}\right)} \tag{21}$$

Using Inequalities (20) and (21) together with Definition 4, M_L satisfies (α, β)-DI if:

$$\frac{1}{P_{min}\left(1-\exp\left(\frac{-\Theta_q}{\lambda}\right)\right)+\exp\left(\frac{-\Theta_q}{\lambda}\right)} \leq 1+\beta \tag{22}$$

Since $P_{min} \leq 1/n$, for all $n > 1$, we have: $1 - P_{min}(1+\beta) > 0$. Then, rearranging yields

$$\lambda \geq \Theta_q \log\left(\frac{(1+\beta)(1-P_{min})}{1-P_{min}(1+\beta)}\right)^{-1} \tag{23}$$

Finally, based on Inequalities (19) and (23), we have:

$$\lambda \geq \Theta_q \times \max\left(\log\left(\frac{1+P_{min}(\alpha-1)}{(1-\alpha)(1-P_{min})}\right)^{-1}, \log\left(\frac{(1+\beta)(1-P_{min})}{1-P_{min}(1+\beta)}\right)^{-1}\right)$$

In contrast to the original Differential Identifiability model propose in [9] which assumes that the prior probability of an individual being in D^s is the same for all individuals, our previous construction defines a Laplace distribution-based mechanism that provides an (α, β)-differentially identifiable outputs for any arbitrary prior distribution.

7 Linking Differential Identifiability and Differential Privacy

In this section, we establish a fundamental connection between DP model and our DI model by showing that the parameter ϵ used in the DP model can be directly translated to the parameters α and β used in our DI model.

Theorem 4. *Let M_L be a mechanisms that satisfies (α, β)-DI for a given query $q : \mathcal{U} \rightarrow \mathbb{R}$ by returning $q(X) + Y$ where Y is drawn i.i.d from $Lap(\lambda)$. M_L satisfies ϵ-DP where*

$$\epsilon = \frac{\Delta_q}{\Theta_q} \times \max\left(\log\left(\frac{1 + P_{min}(\alpha - 1)}{(1-\alpha)(1 - P_{min})}\right)^{-1}, \log\left(\frac{(1+\beta)(1-P_{min})}{1 - P_{min}(1+\beta)}\right)^{-1}\right)^{-1} \tag{24}$$

Proof. Since M_L satisfies (α, β)-DI, then using Theorem 3 we have:

$$\lambda = \Theta_q \times \max\left(\log\left(\frac{1 + P_{min}(\alpha - 1)}{(1-\alpha)(1 - P_{min})}\right)^{-1}, \log\left(\frac{(1+\beta)(1-P_{min})}{1 - P_{min}(1+\beta)}\right)^{-1}\right) \tag{25}$$

In other hand, based on Differential Privacy's Laplace mechanism definition [5], we know that M_L satisfies ϵ-DP when

$$\lambda = \frac{\Delta_q}{\epsilon} \tag{26}$$

Finally using Eqs. (25) and (26), we get (24).

Choosing the appropriate value of ϵ is continuing to be an open problem in DP. The connection we created between ϵ-DP and (α, β)-DI models in Theorem 4 will allow to choose the appropriate ϵ value given the risk of identifying the presence of an individual in the database specified by α and β.

8 Evaluation

We now evaluate the applicability of our model. For this, we use the Adult Database from the UCI Machine Learning Repository [1] as $\mathcal{U}$ (the universe of individuals). The database contains information about 32562 individuals collected from the 1994 U.S. Census. The information about each individual is provided through 9 categorical and 5 numerical attributes. In this evaluation, we consider only numerical attributes. In order to evaluate the applicability of our model, we quantify, for several aggregate queries (e.g., sum, average, max, min, etc.), the error ratio caused by the usage our prior-dependent mechanism (Theorem 3) when the values of α and β are varied. Since, it is not possible to graphically illustrate the variation of the error ratio in function of more than two

variables (i.e., α, β, and $\mathbb{P}$), for the prior distribution $\mathbb{P}$, we will consider two main cases. First, we will consider a very weak adversary $\mathcal{A}_w$. That is, he/she does not have any information about the individuals in the database $D^s \in \mathcal{U}$ (uniform prior distribution: $P_{min} = P_{max} = 1/32562$). Second, we will consider a strong adversary $\mathcal{A}_s$ which have significant prior information about the presence of a subset of individuals in D^s. More precisely, we will suppose that $\mathcal{A}_s$ knows with certainty that some individuals are in D^s. This means that $P_{min} = 0$ since for certain database $D' \in \mathcal{U}$ we have $\mathcal{B}_{\emptyset}^{\mathcal{A}_s}[D' = D^s] = 0$. Moreover, we suppose that, before interacting with the database, $\mathcal{A}_s$'s best confidence on the set of identities present in the database cannot be larger than 1/10. Formally, this means that $P_{max} = 1/10$ since there exists $D' \in \mathcal{U}$ such that $\mathcal{B}_{\emptyset}^{\mathcal{A}_s}[D' = D^s] = 1/10$.

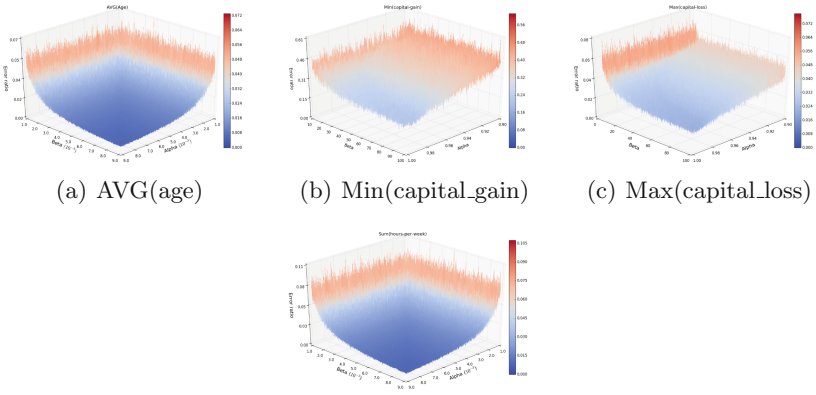
Table 2. Used aggregate queries and their identifiability sensitivity

Attribute	Query	Identifiability sensitivity (Θ)
Age	Average	27×10^{-4}
Capital-gain	Min	114
Capital-loss	Max	445
Hours-per-week	Sum	99

Table 2 shows the set of aggregate queries that we used in the evaluation of our model. For each query, we give the attribute over which it is executed and its corresponding identifiability sensitivity value. We note that counting queries are not considered in this evaluation since by definition, they have an identifiability sensitivity equal to zero which means that revealing the exact result of a counting query performed over the database D^s will not disclose any information to the adversary about the content of the database D^s.

Figure 1 illustrates the error ratio included in the differentially identifiable result of each query when the parameters α and β are varied and when the weak adversary $\mathcal{A}_w$ is considered. The different plots show, first, that the smaller the values of α and β (i.e., the higher the desired privacy), more noise are included in the response. Second, according to Figs. 1(a) and (d), our model provides a very good compromise between privacy and query response precision. For example, for the *Sum(hours_per_week)* query (Fig. 1(d)), our (α, β)-DI construction reduces the error rate to 9×10^{-3} for $\alpha = \beta = 8 \times 10^{-3}$. Third, although the high identifiability sensitivity of the queries *Min(capital-gain)* (Fig. 1(b)) and *Max(capital-loss)* (Fig. 1(c)), our (α, β)-DI construction provides an acceptable compromise between privacy and query response precision. As an example, our model provides an $(0.5, 1)$-differentially identifiable answer for the query *Max(capital_loss)* (Fig. 1(c)) with an error rate of 0.1.

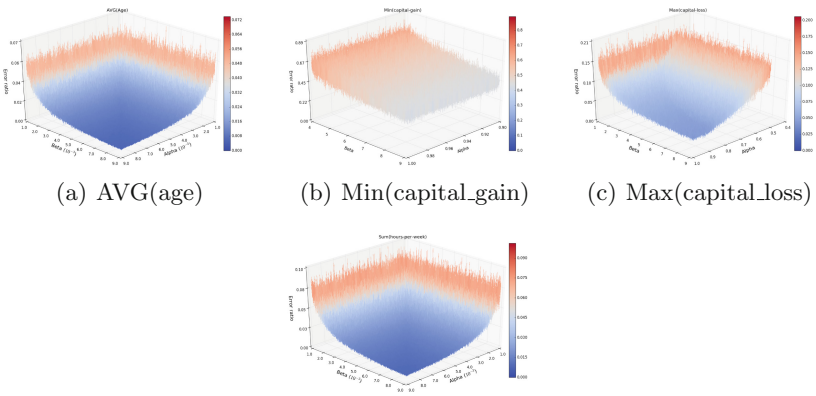
When considering the strong adversary $\mathcal{A}_s$ (Fig. 2), our construction still provides a very close privacy/utility trade-off compared to the one provided when the weak adversary $\mathcal{A}_w$ is considered, except for the query



(a) AVG(age) (b) Min(capital_gain) (c) Max(capital_loss)

(d) Sum(hours_per_week)

Fig. 1. Noise ratio for the adversary $\mathcal{A}_w$



(a) AVG(age) (b) Min(capital_gain) (c) Max(capital_loss)

(d) Sum(hours_per_week)

Fig. 2. Noise ratio for the adversary $\mathcal{A}_s$

Min(capital gain) (Fig. 2(b)). For this query, our construction provides answers with little bit more noise compared to answers provided when the weak adversary $\mathcal{A}_w$ is considered (Fig. 1(b)). This mainly caused by the reduced range of β's values (i.e., α should be less that 9 according to Theorem 3) that can be used without letting $\mathcal{A}_s$ be 100% sure about the content of the database D^s.

9 Related Work

Several privacy definitions have been proposed in last two decades. The most sticking out ones are *k-anonymity* [14], *l-diversity* [13], and *t-closeness* [10]. Dwork pointed out their weaknesses in [4] and argues that privacy problems

should be considered in a more formal way. Following this reflexion, the notion of DP was proposed in [3] and several approaches for satisfying it were developed to support low sensitive queries such as counting, mean, and median queries. Several relaxations of the original DP model have been proposed in order to make DP more efficient for high sensitive queries. (ϵ, δ)-DP was proposed in [5] by introducing new parameter δ that will be used to upper bound the probability that ϵ-DP is not satisfied. Generic DP is a generalization of the DP model proposed in [7]. It allows more flexible definitions for neighboring databases and conditions that the model should satisfy.

In [8], authors showed that for DP, it is not possible to ensure an acceptable privacy/utility compromise without making assumptions about the manner with which the data are generated. In this paper, we provide similar result by showing no possible acceptable privacy/utility compromise can be provided for our DI model without making assumptions about the adversary prior knowledge.

Cormode showed in [2] that DP is not useful for preventing inferential disclosure by demonstrating that one can use differentially private outputs to infer sensitive information with non-trivial accuracy. Lee and Clifton [9] argued that the parameter ϵ used in DP limits only how much one individual can affect the resulting model. It cannot be used to limit how much information is revealed about an individual. They then propose ρ-DI which captures membership disclosure under very specific adversarial background knowledge that we believe seldom satisfied in the real environments. Machanavajjhala et al. [12] proposed a model called ϵ-privacy aiming to limit the impact that one entity can have on the belief of the adversary. Unfortunately ϵ-privacy does not support interactive and adaptive data querying. Membership Privacy [11] proposed a model that uses Bayesian inference to bound the probability of identifying an individual in the database. However, the proposed model fails to bound the probability that an adversary figure out the set of individuals in a database.

10 Conclusion

This paper presents the new differential identifiability model allowing to bound the quantity of disclosed information about the presence of an individual in a database while considering an adversary with arbitrary prior knowledge. We showed that our proposed model can be satisfied using the general Laplace noise addition mechanism used traditionally in differential privacy. We proved that there is a direct connection between our (α, β)-differential identifiability and ϵ-differential privacy, and we showed through a set of experimentations that our model provides a good privacy/utility trade-off for most aggregate queries.

References

1. UCI machine learning repository. https://archive.ics.uci.edu/ml/index.php. Accessed 10 Apr 2018
2. Cormode, G.: Personal privacy vs population privacy: Learning to attack anonymization. In: Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2011, pp. 1253–1261. ACM, New York (2011). https://doi.org/10.1145/2020408.2020598
3. Dwork, C.: Differential privacy. In: Bugliesi, M., Preneel, B., Sassone, V., Wegener, I. (eds.) ICALP 2006. LNCS, vol. 4052, pp. 1–12. Springer, Heidelberg (2006). https://doi.org/10.1007/11787006_1
4. Dwork, C.: An ad omnia approach to defining and achieving private data analysis. In: Bonchi, F., Ferrari, E., Malin, B., Saygin, Y. (eds.) PInKDD 2007. LNCS, vol. 4890, pp. 1–13. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-78478-4_1
5. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Halevi, S., Rabin, T. (eds.) TCC 2006. LNCS, vol. 3876, pp. 265–284. Springer, Heidelberg (2006). https://doi.org/10.1007/11681878_14
6. Dwork, C., Rothblum, G.N., Vadhan, S.: Boosting and differential privacy. In: 2010 51st Annual IEEE Symposium on Foundations of Computer Science (FOCS), pp. 51–60. IEEE (2010)
7. Kifer, D., Lin, B.R.: Towards an axiomatization of statistical privacy and utility. In: Proceedings of the Twenty-Ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 147–158. ACM (2010)
8. Kifer, D., Machanavajjhala, A.: No free lunch in data privacy. In: Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data, SIGMOD 2011, pp. 193–204. ACM, New York (2011). https://doi.org/10.1145/1989323.1989345
9. Lee, J., Clifton, C.: Differential identifiability. In: Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD 2012, pp. 1041–1049. ACM, New York (2012). https://doi.org/10.1145/2339530.2339695
10. Li, N., Li, T., Venkatasubramanian, S.: t-closeness: privacy beyond k-anonymity and l-diversity. In: IEEE 23rd International Conference on Data Engineering, ICDE 2007, pp. 106–115. IEEE (2007)
11. Li, N., Qardaji, W., Su, D., Wu, Y., Yang, W.: Membership privacy: a unifying framework for privacy definitions. In: Proceedings of the 2013 ACM SIGSAC Conference on Computer & #38; Communications Security, CCS 2013, pp. 889–900. ACM, New York (2013). https://doi.org/10.1145/2508859.2516686
12. Machanavajjhala, A., Gehrke, J., Götz, M.: Data publishing against realistic adversaries. Proc. VLDB Endow. **2**(1), 790–801 (2009). https://doi.org/10.14778/1687627.1687717
13. Machanavajjhala, A., Gehrke, J., Kifer, D., Venkitasubramaniam, M.: l-diversity: privacy beyond k-anonymity. In: Proceedings of the 22nd International Conference on Data Engineering, ICDE 2006, pp. 24–24. IEEE (2006)
14. Sweeney, L.: k-anonymity: a model for protecting privacy. Int. J. Uncertain. Fuzziness Knowl.-Based Syst. **10**(05), 557–570 (2002)