The Semantic Discrimination Rate Metric for Privacy Measurements which Questions the Benefit of T-closeness over L-diversity

Louis Philippe Sondeck[1], Maryline Laurent[2] and Vincent Frey[3]
[1]*Orange Labs, Telecom SudParis, Cesson Sevigne, France*
[2]*SAMOVAR, Telecom SudParis, Paris-Saclay University, CNRS, Evry, France*
[3]*Orange Labs, Cesson Sevigne, France*
{*louisphilippe.sondeck, vincent.frey*}*@orange.com, maryline.laurent@telecom-sudparis.eu*

Keywords: anonymity metric, Semantic Discrimination Rate, Discrimination Rate, identifiability, k-anonymity, t-closeness, l-diversity.

Abstract: After a brief description of k-anonymity, l-diversity and t-closeness techniques, the paper presents the Discrimination Rate (DR) as a new metric based on information theory for measuring the privacy level of any anonymization technique. As far as we know, the DR is the first approach supporting fine grained privacy measurement down to attribute's values. Increased with the semantic dimension, the resulting semantic DR (SeDR) enables to: (1) tackle anonymity measurements from the attacker's perspective, (2) prove that t-closeness can give lower privacy protection than l-diversity.

1 INTRODUCTION

An increasing number of services are relying on users' data sharing. City services, weather, mapping... are services based on Open Data operated by many public or private platforms. Several efforts are made by publishers for anonymizing a dataset prior to publishing, based on anonymization techniques (e.g. Differential Privacy, k-anonymity, l-diversity, t-closeness).

So far, there have been no convincing metrics defined for accurately quantifying the anonymity level of a given dataset. The differential privacy metrics (Dwork et al., 2006) rely on an "ε" parameter to capture the anonymity degree. However, it remains specific to data being anonymized in a differential privacy manner; it is difficult to apply to other types of anonymization (Lee and Clifton, 2011) (Hsu et al., 2014); it suffers from the lack of accuracy due to the noise being roughly added for anonymization thus, making the evaluation of the risk of re-identification difficult (Lee and Clifton, 2011). The theoretic *Distortion Rate* based metrics (Sankar et al., 2013) (Rebollo-Monedero et al., 2010), including also Mutual Information metric (Salamatian et al., 2013) (Rebollo-Monedero et al., 2010) (Makhdoumi and Fawaz, 2013), provides more generic measurements, but does not take into account the *semantic dimension* and measurement of *combined key attributes*, while these aspects are both shown to be critical in privacy measurement (refer respectively to Sections 5.1, 6 and 4).

Furthermore, only few works provide measurements according to some threat models (identity and attribute attacks), even if it has been clearly identified that the anonymity problem comes from the correlation between *key attributes* (attributes that are transformed to protect the subject's identity; e.g. 35510 transformed into 35***) and *sensitive attributes* (attributes which sensitive value is of interest to be revealed; e.g. health, salary...).

Contributions

This paper proposes the Semantic Discrimination Rate (SeDR), an attribute-centric metric that enables to quantify anonymity by measuring the identification capability of attributes and that takes into account semantic considerations. The SeDR is based on the DR metric (Sondeck et al., 2017) which objective is to quantify the identification capability by measuring how much attributes can refine a given anonymity set. The DR gives a score scaling from 0 to 1 where an identifier with a DR equal to 1 reduces the anonymity set to a single user. The SeDR adds the semantic dimension to the DR and measures the identification capability according to subsets of subjects, instead of single subjects within an anonymity set, thus, leading to the definition of some semantic domains (cf. Sec-

Table 1: Generalization Table.

	ZIP Code	ZIP Code*	Age	Age*	Age**
1	35567	355**	22	2*	≤ 40
2	35502	355**	22	2*	≤ 40
3	35560	355**	22	2*	≤ 40
4	35817	3581*	45	≥ 40	≥ 40
5	35810	3581*	63	≥ 40	≥ 40
6	35812	3581*	40	≥ 40	≥ 40
7	35502	355**	35	3*	≤ 40
8	35568	355**	35	3*	≤ 40
9	35505	355**	32	3*	≤ 40

tion 5.2).

The SeDR metric enables to:

- **Tackle anonymity measurements from the attacker's perspective** by computing how much information is gained after applying a given attack. Any existing attacks (identity attack and attribute attacks) can be evaluated.
- **Prove that t-closeness as a metric, is not as protective as claimed by the authors** and that, depending on the semantic considerations, **t-closeness can be worse than l-diversity**.

The rest of the paper is organized as follows: Section 2 gives some background on k-anonymity-like privacy techniques and their related attacks, Section 3 presents our critical analysis on t-closeness technique, and points out its irrelevance to quantify privacy. After introducing our Discrimination Rate concepts in Section 4, Section 5 describes the semantic empowered Discrimination Rate together with our *semantic domain* definitions. Then Section 6 presents our measurements and comparison of l-diversity vs t-closeness. Finally Section 7 gives our conclusions.

2 BACKGROUND ON ANONYMIZATION TECHNIQUES

Anonymization techniques aim to protect user's confidential attributes from released datasets while ensuring usability of data. The k-anonymity model, introduced by Samarati (Samarati, 2001) is one of the most notable models to achieve data anonymization. From the k-anonymity point of view, datasets contain three types of data: identifiers (*e.g. social security numbers...*), key-attributes/quasi-identifiers (*e.g. ZIP Code, Age*) and confidential/sensitive attributes (*e.g. Salary, Diseases, ...*). Due to its limitations, many improvements have been proposed. After a brief description of the k-anonymity goals and limitations, we present two of its most important improvements: l-diversity and t-closeness.

2.1 K-anonymity to Mitigate Identity Disclosure

The goal of k-anonymity is to protect sensitive attributes by reducing the link between those sensitive attributes and subjects to whom they belong; mitigating hence the **identity disclosure attack**. For that purpose, k-anonymity makes use of different operations among which *generalization* and *suppression* (cf. Table 1) that are applied on key-attributes and identifiers to reduce their capacity to identify a subject. Generalization consists in replacing every key attribute value by more generic values (e.g. Age 22 replaced with Age 2*) and suppression is applied on identifiers (attributes that can not be generalized and that strongly identify subjects). For k-anonymity, these operations are applied in such a way that every generalized attribute corresponds to at least k subjects. The maximal set of subjects is called *the equivalence class*.

However, if k-anonymity mitigates *identity disclosure*, it fails mitigating *attributes disclosure* attacks, especially: *homogeneity* and *backgroung knowledge attacks* (Machanavajjhala et al., 2007).

Table 2: 3-anonymity Table (Disease).

	Age*	Disease
1	2*	lung cancer
2	2*	lung cancer
3	2*	lung cancer
4	≥ 40	stomach cancer
5	≥ 40	diabetes
6	≥ 40	flu
7	3*	aids
8	3*	aids
9	3*	diabetes

2.2 L-diversity to Mitigate Homogeneity and Background Knowledge Attacks

The l-diversity technique (Machanavajjhala et al., 2007) has been introduced to counteract:

1. **homogeneity attack** which refers to the knowledge gained by correlating *key attributes* and *sensitive attributes* within the table. For instance, in the 3-anonymity table (Table 2), the key attribute value "2*" completely corresponds to the sensitive value "lung cancer"; as such, an attacker only

needs to know that the subject is twenties to link him to the disease "lung cancer".

2. **backgroung knowledge attack** which uses external data to improve subjects identification. For example, in the 3-anonymity table (Table 2), if we consider the third equivalence class (with key attribute value "3*"), the correspondence between "Age*" and "Disease" is not complete, and an attacker will therefore need external information (for example that the subject is less likely to have *diabetes*) to link him to *aids*.

To counteract those attacks, l-diversity adds the restriction that all the sensitive attributes should have at least l "*well represented*" values.

Definition 1. *(The l-diversity principle)*
An equivalence class is said to have l-diversity if there are at least l "well-represented" values for the sensitive attribute. A table is said to have l-diversity if every equivalence class of the table has l-diversity.

This restriction enables to reduce the correlation between key attribute values and sensitive attribute values and helps to mitigate both *homogeneity* and *background knowledge* attacks. For instance, in the 3-diverse table (Table 3), the diversity of sensitive values ("Disease", "Salary") in each class prevents from correlating key attributes and sensitive attributes.

Despite these improvements, l-diversity has been proved (Li et al., 2007) to be inefficient to counteract *attribute disclosure* attacks as it does not take into account the semantic of attributes. This flaw is depicted through two main attacks: *skewness attack* and *similarity attack*.

Table 3: A 3-diverse Table.

	ZIP Code*	Age*	Salary	Disease
1	355**	2*	4K	colon cancer
2	355**	2*	5K	stomach cancer
3	355**	2*	6K	lung cancer
4	3581*	≥ 40	7K	stomach cancer
5	3581*	≥ 40	12K	diabetes
6	3581*	≥ 40	9K	aids
7	355**	3*	8K	aids
8	355**	3*	10K	flu
9	355**	3*	11K	lung cancer

2.3 T-closeness to Mitigate Skewness and Similarity Attacks

The t-closeness technique (Li et al., 2007) has been introduced to counteract:

1. the **skewness attack** which is based on the skewness between the distribution of sensitive attribute values within the original table and the distribution within equivalence classes. Let us consider the following example:

   Example 1. *Suppose we have an original skewness table containing data of 1000 patients with and without cancer; the key attributes are "Age", "ZIP Code" and the sensitive attribute is "Cancer"; and "Cancer" can have two values "Yes" or "No". Suppose we have only 10 "Yes" in the table. A 2-diverse table (formed by equivalence class of 2 subjects) would provide 50% probability of having cancer for each subject within classes instead of 10/1000% in the original table and then, an information gain from the anonymized table.*

2. the **semantic attack** which relies on similarity between sensitive values. Indeed, when the sensitive attribute values are distinct but semantically similar (e.g. "stomach cancer", "colon cancer", "lung cancer"), the *similarity attack* can occur. For example, let us consider the first class of the 3-diverse table (Table 3) with key value "2*". The value "2*" corresponds to the subset of sensitive values: $\{4K, 5K, 6K\}$. Even if those values are diversified, they still contain semantic information as an attacker can infer that all subjects who are twenties have low incomes.

To overcome *skewness* and *similarity attacks*, the t-closeness principle was proposed by Li and al (Li et al., 2007) and states that:

Definition 2. *(The t-closeness principle)*
An equivalence class is said to have t-closeness if the distance between the distribution of a sensitive attribute in this class and the distribution of the attribute in the original table is no more than a threshold t. A table is said to have t-closeness if all equivalence classes have t-closeness.

The t-closeness metric measures therefore the distance between distributions of sensitive values within classes and within the original table to ensure it does not exceed a given threshold. This property is claimed to mitigate both *skewness* and *similarity attacks*.

3 T-CLOSENESS LIMITATIONS AND INABILITY TO QUANTIFY PRIVACY

(Domingo-Ferrer and Torra, 2008) identified the following limitations on the t-closeness metric:

- t-closeness does not provide a computational procedure;
- If such a procedure was available, it would greatly damage the utility of data. Indeed, by definition, t-closeness aims to destroy the correlations between key attributes and sensitive attributes and this, for any combination of key attribute values.

Additionally, we identified another criticism as t-closeness, taken as a metric, does not measure the effective disclosure risk but instead the accomplishment of the anonymization process. Indeed, two attributes with the same t-closeness measurement can have different privacy levels. This comes directly from the definition (cf. section 2.3): *t-closeness computes the distance between the distribution of a sensitive attribute within classes and the distribution of attributes in the original table*. That is, a t-closeness measurement relies on the distribution of attributes in the original table; hence, if two attributes have different distributions in the original table, they can have the same t-closeness measurement, but not the same disclosure risk.

Another concern is that the t-closeness measurement relies on a pre-built hierarchy of attribute values that can differ according to the attacker's model. Indeed, in order to compute a t-closeness measurement, the attribute values should be classified and the measurement relies on this classification (Li et al., 2007) which is subjective. We give more details about semantical subjectivity and attacker's model in Section 5.1.

Finally, there is no direct link between t-closeness measurements and the re-identification process. Indeed, t-closeness computes a distance between distribution sets and the relationship with information gain or loss is unclear as acknowledged by the authors (Li et al., 2007): *"...the relationship between the value t and information gain is unclear"*.

4 BASIC CONCEPTS OF DISCRIMINATION RATE (DR)

The Discrimination Rate metric (Sondeck et al., 2017) refers to the entropy and conditional entropy measurements (Shannon, 2001). The DR objective is to quantify how much an attribute is able to refine an *anonymity set*; the maximum refinement leading to one subject. The *anonymity set* is characterized by a given *sensitive attribute* and the DR measures how much the *key attribute* refines the set of values of this sensitive attribute. More precisely, we consider *key attributes* as discrete random variables (d.r.v.) and the anonymity set as the set of outcomes of another d.r.v. Let us consider 2 d.r.v. X and Y where Y is the attribute we wish to measure the identification capacity and X, the attribute which the set of outcomes is our anonymity set. We then want to compute the amount of information carried by Y according to the refinement of the set of outcomes of X. For that purpose, we consider the amount of uncertainty carried by X (the entropy of X, $H(X)$) as our initial state and compute the entropy of X conditioned on Y ($H(X|Y)$), as we wish to measure the effect of Y on X. This quantity represents the remaining uncertainty within X, after Y is divulged. In order to compute the amount of information carried by Y according to X, we need to subtract that quantity from $H(X)$ and thus we obtain $H(X) - H(X|Y)$, which is the effective amount of identification information carried by attribute Y according to that anonymity set. Finally, we divide that quantity by $H(X)$ to normalize the value.

Table 4: Original Data Table (Salary/Disease).

	ZIP Code	Age	Salary	Disease
1	35567	22	4K	colon cancer
2	35502	22	5K	stomach cancer
3	35560	22	6K	lung cancer
4	35817	45	7K	stomach cancer
5	35810	63	12K	diabetes
6	35812	40	9K	aids
7	35502	35	8K	aids
8	35568	35	10K	flu
9	35505	32	11K	lung cancer

4.1 SDR and CDR Definitions

Let us propose the following definitions for ***Simple Discrimination Rate*** (single key attributes measurement) and ***Combined Discrimination Rate*** (multiple *key attributes* measurement).

Definition 3. *(Simple Discrimination Rate)*
Let $\boldsymbol{X}$ *and* $\boldsymbol{Y}$ *be two d.r.v. The* ***Simple Discrimination Rate*** *of the key attribute* $\boldsymbol{Y}$ *relatively to sensitive attribute* $\boldsymbol{X}$*, is the capacity of the key attribute* $\boldsymbol{Y}$ *to refine the set of outcomes of the sensitive attribute* $\boldsymbol{X}$ *and is computed as follows:*

$$DR_X(Y) = \frac{H(X) - H(X|Y)}{H(X)} = 1 - \frac{H(X|Y)}{H(X)} \quad (1)$$

Definition 4. *(Combined Discrimination Rate)*
Let $\boldsymbol{X}$*,* $Y_1, ..., Y_n$ *be d.r.v. The* ***Combined Discrimination Rate*** *of key attributes* $Y_1, Y_2, ..., Y_n$ *relatively to the sensitive attribute* $\boldsymbol{X}$*, is the capacity of the combination of key attributes* $Y_1, ..., Y_n$ *to refine the set of*

outcomes of the sensitive attribute **X** *and is computed as follows:*

$$DR_X(Y_1,Y_2,...,Y_n) = 1 - \frac{H(X|Y_1,Y_2,...,Y_n)}{H(X)} \quad (2)$$

We can deduce that $0 \leq DR_X(Y_1,Y_2,...,Y_n) \leq 1$.

Also we deduce that, the SDR is a particular case of the CDR.

Note that, $DR_X(Y_1,Y_2,...,Y_n) = 1$ in case $(Y_1,Y_2,...,Y_n)$ is an **identifier**. Indeed, in this case, the remaining information within X is null ($H(X|Y_1,Y_2,...,Y_n) = 0$).

4.2 Illustration of DR over Table 4

Let us consider Table 4. We can compute the DR of key attribute *Age* (and its values) with respect to the sensitive attribute *Disease*.

The computation steps for $DR_X(Y)$ with X = Disease, Y = Age are as follows:

$$DR_X(Y) = 1 - \frac{H(X|Y)}{H(X)} \quad (3)$$

$$= 1 - \frac{1/3\log_2(1/3) + 2/9\log_2(1/2)}{3/9\log_2(1/9) + 6/9\log_2(2/9)} \quad (4)$$

$$= 0.70 \quad (5)$$

For $H(X|Y)$, attribute *Age* can take 6 values - 22, 32, 35, 40, 45, 63 - which help to reduce the outputs of Disease to subsets of 3, 1, 2, 1, 1 and 1 respectively, corresponding to 1/3, 1/9, 2/9, 1/9, 1/9 and 1/9 of the original set respectively. The specific conditional entropies are then: $H(X|Y = 22) = -log_2(1/3), H(X|Y = 32) = 0, H(X|Y = 35) = -log_2(1/2), H(X|Y = 40) = 0, H(X|Y = 45) = 0$ and $H(X|Y = 63) = 0$. $H(X|Y)$ is therefore the sum of $-1/3\ log_2(1/3)$ and $-2/9\ log_2(2/9)$.

We can also compute the DR of a combination of key attributes *Age* and *ZIP Code* using *Disease* as *sensitive attribute*. The result is depicted in Table 5.

As we can see, the DR provides granular measurements and can be performed either on attributes or on attribute's values.

4.3 Risk Measurement of Anonymization Techniques

The DR supports anonymity measurements by quantifying the amount of knowledge gained by an attacker after applying a given attack on a system. The attacker's knowledge in this case is key attributes and the DR measures how much from those key attributes, an attacker is able to link a subject to a sensitive attribute. The DR enables therefore to measure the privacy risk according to *identity disclosure*, *background knowledge* and *homogeneity attacks*. When applied on Tables 1 and 3, we obtain the results depicted in Tables 6, 7 and 8.

Table 5: Discrimination Rate over the Original Data Table 4.

X	Y	$DR_X(Y)$
Disease	22	0.79
Disease	32	1
Disease	35	0.91
Disease	40	1
Disease	45	1
Disease	63	1
Disease	Age	0.70
Disease	ZIP Code & Age	1

The ***identity disclosure attack*** refers to the generalization mechanism (cf. Table 1) and the anonymity measurement process consists in measuring the refinement capacity of generalized attributes over original attributes (cf. Table 6).

The ***homogeneity attack*** refers to the correspondence between sensitive and key attributes (Domingo-Ferrer and Torra, 2008); the anonymity measurement process consists then in measuring the refinement capacity of key attributes over sensitive attributes (Table 7).

Finally, as the ***background knowledge*** attack relies on the *homogeneity attack* (cf. Section 2.2); we compute the resistance measurement to combine homogeneity and background knowledge attacks from the result of the *homogeneity attack* measurement (Table 8).

We observe from Table 7 that the l-diverse table (Table 3) provides more resistance than the k-anonymity table (Table 2) to the *homogeneity attack*, as the DR (the capacity of the attacker) is lower for the l-diverse table (0.52 vs 0.36). However concerning the *identity attack*, both techniques provide the same resistance for attribute ZIP Code but not for attribute Age (0.31 vs 0.38) as the generalization process of Age is different in each case (Table 1).

Table 6: Risk measurement for identity disclosure (Table 1).

X	Y	$DR_X(Y)$
ZIP Code	ZIP Code*	0.31
Age	Age*	0.66
Age	Age**	0.38

Table 7: Risk measurement for homogeneity attack (Table 3).

X	Y	$DR_X(Y)$
k-anonymity Table		
Disease	2*	1
Disease	≥ 40	0.70
Disease	3*	0.83
Disease	Age*	0.52
l-diverse Table		
Disease	2*	0.78
Disease	≥ 40	0.78
Disease	3*	0.78
Disease	Age*	0.36

Table 8: Resistance measurement to combine homogeneity and background knowledge attacks (Table 7).

X	Y	1 - $DR_X(Y)$
k-anonymity Table		
Disease	2*	0
Disease	≥ 40	0.30
Disease	3*	0.17
Disease	Age*	0.48
l-diverse Table		
Disease	2*	0.22
Disease	≥ 40	0.22
Disease	3*	0.22
Disease	Age*	0.64

4.4 Inability for Basic DR to Measure Semantic

The DR does not take into account the semantic behind attribute values. For example, the t-closeness instantiation (Table 9) can provide more semantic privacy than the l-diverse instantiation (Table 3). Indeed, from key attribute value 355** in the l-diverse instantiation, an attacker can infer with 50% success that the user's salary is low (between 4K and 6K). This reflects the semantic aspect of attributes which is not taken into account by l-diversity, but is included in the t-closeness approach (Table 9).

5 SEMANTIC EMPOWERED DISCRIMINATION RATE

This section presents the semantic DR (SeDR) that supports semantic measurements. After arguing that the semantic measurement is a subjective measurement, we define our *semantic domains* that permit to capture this subjectivity. Then, we present our semantic Discrimination Rate (SeDR), along with illustration of SeDR computation.

Table 9: An 0.167-closeness w.r.t. Salary and 0.278-closeness w.r.t. Disease.

	ZIP Code*	Age**	Salary	Disease
1	3556*	≤ 40	4K	colon cancer
3	3556*	≤ 40	6K	lung cancer
8	3556*	≤ 40	10K	flu
4	3581*	≥ 40	7K	stomach cancer
5	3581*	≥ 40	12K	diabetes
6	3581*	≥ 40	9K	aids
2	3550*	≤ 40	5K	stomach cancer
7	3550*	≤ 40	8K	aids
9	3550*	≤ 40	11K	lung cancer

5.1 Semantic as a Subjective Measurement with Regard to Attacker's Model

The term *semantic* refers to the meaning of attributes or attributes values, which is fully subjective. Indeed, an attribute's value can have different meanings according to the **attacker's model**. The attacker's model here refers to **the attacker's goal** and **previous knowledge to achieve this goal**. The attacker's model can be specified according to the categories an attacker is classifying the sensitive values.

For instance, let us consider the following three attacker's models over Tables 3 and 9 where the **attacker's knowledge** is made of the *key attributes* Age* and ZIP Code*:

1. The attacker needs to know the exact Salary's value of a subject;
2. The attacker needs to know which Salary category the subject belongs to: low (4K-6K), medium (7K-9K) or high (10K-12K);
3. The attacker wants to link the subject to one of the following Salary's subsets: {4K, 6K, 10K}, {7K, 12K, 9K} and {5K, 8K, 11K}.

For the **attacker's model 1**, the attacker needs to know the exact value. The set of categories contains therefore single values: {4K}, {5K},...,{12K}. Hence, the similarity between values is not taken into account for this model as the attacker is interested in single values. Therefore, the t-closeness instantiation provides the same semantic security than the l-diverse instantiation and the current DR is enough to compute the disclosure risk for both techniques.

For the **attacker's model 2**, the attacker's needs are not as restrictive as for attacker's model 1 as the attacker only wants to know the average salary. For this attacker's model, the similarity between values

is worthwhile and is a privacy risk. As such, the t-closeness metric and the l-diversity metric do not provide the same semantic security, and adaptation of the Discrimination Rate is therefore necessary to measure that disclosure risk.

The **attacker's model 3** is somewhat interesting as it refers to the subsets of Salaries within classes of the t-closeness table (Table 9). As shown in Section 6, for this model, the t-closeness instantiation (Table 9) is proved to be worse than the l-diverse instantiation (Table 3).

5.2 Semantic Domain Definitions

This section gives our definitions about *semantic partition* and *semantic domains*, which help to capture the subjectivity of semantic based on attacker's models of section 5.1. These definitions are illustrated through an example.

Definition 5. *(Semantic Partition)*
*Let X be an attribute and $\mathcal{X}$ be the set of all possible values of X. A **Semantic Partition** of X is a partition of $\mathcal{X}$ according to a given attacker's model.*

Definition 6. *(Semantic Domain)*
*A **Semantic Domain** is an element of a Semantic Partition.*

The semantic domains refer to the classification of sensitive values with respect to their sensitivity similarity as identified by the attacker's model (Section 5.1). We refer to the set of *semantic domains* as the *semantic partition*. Indeed, for the purpose of this work, we suppose the semantic domains to be disjoint and the *semantic partition* to be a partition[1] of the set of sensitive values.

The corresponding *semantic partitions* of attacker's models in Section 5.1 are:

- Attacker's model 1: SP_1 = {{4K}, {5K},..., {12K}}.
- Attacker's model 2: SP_2 = {{4K, 5K, 6K}, {7K, 8K, 9K}, {10K, 11K, 12K}}.
- Attacker's model 3: SP_3 = {{4K, 6K, 10K}, {7K, 12K, 9K} {5K, 8K, 11K}}.

Note that, the methodology for getting a *semantic partition* is out of scope of this paper. Our objective is only to show how subjective are the anonymity measurements and how semantic can be introduced in our DR metric. There are however some works (Erola et al., 2010) (Abril et al., 2010) proposing a way to cluster values according to their semantic similarity, and therefore, a way to build semantic partitions.

[1]Partition of a set A: is a subdivision of A into subsets that are disjoint, non-empty and which the union equals to A.

5.3 SeDR as DR with Semantic Measurement

To cope with the DR's inability to handle semantic dimension as explained in Section 4.4, this section defines the Semantic DR (SeDR) which supports semantic measurements based on the *semantic domains* (Section 5.2).

Thanks to the *semantic domains*, the SeDR has the objective to measure how much an attacker provided with key attributes, is able to refine the set of semantic domains (the *semantic partition*) instead of the set of single values. As such, with SeDR, it is possible to know the attacker's capacity to infer **subsets** of user's sensitive values from a key attribute value.

Before applying the SeDR, we should first transform the sensitive attribute X according to a given semantic partition SP. Let sX be the result of the transformation. We define the **semantic partition transformation** f_{SP} as follows:

$$f_{SP} : X \rightarrow sX. \tag{6}$$

The SeDR is then defined as follows:

Definition 7. ***(Semantic Discrimination Rate)***
*Let X be a sensitive attribute and SP a semantic partition of X. Let $sX = f_{SP}(X)$ and $Y_1,...,Y_n$ be a set of key attributes. The **Semantic Discrimination Rate** (SeDR) of $Y_1,...,Y_n$ relatively to X is the DR of $Y_1,...,Y_n$ relatively to sX and is computed as follows:*

$$SeDR_X(Y_1,Y_2,...,Y_n) = DR_{sX}(Y_1,Y_2,...,Y_n) \tag{7}$$

Therefore, the original DR is a particular case of the SeDR with a *semantic partition* composed of single sensitive values.

5.4 Illustration of the SeDR Computation and Comparison with the DR

Let us illustrate the SeDR over the original data Table 4 with the *semantic partition* SP_4 = {{diabetes, flu, aids}, {colon cancer, lung cancer, stomach cancer}}.

The semantic partition transformation f_{SP} is applied on X by replacing the set of values $\mathcal{X}$ (of X) by the set of values $s\mathcal{X}$ (of sX). For example, for *sensitive attribute* "Disease", we transform $\mathcal{X}$ = {*colon cancer, stomach cancer, lung cancer, stomach cancer, diabetes, aids, aids, flu, lung cancer*} using SP_4 = {{*diabetes, flu, aids*}, {*colon cancer, lung cancer,*

Table 10: Semantic DR in Table 4.

X	Y	$DR_X(Y)$
Disease	22	1
Disease	32	1
Disease	35	1
Disease	40	1
Disease	45	1
Disease	63	1
Disease	Age	1

stomach cancer}} into $s\mathcal{X}$ = {*cancer, cancer, cancer, cancer, other disease, other disease, other disease, other disease, cancer*}.

When applying the previous transformation on the sensitive attribute Disease in Table 4, and computing the SeDR according to the key attribute "Age*", we obtain the results in Table 10.

As shown in Table 10 vs Table 5, the SeDR is able to extract more information from the same database than the non-semantic DR, as higher values are obtained in Table 10. For instance, for key attribute value 22, the SeDR is 1 compared to 0.79 for the DR, as this key attribute fully corresponds to the semantic domain {colon cancer, lung cancer, stomach cancer} of the original data table (Table 4).

6 MEASUREMENT AND COMPARISON OF L-DIVERSITY VS T-CLOSENESS WITH SeDR

This section shows first how the semantic attacks - skewness attack and the similarity attack (Section 2.3) - can be measured with either the DR or the SeDR. Then it proves through the SeDR, for the similarity attack only, that t-closeness is not as privacy protective as claimed by the authors, and that it can provide lower privacy protection than l-diversity. Both t-closeness and l-diversity techniques are instantiated over the original data Table 4 to give Tables 3 and 9 respectively. Note that these tables are similar to the ones of the original paper related to the t-closeness metric (Li et al., 2007).

6.1 Skewness Attack - Measurement with DR

The original DR is enough to evaluate this attack as only the skewness between the original distribution of sensitive values and their distribution within equivalence classes needs to be measured. For explaining this measurement, let us consider Example 1 of Section 2.3. The objective of the attack is to improve the attacker's knowledge within the equivalence classes. As such, the DR enables to quantify how much information is gained by an attacker from equivalence classes, according to the original table.

Therefore, for evaluating this attack, we compute the difference between the DRs of the involved key attributes in the original table and in the equivalence classes. Based on the skewness table of Example 1 (Section 2.3), we compute the DR of key attributes "Age" and "ZIP Code" using "Cancer" as the *sensitive attribute* in the original table ($DR_{Cancer}(Age)$ & $DR_{Cancer}(ZIPCode)$) and the DR of the key attributes "Age*" and "ZIP Code*" within equivalence classes ($DR_{Cancer}(Age*)$ & $DR_{Cancer}(ZIPCode*)$). Finally the actual information gain related to *skewness attack* is:

- $DR_{cancer}(Age) - DR_{cancer}(Age*)$ for key attribute Age.
- $DR_{cancer}(ZIPCode) - DR_{cancer}(ZIPCode*)$ for key attribute ZIP Code.

This computation can also be performed on attribute's values instead of the attributes.

This evaluation through DR computation gives far more results than merely computing the ratio between probabilities (50% and 10/1000%), as the DR takes into account the correlation between key attributes and sensitive attributes and since the attacker's knowledge refers to key attributes, the DR quantifies the actual information gain.

6.2 Similarity Attack - Measurement with SeDR

The SeDR is computed to evaluate the similarity between values of sensitive attributes. The similarity between values is formalized through some defined *semantic partitions*.

We consider three *semantic partitions*; two partitions of "Salary" (according to the attacker's models 2 and 3, Section 5.2) and one partition of "Disease":

- SP_2 = {{4K, 5K, 6K}, {7K, 8K, 9K}, {10K, 11K, 12K}} for "Salary".
- SP_3 = {{4K, 6K, 10K}, {7K, 12K, 9K} {5K, 8K, 11K}}
- SP_4 = {{diabetes, flu, aids}, {colon cancer, lung cancer, stomach cancer}}

We then use these *semantic partitions* and each key attribute ("Age*" and "ZIP Code*") to compute

Table 11: Risk measurement for Tables 3 & 9 for the similarity attack using SP_2 as the semantic partition and Age* & ZIP Code* as key attributes.

X	Y	$SeDR_X(Y)$
3-diverse Table		
SP_2	2*	1
SP_2	≥ 40	0.81
SP_2	3*	0.81
SP_2	Age*	0.61
SP_2	355**	0.39
SP_2	3581*	0.81
SP_2	ZIP Code*	0.19
t-closeness Table		
SP_2	≤ 40	0.39
SP_2	≥ 40	0.81
SP_2	Age**	0.19
SP_2	3550*	0.67
SP_2	3581*	0.81
SP_2	3556*	0.81
SP_2	ZIP Code*	0.28

the SeDR for the l-diverse and the t-closeness instantiations (Tables 3 and 9). The results are depicted in Tables 11, 12 and 13.

Table 12: Risk measurement for Tables 3 & 9 for the similarity attack using SP_3 as the semantic partition and Age* & ZIP Code* as key attributes.

X	Y	$SeDR_X(Y)$
3-diverse Table		
SP_3	2*	0.81
SP_3	≥ 40	1
SP_3	3*	0.81
SP_3	Age*	0.61
SP_3	355**	0.58
SP_3	3581*	1
SP_3	ZIP Code*	0.58
t-closeness Table		
SP_3	≤ 40	0.58
SP_3	≥ 40	1
SP_3	Age**	0.58
SP_3	3550*	1
SP_3	3581*	1
SP_3	3556*	1
SP_3	ZIP Code*	1

6.3 Results Proving the Lower Privacy Protection of T-closeness vs L-diversity

As reported in Section 5.3, the SeDR can compute the refinement capacity of a given *key attribute* over a *semantic partition* of a *sensitive attribute* (Section 5.2). The *semantic partition* reflects the subjectivity related to the semantic interpretation. The semantic risk measurement consists therefore in measuring how much from a given *key attribute*, an attacker is able to refine the *semantic partition* of a *sensitive attribute*. The more an attacker is able to refine the *semantic partition*, the higher the related risk.

Each computation is therefore performed according to a given *key attribute* and a given *semantic partition*.

Tables 11, 12 and 13 depict the risk measurements related to the *similarity attack* performed over the l-diverse table (Table 3) and the t-closeness table (Table 9). The considered key attributes are ZIP Code* and Age* and their SeDR are computed over the semantic partitions SP_2, SP_3 (related to "Salary") and SP_4 (related to "Disease").

Hereafter, we prove that **the assertion that t-closeness is semantically more secure than l-diversity is wrong**:

1. Table 11 shows that **an attacker is more able to refine the semantic partition SP_2 within the t-closeness table based on key attribute ZIP Code* than with the l-diversity table**. ZIP Code* gives a SeDR of 0.28 for t-closeness vs 0.19 for l-diversity.
2. Table 12 proves that **the t-closeness instantiation is weaker than the l-diverse instantiation against the *similarity attack* for the semantic partition SP_3**. Based on attribute ZIP Code*, an attacker is able to completely refine the *semantic*

Table 13: Risk measurement for Tables 3 & 9 for the similarity attack using SP_4 as the semantic partition and Age* & ZIP Code* as key attributes.

X	Y	$SeDR_X(Y)$
3-diverse Table		
SP_4	2*	1
SP_4	≥ 40	0.69
SP_4	3*	0.69
SP_4	Age*	0.38
SP_4	355**	0.38
SP_4	3581*	0.69
SP_4	ZIP Code*	0.07
t-closeness Table		
SP_4	≤ 40	0.38
SP_4	≥ 40	0.69
SP_4	Age**	0.07
SP_4	3550*	0.69
SP_4	3581*	0.69
SP_4	3556*	0.69
SP_4	ZIP Code*	0.07

partition SP_3 (DR = 1), as the ZIP Code*'s values directly refer to the considered *semantic domains*.

3. Table 13 shows that **an attacker is more able to refine the semantic partition** SP_4 **with some key attribute ZIP Code* values (3550* and 3556***). For these two values, the computed SeDR is higher for the t-closeness instantiation (0.69 vs 0.38).

7 CONCLUSION

Data publishing promises significant progress for emergence and improvement of new services. However, to mitigate privacy leakages due to poor anonymization procedures, there is a strong need for publishers to have a practical and precise metric to assess the data anonymity level prior to publishing datasets. In this paper, we propose the Semantic Discrimination Rate which is a new practical metric for getting fine grained measurement of the anonymity level of an anonymized dataset. It enables to tackle the de-anonymization issue from the attacker's perspective, and to precisely compute the attacker's capacity according to any existing anonymity attacks. Illustration of that metric is given over some classical anonymization techniques (t-closeness and l-diversity), and proves that t-closeness is not as privacy protective as it was originally claimed to be as it can behave worse than l-diversity.

REFERENCES

Abril, D., Navarro-Arribas, G., and Torra, V. (2010). Towards semantic microaggregation of categorical data for confidential documents. In *International Conference on Modeling Decisions for Artificial Intelligence*, pages 266–276. Springer.

Domingo-Ferrer, J. and Torra, V. (2008). A critique of k-anonymity and some of its enhancements. In *Availability, Reliability and Security, 2008. ARES 08. Third International Conference on*, pages 990–993. IEEE.

Erola, A., Castellà-Roca, J., Navarro-Arribas, G., and Torra, V. (2010). Semantic microaggregation for the anonymization of query logs. In *International Conference on Privacy in Statistical Databases*, pages 127–137. Springer.

Hsu, J., Gaboardi, M., Haeberlen, A., Khanna, S., Narayan, A., Pierce, B. C., and Roth, A. (2014). Differential privacy: An economic method for choosing epsilon. In *Computer Security Foundations Symposium (CSF), 2014 IEEE 27th*, pages 398–410. IEEE.

Lee, J. and Clifton, C. (2011). How much is enough? choosing ε for differential privacy. In *International Conference on Information Security*, pages 325–340. Springer.

Li, N., Li, T., and Venkatasubramanian, S. (2007). t-closeness: Privacy beyond k-anonymity and l-diversity. In *Data Engineering, 2007. ICDE 2007. IEEE 23rd International Conference on*, pages 106–115. IEEE.

Machanavajjhala, A., Kifer, D., Gehrke, J., and Venkitasubramaniam, M. (2007). l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3.

Makhdoumi, A. and Fawaz, N. (2013). Privacy-utility tradeoff under statistical uncertainty. In *Communication, Control, and Computing (Allerton), 2013 51st Annual Allerton Conference on*, pages 1627–1634. IEEE.

Rebollo-Monedero, D., Forne, J., and Domingo-Ferrer, J. (2010). From t-closeness-like privacy to postrandomization via information theory. *Knowledge and Data Engineering, IEEE Transactions on*, 22(11):1623–1636.

Salamatian, S., Zhang, A., du Pin Calmon, F., Bhamidipati, S., Fawaz, N., Kveton, B., Oliveira, P., and Taft, N. (2013). How to hide the elephant- or the donkey-in the room: Practical privacy against statistical inference for large data. In *GlobalSIP*, pages 269–272.

Samarati, P. (2001). Protecting respondents identities in microdata release. *Knowledge and Data Engineering, IEEE Transactions on*, 13(6):1010–1027.

Sankar, L., Rajagopalan, S. R., and Poor, H. V. (2013). Utility-privacy tradeoffs in databases: An information-theoretic approach. *IEEE Transactions on Information Forensics and Security*, 8(6):838–852.

Shannon, C. E. (2001). A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55.

Sondeck, L., Laurent, M., and Frey, V. (2017). Discrimination rate: an attribute-centric metric to measure privacy. *Annals of Telecommunications journal DOI: 10.1007/s12243-017-0581-8.*