From *t*-Closeness to PRAM and Noise Addition via Information Theory

David Rebollo-Monedero[1], Jordi Forné[1], and Josep Domingo-Ferrer[2]

[1] Telematics Engineering Dept., Technical University of Catalonia
C. Jordi Girona 1-3, E-08034 Barcelona, Catalonia
[2] UNESCO Chair in Data Privacy, Dept. of Computer Engineering and Maths,
Rovira i Virgili University
Av. Països Catalans 26, E-43007 Tarragona, Catalonia

Abstract. t-Closeness is a privacy model recently defined for data anonymization. A data set is said to satisfy t-closeness if, for each group of records sharing a combination of key attributes, the distance between the distribution of a confidential attribute in the group and the distribution of the attribute in the data is no more than a threshold t. We state here the t-closeness property in terms of information theory and then use the tools of that theory to show that t-closeness can be achieved by the PRAM masking method in the discrete case and by a form of noise addition in the general case.
Keywords: t-closeness, Microdata anonymization, Information theory, Rate distortion theory, PRAM, Noise addition.

1 Introduction

A microdata set is a data set whose records carry information on invidual respondents, like people or enterprises. The attributes in a microdata set can be classified as follows:

- *Identifiers.* These are attributes that *unambiguously* identify the respondent. Examples are passport number, social security number, full name, etc. Since our objective is to prevent confidential information from being linked to specific respondents, we will assume in what follows that, in a pre-processing step, identifiers have been removed/encrypted.
- *Key attributes.* Borrowing the definition from [1, 2], key attributes are those that, in combination, can be linked with external information to re-identify (some of) the respondents to whom (some of) the records in the microdata set refer. Examples are job, address, age, gender, etc. Unlike identifiers, key attributes cannot be removed, because any attribute is potentially a key attribute.
- *Confidential outcome attributes.* These are attributes which contain sensitive information on the respondent. Examples are salary, religion, political affiliation, health condition, etc.

There are several privacy models to anonymize microdata sets. k-Anonymity [2,3] is probably the best known. However, it presents several shortcomings which have motivated the appearance of enhanced privacy models reviewed below. t-Closeness [4] is one of those recent proposals. Despite its conceptual appeal, t-closeness lacks computational procedures which allow to reach it with minimum data utility loss.

1.1 Contribution and plan of this paper

We state here t-closeness as an information-theoretic problem, in such a way that the knowledge body of information theory can be used to find a solution to it. The resulting solution turns out to be the PRAM masking method [5, 6] in the discrete case and a form of noise addition in the general case.

Sec. 2 reviews the state of the art in k-anonymity-based privacy models. Sec. 3 gives an information-theoretic formulation of t-closeness. Sec. 4 is a theoretical analysis of the solution to t-closeness. Empirical results are reported in Sec. 5. Conclusions are drawn in Sec. 6.

2 Background and motivation

k-Anonymity requires that each combination of key attribute values should be shared by at least k records in the data set. To enforce k-anonymity, at least there are two computational procedures: the original approach based on generalization and recoding of the key attributes and a microaggregation-based approach described in [7] and illustrated in Fig 1. While k-anonymity prevents identity disclosure (re-identification is infeasible within a group sharing the same key attribute values), it may fail to protect against identity disclosure: such is the case if the k records sharing a combination of key attribute values also share the value of a confidential attribute. Several enhancements of k-anonymity have been proposed to address the above and other shortcomings. Some of them are mentioned in what follows.

In [8], an evolution of k-anonymity called p-sensitive k-anonymity was presented. Its purpose is to protect against attribute disclosure by requiring that there be at least p different values for each confidential attribute within the records sharing a combination of key attributes. p-Sensitive k-anonymity has the limitation of implicitly assuming that each confidential attribute takes values uniformly over its domain, that is, that the frequencies of the various values of a confidential attribute are similar. When this is not the case, achieving p-sensitive k-anonymity may cause a huge data utility loss.

Like p-sensitive k-anonymity, l-diversity [9] was defined with the aim of solving the attribute disclosure problem that can arise with k-anonymity. A data set is said to satisfy l-diversity if, for each group of records sharing a combination of key attributes, there are at least l "well-represented" values for each confidential attribute. Depending on the definition of "well-represented", l-diversity can reduce to p-sensitive k-anonymity or be a bit more complex. However, it

shares with the latter the problem of huge data utility loss. Also, it is insufficient to prevent attribute disclosure, because at least the following two attacks are conceivable:

- *Skewness attack.* If, within a group of records sharing a combination of key attributes, the distribution of the confidential attribute is very different from its distribution in the overall data set, then an intruder linking a specific respondent to that group may learn confidential information (*e.g.* imagine that the proportion of respondents with AIDS within the group is much higher than in the overall data set).
- *Similarity attack.* If values of a confidential attribute within a group are l-diverse but semantically similar (*e.g.* similar diseases or similar salaries), attribute disclosure also takes place.

t-Closeness [4] tries to overcome the above attacks. A microdata set is said to satisfy t-closeness if, for each combination of key attributes, the distance between the distribution of the confidential attribute in the group and the distribution of the attribute in the whole data set is no more than a threshold t. t-Closeness can be argued to protect against skewness and similarity (see [10] for a more detailed analysis):

- To the extent to which the within-group distribution of confidential attributes resembles the distribution of those attributes for the entire dataset, skewness attacks will be thwarted.
- Again, since the within-group distribution of confidential attributes mimics the distribution of those attributes over the entire dataset, no semantic similarity can occur within a group that does not occur in the entire dataset. (Of course, within-group similarity cannot be avoided if all patients in a data set have similar diseases.)

The main limitation of the original t-closeness paper is that no computational procedure to reach t-closeness was specified. This is what we address in the remainder of this paper by leaning on the framework of information theory.

3 Information-theoretic formulation of t-closeness

3.1 Conventions

Throughout the paper, the measurable space in which a random variable (r.v.) takes on values will be called an *alphabet*. All alphabets are assumed to be Polish spaces to ensure the existence of regular conditional probabilities, for example, any discrete space or the k-dimensional Euclidean space $\mathbb{R}^k$. We shall follow the convention of using uppercase letters for r.v.'s, and lowercase letters for particular values they take on. Probability density functions (PDFs) and probability mass functions (PMFs) are denoted by p, subindexed by the corresponding r.v. in case of ambiguity risk. For example, both $p_X(x)$ and $p(x)$ denote the value of the function p_X at x. The notation for information-theoretic quantities follows [11].

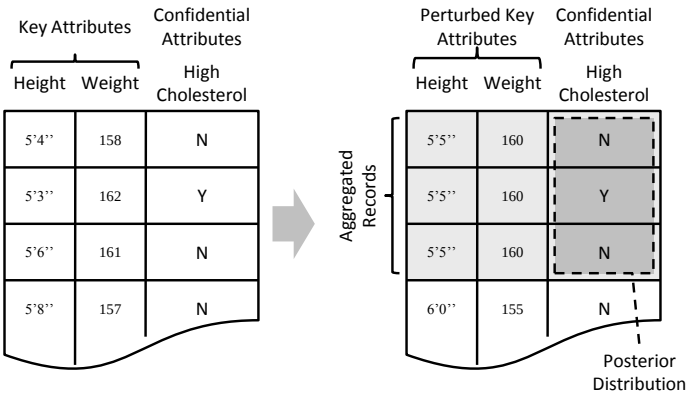

Fig. 1: Perturbation of key attributes to attain k-anonymity, t-closeness and similar privacy properties

3.2 Problem statement

Let W and X be jointly distributed r.v.'s in arbitrary alphabets, possibly discrete, continuous, or mixed Cartesian products. In the problem of database t-closeness described above and depicted in Fig. 1, X represents (the tuple of) key attributes to be perturbed, which could otherwise be used to identify an individual. In the same application, confidential attributes containing sensitive information are denoted by W. Assume that the joint distribution of X and W is known, for instance, an empirical distribution directly drawn from a table, or a parametric statistical model inferred from a subset of records.

A *distortion measure* $d(x, \hat{x})$ is any measurable, nonnegative, real-valued function representing the distortion between the original data X and a perturbed version $\hat{X}$, the latter also a r.v., commonly but not necessarily in the same alphabet of X. The associated expected distortion $\mathcal{D} = \mathrm{E}\, d(X, \hat{X})$ provides a measure of utility of the perturbed data, in the intuitive sense that low distortion approximately preserves the values of the original data, and their joint statistical properties with respect to any other data of interest, in particular W. For example, if $d(x, \hat{x}) = \|x - \hat{x}\|^2$, then $\mathcal{D}$ is the mean-square error (MSE).

Consider now, on the one hand, the distribution p_W of the confidential information W, and on the other, the conditional distribution $p_{W|\hat{X}}$ given the observation of the perturbed attributes $\hat{X}$. In the database k-anonymization problem, whenever the posterior distribution $p_{W|\hat{X}}$ differs from the prior distribution p_W, we have actually gained some information about individuals statistically linked to the perturbed key attributes $\hat{X}$, in contrast to the statistics of the general population. Concordantly, define the *privacy risk* $\mathcal{R}$ as the Kullback-Leibler (KL) divergence D between the posterior and the prior distributions, that is, $\mathcal{R} = \mathrm{D}(p_{W|\hat{X}} \| p_W)$, which is one of the measures proposed in the original t-closeness paper [4]. Simple information-theoretic manipulations show that the privacy risk thus defined coincides with the mutual information [11] $\mathcal{R} = \mathrm{I}(W; \hat{X})$, and that both the KL divergence and the mutual information

may be equivalently defined exchanging the roles of W and $\hat{X}$. Recall that the KL divergence vanishes (that is, one has 0-closeness) if, and only if, the distributions match (almost surely), which in turn is equivalent to requiring that W and $\hat{X}$ be statistically independent. Of course, in this extreme case, the utility of the published data, represented by the distribution $p_{W\hat{X}}$, usually by means of the corresponding table, is severely compromised. In the other extreme, leaving the original data undistorted, i.e., $\hat{X} = X$, compromises privacy, because in general $p_{W|X}$ and p_W differ.

We would like to remark at this point that the use of an information-theoretic quantity for database privacy assessment is by no means new. In addition to the t-closeness work already cited, [12–14] already used Shannon entropy as a measure of information loss, pointing out limitations affecting specific applications. We would like to stress out that we use a KL divergence as a measure of *information disclosure* (rather than loss), consistently with the equivalence between the case when $p_{W|\hat{X}} = p_W$ and the complete absence of privacy risk. On the other hand, the flexibility in our definition of distortion measure as a measure of *information loss* may enable us to preserve the statistical properties of the perturbed data to an arbitrary degree, possibly with respect to any other data of interest. Of course, the choice of distortion measure should ultimately rely on each particular application.

Consequently, we are interested in the tradeoff between two contrasting quantities, privacy and distortion, by means of perturbation of the original data. More precisely, consider *randomized perturbation rules* on the original data X, determined by the conditional distribution $p_{\hat{X}|X}$ of the perturbed data $\hat{X}$ given X. In the special case when the alphabets involved are finite, $p_{\hat{X}|X}$ may be regarded as a transition probability matrix, such as the one that appears in the PRAM masking method [5, 6]. The Markov chain $W \leftrightarrow X \leftrightarrow \hat{X}$, stating the conditional independence of $\hat{X}$ and W given X, emphasizes that this randomized rule has only X as input, but not W. Two remarks are in order. First, we consider randomized rules because deterministic quantizers are a particular case, and at this point we may not discard the possibility that more general rules attain a better tradeoff. Secondly, we consider rules that affect and depend on X only, but not W, for simplicity. Specifically, implementing and estimating convenient conditional distributions $p_{\hat{X}|WX}$ rather than $p_{\hat{X}|X}$ will usually be more complex, and require large quantities of data to prevent overfitting issues.

To sum up, we are interested in a randomized perturbation minimizing the privacy risk given a distortion constraint (or viceversa). In mathematical terms, we consistently define the *privacy-distortion function* as

$$\mathcal{R}(\mathcal{D}) = \inf_{\substack{p_{\hat{X}|X} \\ \mathrm{E}\, d(X,\hat{X}) \leqslant \mathcal{D}}} \mathrm{I}(W;\hat{X}). \tag{1}$$

For conceptual convenience, we provide an equivalent definition introducing an auxiliary r.v. Q, playing the role of randomized quantization index, a randomized

quantizer $p_{Q|X}$, and a reconstruction function $\hat{x}(q)$:

$$\mathcal{R}(\mathcal{D}) = \inf_{\substack{p_{Q|X},\,\hat{x}(q) \\ \mathrm{E}\,d(X,\hat{X})\leqslant\mathcal{D}}} \mathrm{I}(W;Q).$$

It can be shown [15] that there is no loss of generality in assuming that Q and $\hat{X}$ are related bijectively, thus $\mathrm{I}(W;Q) = \mathrm{I}(W;\hat{X})$, and that both definitions indeed lead to the same function. The elements involved in the definition of the privacy-distortion function are depicted in Fig. 2.

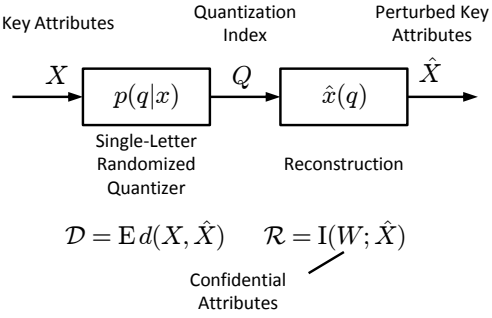

Fig. 2: Information-theoretic formulation of the privacy-distortion problem.

Even though the motivating application for this work is the problem of database t-closeness, it is important to notice that our formulation in principle addresses any applications where perturbative methods for privacy are of interest. Another illustrative application is privacy for location-based services (LBS). In this scenario, private information such as the user's location (or a sequence thereof) may be modeled by the r.v. X, to be perturbed, and W may represent a user ID. The posterior distribution $p_{\hat{X}|W}$ now becomes the distribution of the user's perturbed location, and the prior distribution $p_{\hat{X}}$, the population's distribution.

3.3 Connection with information theory

Perhaps the most attractive aspect of the formulation of the privacy-distortion problem in Sec. 3.2 is the strong resemblance it bears with the *rate-distortion problem* in the field of information theory. We shall see that our formulation is a generalization of a well-known, extensively studied information-theoretic problem with half a century of maturity. Namely, the problem of lossy compression of source data with a distortion criterion, first proposed by Shannon in 1959 [16].

To emphasize the connection, briefly recall that the simplest version of the problem of lossy data compression, shown in Fig. 3, involves coding of identically distributed (i.i.d.) copies $X_1, X_2, \ldots$ of a generic r.v. X. To this end, an n-letter deterministic quantizer maps blocks of n copies $X_1, \ldots, X_n$ into quantization indices Q in the set $\{1, \ldots, \lfloor 2^{n\mathcal{R}} \rfloor\}$, where $\mathcal{R}$ represents the coding rate in bits per sample. An estimation $\hat{X}_1, \ldots, \hat{X}_n$ of the source data vector is recovered to minimize the expected distortion per sample $\mathcal{D} = \frac{1}{n}\sum_i \mathrm{E}\,d(X_i, \hat{X}_i)$, according

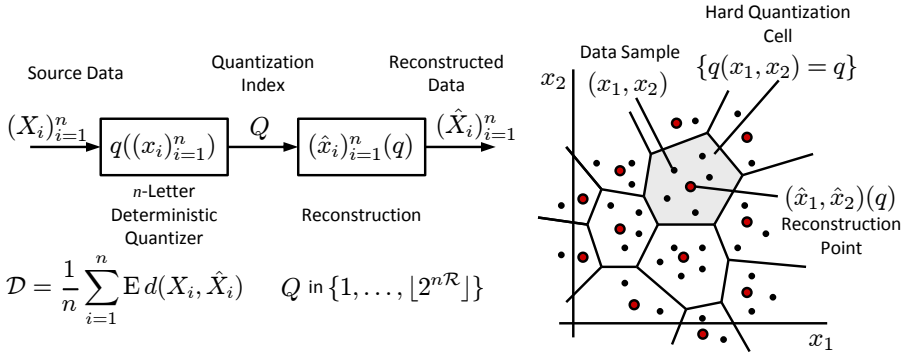

Fig. 3: Information-theoretic formulation of the rate-distortion problem.

to some distortion measure $d(x, \hat{x})$. Intuitively, a rate of zero bits may only be achieved in the uninteresting case when no information is conveyed, whereas in the absence of distortion, the rate is maximized. Rate-distortion theory deals with the characterization of the optimal tradeoff between the rate $\mathcal{R}$ and the distortion $\mathcal{D}$, allowing codes with arbitrarily large block length n. Accordingly, the *rate-distortion function* is defined as the infimum of the rates of codes satisfying a distortion constraint.

A surprising and fundamental result of rate-distortion theory is that such function, defined in terms of blocks of samples, can be expressed in terms of a single copy of the source data vector [11], often more suitable for theoretical analysis. More precisely, the *single-letter characterization of the rate-distortion function* is

$$\mathcal{R}(\mathcal{D}) = \inf_{\substack{p_{\hat{X}|X} \\ \mathrm{E}\, d(X,\hat{X}) \leqslant \mathcal{D}}} \mathrm{I}(X; \hat{X}) = \inf_{\substack{p_{Q|X},\, \hat{x}(q) \\ \mathrm{E}\, d(X,\hat{X}) \leqslant \mathcal{D}}} \mathrm{I}(X; Q), \tag{2}$$

represented in Fig. 4. Aside from the fact that the equivalent problem is expressed

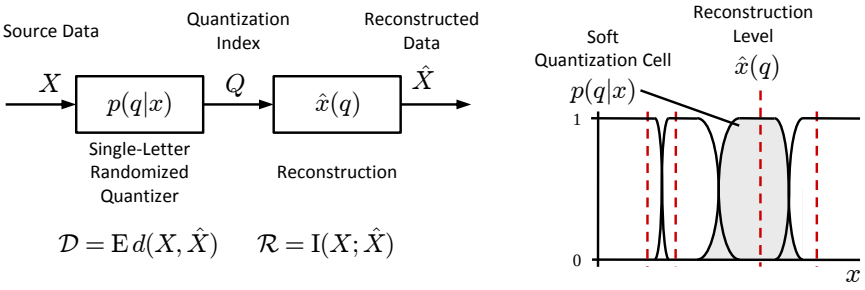

Fig. 4: Single-letter characterization of the rate-distortion problem.

in terms of a single letter X rather than n copies, there are two additional differences. First, the quantizer is randomized, and determined by a conditional distribution $p_{Q|X}$. Secondly, the rate is no longer the number of bits required to index quantization cells, or even the lowest achievable rate using an ideal

entropy coder, namely the entropy of the quantization index $\mathrm{H}(Q)$. Instead, the rate is a mutual information $\mathcal{R} = \mathrm{I}(X;\hat{X})$.

Interestingly, the single-letter characterization of the rate-distortion function (2) is almost identical to our definition of privacy-distortion function (1), except for the fact that in the latter there is an extra variable W, the confidential attributes, in general different from X, the key attributes. It turns out that some of the information-theoretic results and methods for the rate-distortion problem can be extended, with varying degrees of effort, to the privacy-distortion problem formulated in this work. Some of these extensions are discussed in the next section.

4 Theoretical analysis

All theoretical claims in this section are detailed and proven in [15].

Similarly to the rate-distortion function, the privacy-distortion function (1) is decreasing, convex, and continuous in the interior of its domain. Furthermore, the optimization problem determining (1), with $p_{\hat{X}|W}$ as unknown variable, is itself convex. This means that any local minimum is also global, and makes the powerful tools of convex optimization [17] applicable to compute numerically but efficiently the privacy-distortion function. In Sec. 5, an example of numerical computation will be discussed.

While a general closed-form expression for privacy-distortion function has not been provided, the Shannon lower bound for the rate-distortion function can be extended to find a closed-form lower bound under certain assumptions. Furthermore, the techniques used to prove this bound may yield an exact closed formula in specific cases. A closed-form upper bound is also presented in this section.

Suppose that W and X are real-valued r.v.'s (random scalars), and that MSE is used as distortion measure, thus $\mathcal{D} = \mathrm{E}(X-\hat{X})^2$. Define the normalized distortion $d = \frac{\mathcal{D}}{\sigma_X^2}$, where σ_X^2 denotes the variance of X. Let σ_W^2 be the variance of W, ρ_{WX} the correlation coefficient of W and X, and $\mathrm{h}(W)$ the differential entropy [11] of W. Then,

$$\mathcal{R}(\mathcal{D}) \geqslant \mathcal{R}_{\mathrm{QGLB}}(\mathcal{D}) = \mathrm{h}(W) - \frac{1}{2}\log\left(2\pi e\left(1-(1-d)\rho_{WX}^2\right)\sigma_W^2\right) \qquad (3)$$

for $0 \leqslant d \leqslant 1$ (for $d \geqslant 1$, clearly $\mathcal{R} = 0$). We shall call the bounding function $\mathcal{R}_{\mathrm{QGLB}}(\mathcal{D})$ the *quadratic-Gaussian lower bound* (QGLB).

With the same assumptions, namely scalar r.v.'s and MSE distortion measure, consider the two trivial cases $d = 0$ and $d = 1$. The former can be achieved with $\hat{X} = X$, yielding $\mathcal{R}(\mathcal{D}) = \mathrm{I}(W;X)$, and the latter with $\hat{X} = \mu_X$, the mean of X, for which $\mathcal{R}(\mathcal{D}) = 0$. Now, for any $0 \leqslant d \leqslant 1$, set $\hat{X} = X$ with probability $1-d$, and $\hat{X} = \mu_X$ with probability d. Convexity properties of the mutual information guarantee that the privacy-distortion performance of this setting cannot lie above the segment connecting the two trivial cases. Since the setting is not necessarily optimal, it may be concluded that

$$\mathcal{R}(\mathcal{D}) \leqslant \mathcal{R}_{\mathrm{MIUB}}(\mathcal{D}) = \mathrm{I}(W;X)(1-d). \qquad (4)$$

We shall call this bounding function the *mutual-information upper bound* (MIUB). The $p_{\hat{X}|X}$ determined by the combination of the two trivial cases for intermediate values of d may be a simple yet effective way to initialize numerical search methods to compute the privacy-distortion function, as it will be shown in Sec. 5.

Provided that W and X are jointly Gaussian, real-valued r.v.'s, and that MSE is used as distortion measure, the QGLB (3) is tight:

$$\mathcal{R}(\mathcal{D}) = -\frac{1}{2}\log\left(1-(1-d)\rho_{WX}^2\right), \tag{5}$$

with $d = \frac{\mathcal{D}}{\sigma_X^2} \leqslant 1$ as before. The optimal randomized perturbation rule achieving this privacy-distortion performance is represented in Fig. 5. Observe that the

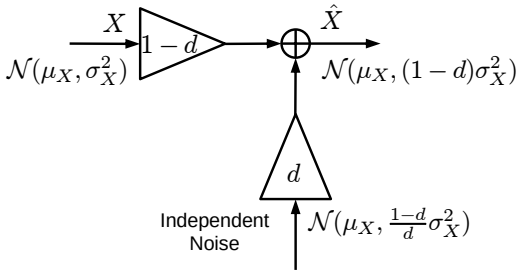

Fig. 5: Optimal randomized perturbation in the quadratic-Gaussian case.

perturbed data $\hat{X}$ is a convex combination of the source data X and independent noise, in a way such that the final variance achieves the distortion constraint with equality.

5 Numerical computation example

In this section, we illustrate the theoretical analysis of Sec. 4 with experimental results for a simple, intuitive case. Specifically, W and X are jointly Gaussian random scalars with correlation coefficient ρ (after zero-mean, unit-variance normalization). In terms of the database microaggregation problem, W represents sensitive information, and X corresponds to key attributes that can be used to identify specific individuals. These variables could model, for example, the plasma concentration of LDL cholesterol in adults, which is approximately normal, and their weight, respectively. MSE is used as a distortion measure. For convenience $\sigma_X^2 = 1$, thus $\mathcal{D} = d$. Since the privacy-distortion function is convex, minimization of one objective with a constraint on the other is equivalent to the minimization of the Lagrangian cost $\mathcal{C} = \mathcal{D} + \lambda\mathcal{R}$, for some positive multiplier λ. We wish to design randomized perturbation rules $p_{\hat{X}|X}$ minimizing $\mathcal{C}$ for several values of λ, to investigate the feasibility of numerical computation of the privacy-distortion curve, and to verify the theoretic results for the quadratic-Gaussian case of Sec. 4.

We implement a slight modification of a simple optimization technique, namely the steepest descent algorithm, operating on a sufficiently fine discretization of

the variables involved. More precisely, p_{WX} is the joint PMF obtained by discretizing the PDF of W and X, where each variable is quantized with 31 samples in the interval $[-3,3]$. The starting values for $p_{\hat{X}|X}$ are convex combinations of the extreme cases corresponding to $d=0$ and $d=1$, as described in Sec. 4 when the MIUB (4) was discussed. Only results corresponding to the correlation coefficient $\rho=0.95$ are shown, for two reasons. First, because of their similarity with results for other values of ρ. Secondly, because for high correlation, the gap between the MIUB (which approximates the performance of the starting solutions) and the QGLB (3) is wider, leading to a more challenging problem.

The definitions of distortion and privacy risk in Sec. 3 for the finite-alphabet case become

$$\mathcal{D}=\sum_x\sum_{\hat{x}} p(x)p(\hat{x}|x)d(x,\hat{x}), \qquad \mathcal{R}=\sum_w\sum_{\hat{x}} p(w)p(\hat{x}|w)\ln\frac{p(\hat{x}|w)}{p(\hat{x})}.$$

The conditional independence assumption in the same section enables us to express the PMFs of $\hat{X}$ in the expression for $\mathcal{R}$ as $p(\hat{x})=\sum_x p(\hat{x}|x)p(x)$ and $p(\hat{x}|w)=\sum_x p(\hat{x}|x)p(x|w)$, in terms of the optimization variables $p(\hat{x}|x)$. Our implementation of the steepest descent algorithm uses the exact gradient with components $\frac{\partial\mathcal{C}}{\partial p(\hat{x}|x)}=\frac{\partial\mathcal{D}}{\partial p(\hat{x}|x)}+\lambda\frac{\partial\mathcal{R}}{\partial p(\hat{x}|x)}$, where $\frac{\partial\mathcal{D}}{\partial p(\hat{x}|x)}=p(x)d(x,\hat{x})$ and

$$\frac{\partial\mathcal{R}}{\partial p(\hat{x}|x)}=p(x)\left(\sum_w p(w|x)\ln p(\hat{x}|w)-\ln p(\hat{x})\right)$$

(after simplification [15]).

Two modifications of the standard version of the steepest descent algorithm [17] were applied. First, rather than updating $p_{\hat{X}|X}$ directly according to the negative gradient multiplied by a small factor, we used its projection onto the affine set of conditional probabilities satisfying $\sum_{\hat{x}} p(\hat{x}|x)=1$ for all x, which in fact gives the steepest descent within that set. Secondly, rather than using a barrier or a Lagrangian function to consider the constraint $p(\hat{x}|x)\geqslant 0$ for all x and $\hat{x}$, after each iteration, we reset possible negative values to 0 and renormalized the probabilities accordingly. This may seem unnecessary since the theoretical analysis in Sec. 4 gives a strictly feasible solution (i.e., probabilities are strictly positive), and consequently the constraints are inactive. However, the algorithm operates on a discretization of the joint distribution of W and X in a machine with finite precision. The fact is that precision errors in the computation of gradient components corresponding to very low probabilities activated the nonnegativity constraints. Finally, we observed that the ratio between the largest and the smallest eigenvalue of the Hessian matrix was large enough for the algorithm to require a fairly small update factor, 10^{-4}, to prevent significant oscillations.

The privacy-distortion performance of the randomized perturbation rules $p_{\hat{X}|X}$ found by our modification of the steepest descent algorithm is shown in Fig. 6, along with the bounds established in Sec. 4, namely the QGLB (3) and the MIUB (4). On account of (5), it can be shown that $\lambda=2\sigma_X^2\left(1/\rho^2-1+d\right)$. Accordingly, we set λ approximately to 0.72, 1.22 and 1.72, which theoretically corresponds to $d=0.25,0.5,0.75$. A total of 32000 iterations were computed for

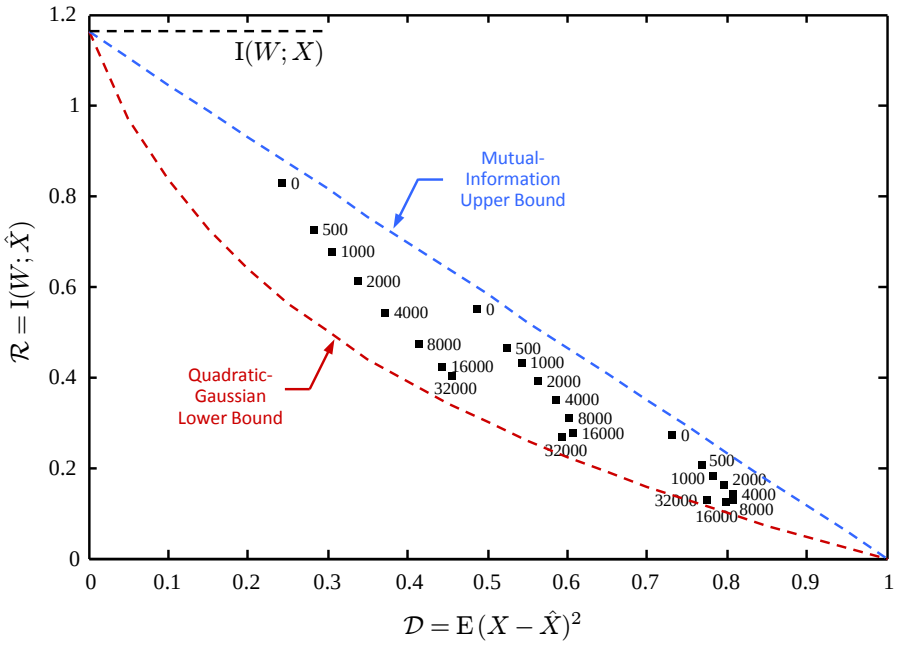

Fig. 6: Privacy-distortion performance of randomized perturbation rules found by a modification of the steepest descent algorithm.

each value of λ, at about 16 iterations per second on a modern computer[(a)]. The large number of iterations is consistent with the fact that the Hessian is ill-conditioned and the small updating step size. Obviously, one would expect that methods based on Newton's technique [17] converge to the optimal solution in less iterations (at the cost of higher computational complexity per iteration), but our goal was to check the performance of one of the simplest optimization algorithms. In all cases, the conditional PMFs found had a performance very close to that described by (5) in Sec. 4. Their shape, depicted in Fig. 7, roughly resembled the Gaussian shape predicted by the theoretical analysis as the number of iterations increased. Specifically, Fig. 7 corresponds to $\lambda \simeq 1.22$, was obtained after 32000 iterations, and the number of discretized samples of X and W was increased from 31 to 51. Increasing the number of iterations to 128000 resulted in an experimental solution shaped almost identically to the optimal one, although the one in Fig. 7, corresponding to a fourth of the number of iterations, already achieves values of $\mathcal{C}$ reasonably optimal.

6 Conclusion

An information-theoretic formulation of the privacy-distortion tradeoff in applications such as microdata anonymization and location privacy in location-based

(a) Implementation used Matlab R2007b on Windows Vista SP1, on an Intel Core2 Quad Q6600 CPU at 2.4 GHz.

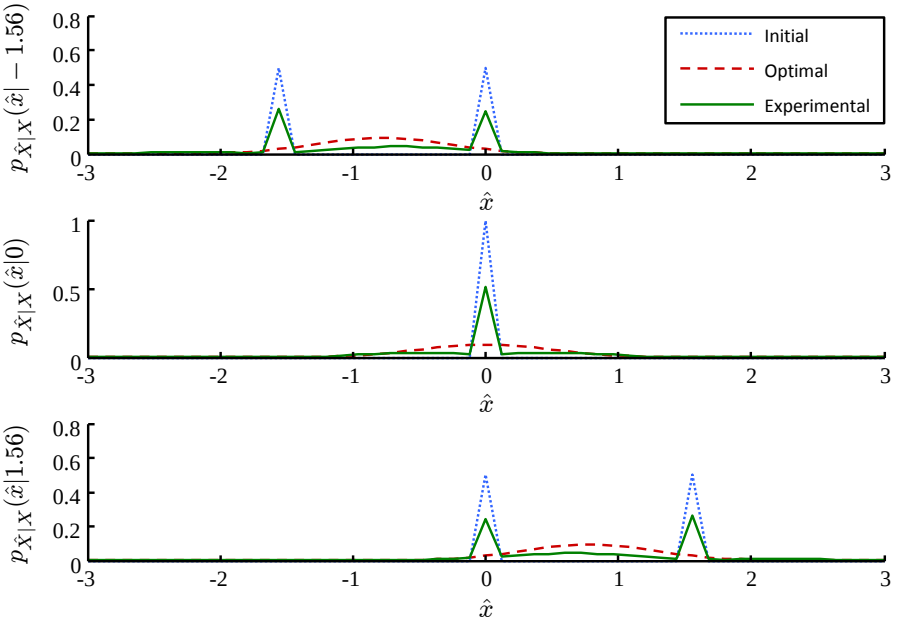

Fig. 7: Shape of initial, optimal, and experimental randomized perturbation rules $p_{\hat{X}|X}$ found by the steepest descent algorithm.

services is provided. Following the t-closeness model, the privacy risk is measured as the mutual information between perturbed key attributes and confidential attributes, equivalent to the KL divergence between posterior and prior distributions. We consider the problem of maximizing privacy (that is, minimizing the above mutual information) while keeping the perturbation of data within a pre-specified bound to ensure that data utility is not too damaged. We establish a strong connection between this privacy-perturbation problem and the rate-distortion problem of information theory and extend of a number of results, including convexity of the privacy-distortion function and the Shannon lower bound. A closed formula is obtained for the quadratic-Gaussian case, proving that the optimal perturbation is randomized rather than deterministic, which justifies the use of PRAM in the case of attributes with finite alphabets or noise addition in the general case.

Acknowledgments and disclaimer

We would like to thank the anonymous reviewers for their valuable comments. This work was partly supported by the Spanish Government through projects CONSOLIDER INGENIO 2010 CSD2007-00004 “ARES”, TSI2007-65393-C02-02 “ITACA” and TSI2007-65406-C03-01 “E-AEGIS”, and by the Government of Catalonia under grants 2005 SGR 00446 and 2005 SGR 01015. The third author is with the UNESCO Chair in Data Privacy, but his views do not necessarily reflect the position of UNESCO nor commit that organization.

References

1. Dalenius, T.: Finding a needle in a haystack - or identifying anonymous census records. Journal of Official Statistics **2**(3) (1986) 329–336
2. Samarati, P.: Protecting respondents' identities in microdata release. IEEE Transactions on Knowledge and Data Engineering **13**(6) (2001) 1010–1027
3. Samarati, P., Sweeney, L.: Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical report, SRI International (1998)
4. Li, N., Li, T., Venkatasubramanian, S.: t-closeness: Privacy beyond k-anonymity and l-diversity. In: Proc. IEEE Int. Conf. Data Eng. (ICDE), Istanbul, Turkey (April 2007) 106–115
5. Gouweleeuw, J.M., Kooiman, P., Willenborg, L.C.R.J., DeWolf, P.P.: Post randomisation for statistical disclosure control: Theory and implementation (1997) Research paper no. 9731 (Voorburg: Statistics Netherlands).
6. de Wolf, P.P.: Risk, utility and PRAM. In: Privacy Stat. Databases (PSD). Volume 4302 of Lecture Notes Comput. Sci. (LNCS)., Rome, Italy, Springer-Verlag (December 2006) 189–204
7. Domingo-Ferrer, J., Torra, V.: Ordinal, continuous and heterogenerous k-anonymity through microaggregation. Data Mining and Knowledge Discovery **11**(2) (2005) 195–212
8. Truta, T.M., Vinay, B.: Privacy protection: p-sensitive k-anonymity property. In: 2nd International Workshop on Privacy Data Management PDM 2006, Atlanta, GA, IEEE Computer Society (2006) p. 94
9. Machanavajjhala, A., Gehrke, J., Kiefer, D., Venkitasubramanian, M.: L-diversity: privacy beyond k-anonymity. In: Proceedings of the IEEE ICDE 2006. (2006)
10. Domingo-Ferrer, J., Torra, V.: A critique of k-anonymity and some of its enhancements. In: Proceedings of ARES/PSAI'2008, Los Alamitos CA, IEEE Computer Society (2008) 990–993
11. Cover, T.M., Thomas, J.A.: Elements of Information Theory. Wiley, New York (1991)
12. Kooiman, P.L., Willenborg, L., Gouweleeuw, J.: PRAM: A method for disclosure limitation of microdata. Research Rep. 9705, Statistics Netherlands, Voorburg, NL (1998)
13. de Waal, T., Willenborg, L.: Information loss through global recoding and local suppression. Netherlands Official Stat. **14** (1999) 17–20
14. Willenborg, L., de Waal, T.: Elements of Statistical Disclosure Control. Springer-Verlag, New York (2001)
15. Rebollo-Monedero, D., Forné, J.: An information-theoretic formulation of the privacy-distortion tradeoff. Research rep., Tech. Univ. of Catalonia (UPC) (June 2008)
16. Shannon, C.E.: Coding theorems for a discrete source with a fidelity criterion. In: IRE Nat. Conv. Rec. Volume 7 Part 4. (1959) 142–163
17. Boyd, S., Vandenberghe, L.: Convex Optimization. Cambridge University Press, Cambridge, UK (2004)