Worst- and Average-Case Privacy Breaches in Randomization Mechanisms

Michele Boreale, Michela Paolini

Worst- and average-case privacy breaches in randomization mechanisms*

Michele Boreale[1] and Michela Paolini[2]

[1] Università di Firenze, Italy [2] IMT - Institute for Advanced Studies, Lucca, Italy

Abstract In a variety of contexts, randomization is regarded as an effective technique to conceal sensitive information. We model randomization mechanisms as information-theoretic channels. Our starting point is a semantic notion of security that expresses absence of any privacy breach above a given level of seriousness ϵ, irrespective of any background information, represented as a prior probability on the secret inputs. We first examine this notion according to two dimensions: worst vs. average case, single vs. repeated observations. In each case, we characterize the security level achievable by a mechanism in a simple fashion that only depends on the channel matrix, and specifically on certain measures of "distance" between its rows, like norm-1 distance and Chernoff Information. We next clarify the relation between our worst-case security notion and differential privacy (DP): we show that, while the former is in general stronger, the two coincide if one confines to background information that can be factorised into the product of independent priors over individuals. We finally turn our attention to expected utility, in the sense of Ghosh et al., in the case of repeated independent observations. We characterize the exponential growth rate of any reasonable utility function. In the particular case the mechanism provides ϵ-DP, we study the relation of the utility rate with ϵ: we offer either exact expressions or upper-bounds for utility rate that apply to practically interesting cases, such as the (truncated) geometric mechanism.

Keywords: Foundations of security, quantitative information flow, differential privacy, utility, information theory.

1 Introduction

In a variety of contexts, randomization is regarded as an effective means to conceal sensitive information. For example, anonymity protocols like Crowds [24] or the Dining Cryptographers [11] rely on randomization to "confound" the adversary as to the true actions undertaken by each participant. In the field of Data Mining, techniques have been proposed by which datasets containing personal information that are released for business or research purposes are perturbed with noise, so as to prevent an adversary from re-identifying individuals or learning sensitive information about them (see e.g. [15] and references therein).

* Work partially supported by the EU funded project ASCENS. Corresponding author: Michele Boreale, Università di Firenze, Dipartimento di Sistemi e Informatica, Viale Morgagni 65, I-50134 Firenze, Italy. E-mail: michele.boreale@unifi.it.

In the last few years, interest in the theoretical principles underlying randomization-based information protection has been steadily growing. Two major areas have by now clearly emerged: *Quantitative Information Flow* (QIF) [8,19,5,6,9,10,26] and *Differential Privacy* (DP) [13,14,21,22,16,17]. As discussed in [4], QIF is mainly concerned with quantifying the degree of protection offered against an adversary trying to guess the whole secret; DP is rather concerned with protection of individual bits of the secret, possibly in the presence of background information, like knowledge of the remaining bits. The areas of QIF and DP have grown separately for some time: only very recently researchers have begun investigating the relations between these two notions [1,2,3,4].

The present paper is an attempt at distilling and systematizing the notions of security breach underlying QIF and DP. We view a randomization mechanism as an information-theoretic channel with inputs in $\mathcal{X}$ and outputs in $\mathcal{Y}$. The starting point of our treatment is a semantical notion of breach. Assume $\mathcal{X}$ is a finite set of items containing the secret information X, about which the adversary has some background knowledge or belief, modeled as a prior probability distribution $p(x)$. Consider a predicate $Q \subseteq \mathcal{X}$ – in a dataset about individuals, one may think of Q as gender, or membership in a given ethnical group etc. The mere fact that X is in Q or not, if ascertained, may convey sensitive information about X. Henceforth, any observation $y \in \mathcal{Y}$ that causes a significant change in the adversary's posterior belief about $X \in Q$ must be regarded as dangerous. In probabilistic terms, Q is a *breach* if, for some prior probability on $\mathcal{X}$, the posterior probability of Q after interaction with the randomization mechanism exhibits a significant change, compared to its prior probability. We decree a randomization mechanism as secure at level ϵ, if it exhibits *no breach* of level $> \epsilon$, independently of the prior distribution on the set of secret data $\mathcal{X}$. The smaller ϵ, the more secure the mechanism. This simple idea, or variations thereof, has been proposed elsewhere in the Data Mining literature – see e.g. [15]. Here, we are chiefly interested in analyzing this notion of breach according to the following dimensions.

1. Worst- vs. average-case security. In the worst-case approach, one is interested in bounding the level of any breach, independently of how likely the breach is. In the average-case, one takes into account the probability of the observations leading to the breach.
2. Single vs. repeated, independent executions of the mechanism.
3. Expected utility of the mechanism and its asymptotic behavior, depending on the number of observations and on a user-defined loss function.

To offer some motivations for the above list, we observe that worst-case is the type of breach considered in DP, while average-case is the type considered in QIF. In the worst-case scenario, another issue we consider is resistance to background information. In the case of DP, this is often stated in the terms that [13]: *Regardless of external knowledge, an adversary with access to the sanitized database draws the same conclusions whether or not my data is included*, and formalized as such [17]. We investigate how this relates to the notion of privacy breach we consider, which also intends to offer protection against arbitrary background knowledge.

Concerning the second point, a scenario of repeated observations seems to arise quite naturally in many applications. For instance, an online, randomized data-releasing

mechanism might offer users the possibility of asking the same query a number of times. This allows the user to compute more accurate answers, but also poses potential security threats, as an adversary could remove enough noise to learn valuable information about the secret. This is an instance of the *composition* attacks which are well known in the context of DP, where they are thwarted by allotting each user or group of users a *privacy budget* that limits the overall number of queries to the mechanism; see e.g. [21,16]. For another example, in a de-anonymization scenario similar to [23], [6] shows that gathering information about a target individual can be modeled as collecting multiple observations from a certain randomization mechanism. In general, one would like to assess the security of a mechanism in these situations. In particular, one would like to determine exactly *how fast* the level of any potential breach grows, as the number n of independent observations grows.

The third point, concerning utility, has been the subject of intensive investigation lately (see related work paragraph). Here, we are interested in studying the growth of expected utility in the model of Ghosh et al. [18] as the number of independent observations grows, and to understand how this is related to security.

In summary, the main results we obtain are the following.

- In the scenario of a single observation, both in the average and in the worst case, we characterize the security level (absence of breach above a certain threshold) of the randomization mechanism in a simple way that only depends on certain row-distance measures of the underlying matrix.
- We prove that our notion of worst-case security is stronger than DP. However, we show the two notions coincide when one confines to background information that factorises as the product of independent measures over all individuals. This, we think, sheds further light on resistance of DP against background knowledge.
- In the scenario of repeated, independent observations, we determine the exact asymptotic growth rate of the (in)security level, both in the worst and in the average case.
- In the scenario of repeated, independent observations, we determine the exact asymptotic growth rate of any reasonable expected utility. We also give bounds relating this rate to ϵ-DP, and exact expressions in the case of the geometric mechanisms. In this respect, we argue that the geometric mechanism is superior to its *truncated* version [18].

Related work There is a large body of recent literature on QIF [8,19,5,6] and DP [13,14]. The earliest proposal of a worst-case security notion is, to the best of our knowledge, found in [15]. As mentioned, the investigation of the relations between QIF and DP has just begun. Both [4] and [2,3] discuss the implication of ϵ-DP on information leakage guarantees, and vice-versa, in the case of a single observation. In the present work, we propose and characterize both worst- and average-case semantic notions of privacy breach, encoding resistance to arbitrary side-information, and clarify their relationships with QIF and DP. We also study the asymptotic behavior of privacy breaches depending on the number of observations.

The notion of utility has been the subject of intensive investigation in the field of DP, see e.g. [22,18,1,2,3] and references therein. A general goal is that of designing mechanisms achieving optimal expected utility given a certain security level ϵ. Ghosh et al.

[18] propose a model of expected utility based on user preferences, and show that both the geometric mechanism and its truncated version achieve universal optimality. Here we provide the growth rate of utility, and we highlight a difference between a mechanism and its truncated version, in the presence of repeated observations. Alvim et al. [1] have shown the tight connection between utility and Bayes risk, hence information leakage, in the case of a single observation. A different, somewhat stronger notion of utility, called *accuracy*, is considered by McSherry and Talwar [22]. They do not presuppose any user-specific prior over the set of possible answers; rather, they show that, in the exponential mechanism they propose, for any database, the expected *score* of the answer comes close to the maximum.

Structure of the paper The rest of the paper is organized as follows. In Section 2 we review some terminology and basic concepts about Bayesian hypothesis testing and information leakage. Section 3 characterizes the semantic security of randomization mechanisms, both in the worst and in the average case, but limited to a single observation on the part of the adversary. Section 4 discusses the relation between DP and our worst-case security. Section 5 discusses the asymptotic behavior of the security level in the case of n independent observations where the secret input remains fixed, again both in the worst and in the average case. In the worst case, we also offer a result characterizing the probability, depending on n, that *some* sequence of observations causes a breach. In Section 6 we deal with utility in the case of repeated observations. Section 7 discusses further work and draws some concluding remarks. Due to space limitations proofs have been omitted; they can be found in a full version available online [7].

2 Preliminaries

We review some notation and basic concepts about Bayesian hypothesis testing and information leakage.

2.1 Basic terminology

Let $\mathcal{X}$ be a finite nonempty set. A probability distribution on $\mathcal{X}$ is a function $p : \mathcal{X} \to [0, 1]$ such that $\sum_{x\in\mathcal{X}} p(x) = 1$. The *support* of p is defined as $\mathrm{supp}(p) \stackrel{\triangle}{=} \{x \in \mathcal{X} | p(x) > 0\}$. For any $Q \subseteq \mathcal{X}$ we let $p(Q)$ denote $\sum_{x\in Q} p(x)$. Given $n \geq 0$, we let $p^n : \mathcal{X}^n \to [0, 1]$ denote the n-th extension of p, defined as $p^n(x_1, \ldots, x_n) \triangleq \prod_{i=1}^{n} p(x_i)$; this is in turn a probability distribution on the set X^n. When $Q \subseteq \mathcal{X}^n$ and n is clear from the context, we shall abbreviate $p^n(Q)$ as just $p(Q)$. For $n = 0$, we set $p^0(\epsilon) = 1$, where ϵ denotes the empty tuple. $\Pr(\cdot)$ will generally denote a probability measure defined on some probability space (understood from the context). Given a random variable X taking values in $\mathcal{X}$, we write $X \sim p(x)$ if X is distributed according to $p(x)$, that is for each $x \in \mathcal{X}$, $\Pr(X = x) = p(x)$. We shall only consider discrete random variables. Suppose we are given random variables X, Y,... taking values in $\mathcal{X}$, $\mathcal{Y}$,... and defined on the same probability space. We shall use abbreviations such as $p(y|x)$ for $\Pr(Y = y|X = x)$, $p(y|Q)$ for $\Pr(Y = y|X \in Q)$, and so on, whenever no confusion arises about the involved random variables X and Y. Finally, when notationally convenient, we shall denote the

conditional probability distribution on $\mathcal{Y}$ $p(\cdot|x)$ ($x \in \mathcal{X}$) as $p_x(\cdot)$. Randomization mechanisms are information-theoretic channels. The use of this concept in the field of QIF has been promoted by Chatzikokolakis, Palamidessi and collaborators [9,10,8]; the systems amenable to this form of representation are sometimes referred to as *information hiding systems* (see also [5,6]).

Definition 1 (randomization mechanism). *A* randomization mechanism *is a triple* $\mathcal{R} = (\mathcal{X}, \mathcal{Y}, p(\cdot|\cdot))$*, composed by a finite set of* inputs $\mathcal{X}$ *representing the secret information, a finite set of* observables $\mathcal{Y}$ *representing the observable values, and a* conditional probability matrix, $p(\cdot|\cdot) \in [0,1]^{\mathcal{X}\times\mathcal{Y}}$*, where each row sums up to 1.*

The entry of row x and column y of the channel's matrix will be written as $p(y|x)$, and represents the probability of observing y, given that x is the (secret) input of the system. For each x, the x-th row of the matrix is identified with the probability distribution on $\mathcal{Y}$ given by $y \mapsto p(y|x)$, which is denoted by p_x. We say $\mathcal{R}$ is *non-degenerate* if $x \neq x'$ implies $p_x \neq p_{x'}$, and *strictly positive* if $p(y|x) > 0$ for each x and y. Note that $p(\cdot)$ on $\mathcal{X}$ and the conditional probability matrix $p(y|x)$ together induce a probability distribution q on $\mathcal{X} \times \mathcal{Y}$ defined as $q(x,y) \triangleq p(x) \cdot p(y|x)$, hence a pair of discrete random variables $(X,Y) \sim q(x,y)$, with X taking values in $\mathcal{X}$ and Y taking values in $\mathcal{Y}$. Of course, one has $X \sim p(x)$ and, for each $x \in \mathcal{X}$ and $y \in \mathcal{Y}$ s.t. $p(x) > 0$, $\Pr(Y = y|X = x) = p(y|x)$.

2.2 Bayesian hypothesis testing, min-entropy, leakage

Assume we are given a randomization mechanism $\mathcal{R} = (\mathcal{X}, \mathcal{Y}, p(\cdot|\cdot))$ and an a priori distribution $p(x)$ on $\mathcal{X}$. Assume an attacker wants to identify X on the basis of the observation Y, where, as explained above, $(X,Y) \sim p(x) \cdot p(y|x)$. This scenario can be formalized in terms of Bayesian hypothesis testing, as follows. The attacker's strategy is represented by a *guessing function* $g : \mathcal{Y} \to \mathcal{X}$. The *success probability after 1 observation* (relative to g) is defined as by

$$P_{succ}^{(g)} \triangleq \Pr(g(Y) = X)\,. \tag{1}$$

Correspondingly, the error probability is $P_e^{(g)} \triangleq 1 - P_{succ}^{(g)}$. It is well-known (see e.g. [12]) that optimal strategies, that is strategies maximizing the success probability, are those obeying the following *Maximum A Posteriori* (MAP) criterion: for each $y \in \mathcal{Y}$ and $x \in \mathcal{X}$ $g(y) = x$ implies $p(y|x)p(x) \geq p(y|x')p(x')$ $\forall x' \in \mathcal{X}$.In what follows, we shall always assume that g is MAP and consequently omit the superscript $^{(g)}$. The quantity P_{succ} admits a number of equivalent formulations. For example, it is straightforward to check that (cf. e.g. [26,5,6]; the sums below run over y of positive probability)

$$P_{succ} = \sum_y p(y) \max_x p(x|y) \tag{2}$$

$$= \sum_y \max_x p(y|x)p(x)\,. \tag{3}$$

Equation (2) shows clearly that P_{succ} results from an *average* over all observations $y \in \mathcal{Y}$. This equation also establishes a connection with Rényi's *min-entropy* [25]. This, for

a random variable $X \sim p(x)$, is defined thus (in the following, all the log's are taken with base 2): $H_\infty(X) \triangleq -\log \max_x p(x)$ .Conditional min-entropy of X given Y is defined as: $H_\infty(X|Y) \triangleq -\log \sum_y p(y) \max_x p(x|y)$ .Therefore from (2)

$$P_{succ} = 2^{-H_\infty(X|Y)} . \tag{4}$$

Success probability is the key to defining *information leakage* of X given Y. This quantity expresses, in bits, how much, on the average, one observation increases the success probability of the attacker. The intuition is that a gain of one bit of leakage corresponds to doubling the a priori success probability: $\mathcal{L}(X;Y) \triangleq H_\infty(X) - H_\infty(X|Y) = \log \frac{P_{succ}}{\max_x p(x)}$.

2.3 Asymptotic behavior

The scenario of a single observation generalizes to the case of several, say n, independent observations as follows. Given a prior $p(x)$ and fixed any $n \geq 0$, the adversary gets to know the observations corresponding to n independent executions of the mechanism $\mathcal{R}$, say $y^n = (y_1, ..., y_n) \in \mathcal{Y}^n$, throughout which the secret state x is kept fixed. Formally, the adversary knows a random vector of observations $Y^n = (Y_1, ..., Y_n)$ such that, for each $i = 1, ..., n$, Y_i is distributed like Y and the individual Y_i are *conditionally independent given* X. That is, the following equality holds true for each $y^n \in \mathcal{Y}^n$ and $x \in \mathcal{X}$ s.t. $p(x) > 0$ $\Pr(Y^n = (y_1, \ldots, y_n) \,|\, X = x) = \prod_{i=1}^n p(y_i|x)$ .We will often abbreviate the right-hand side of the last expression as $p(y^n|x)$. Again, for any n, the attacker's strategy is modeled by a guessing function $g : \mathcal{Y}^n \to \mathcal{X}$; the optimal strategy, that we will assume throughout the paper, is when g is MAP. The corresponding success and error probabilities, which depend on n, will be denoted by P^n_{succ} and P^n_e, respectively[1]. It is quite expected that, as $n \to +\infty$, $P^n_{succ} \to 1$, and this is indeed the case, under very mild conditions. What is important, though, is to characterize *how fast* the probability of success approaches 1. Intuitively, we want be able to determine an exponent $\rho \geq 0$ such that, for large n, $P^n_{succ} \approx 1 - 2^{-n\rho}$. To this purpose, we introduce some concepts in what follows.

Let $\{a_n\}_{n\geq 0}$ be a sequence of nonnegative reals. Assume that $\tau = \lim_{n\to+\infty} a_n$ exists and that $a_n \leq \tau$ for each n. We define the *rate* of $\{a_n\}_{n\geq 0}$ as follows:

$$\text{rate}(\{a_n\}) \triangleq \lim_{n\to+\infty} -\frac{1}{n} \log(\tau - a_n) \tag{5}$$

provided this limit exists[2]. When $\text{rate}(\{a_n\}) = \rho$ we also say that a_n *reaches* τ *at rate* ρ, and write this as $a_n \doteq \tau - 2^{-n\rho}$ .Intuitively, for large values on of n, this $\doteq$ can be interpreted as $\approx$. The above definition is modified as expected for the case when $a_n \geq \tau$ for each n: we set $\text{rate}(\{a_n\}) \triangleq \lim_{n\to+\infty} -\frac{1}{n} \log(a_n - \tau)$ and write $a_n \doteq \tau + 2^{-n\rho}$if $\rho = \text{rate}(\{a_n\})$. Note that we do allow $\text{rate}(\{a_n\}) = +\infty$, a case that arises for example when $\{a_n\}_{n\geq 0}$ is a constant sequence.

[1] For the case $n = 0$, we set for uniformity $y^n \triangleq \epsilon$ (empty tuple) and $p(\epsilon|x) \triangleq 1$. With this choice, $P^0_{succ} = \max_x p(x)$.

[2] More generally, we define the upper-rate (resp. lower-rate) $\overline{\text{rate}}(\{a_n\})$ (resp. $\underline{\text{rate}}(\{a_n\})$) by replacing the lim in (5) by lim sup (resp. lim inf).

The rate of growth of P^n_{succ} is given by *Chernoff Information*. Given two probability distributions p, q on $\mathcal{Y}$, we let their Chernoff Information be

$$C(p,q) \stackrel{\triangle}{=} -\min_{0\le\lambda\le 1} \log\Big(\sum_{y\in \mathrm{supp}(p)\cap\mathrm{supp}(q)} p^{\lambda}(y)q^{1-\lambda}(y)\Big) \tag{6}$$

where we stipulate that $C(p,q) = +\infty$ if $\mathrm{supp}(p) \cap \mathrm{supp}(q) = \emptyset$. Here $C(p,q)$ can be thought of as a sort of distance[3] between p and q: the more p and q are far apart, the the less observations are needed to discriminate between them. More precisely, assume we are in the binary case $\mathcal{X} = \{x_1, x_2\}$ (binary hypothesis testing) and let $p_i = p(\cdot|x_i)$ for $i = 1, 2$. Then a well-known result gives us the rate of convergence for the probabilities of success and error, with the proviso that $p(x_1) > 0$ and $p(x_2) > 0$ (cf. [12]): $P^n_{succ} \doteq 1-2^{-nC(p_1,p_2)}$ and $P^n_e \doteq 2^{-nC(p_1,p_2)}$(here we stipulate $2^{-\infty} = 0$). Note that this rate does not depend on the prior distribution $p(x)$ on $\{x_1, x_2\}$, but only on the probability distributions p_1 and p_2. This result extends to the general case $|\mathcal{X}| \ge 2$. Provided $\mathcal{R}$ is non-degenerate, it is enough to replace $C(p_1, p_2)$ by $\min_{x\neq x'} C(p_x, p_{x'})$, thus (see [5,20]):

$$P^n_{succ} \doteq 1 - 2^{-n \min_{x\neq x'} C(p_x,p_{x'})} \tag{7}$$

$$P^n_e \doteq 2^{-n \min_{x\neq x'} C(p_x,p_{x'})} \tag{8}$$

(with the understanding that, in the min, $p(x) > 0$ and $p(x') > 0$).

3 Semantic security of randomization mechanisms

We shall consider two scenarios. In the worst-case scenario, one is interested in the seriousness of a breach, independently of how much the breach is likely; this is also the scenario underlying differential privacy, which we will examine in Section 6. In the average-case scenario, one considers, so to speak, the seriousness of the breach averaged on the probability of the observed Y. In each scenario, our aim is to characterize when a randomization mechanism can be considered secure both in a semantic and in an operational fashion. We fix a generic randomization mechanism $\mathcal{R}$ for the rest of the section.

3.1 The worst-case scenario

In the worst-case definition, we compare the probability of predicates $Q \subseteq \mathcal{X}$ of the inputs, prior and posterior to one observation $y \in \mathcal{Y}$: a large variation in the posterior probability relative to any y implies a breach. Note that even the situation when the posterior probability is small compared to the prior is considered as dangerous, as it tells the adversary that $X \in Q^c$ is likely.

Definition 2 (worst-case breach). *Let $\epsilon \ge 0$. A* ϵ-breach *(*privacy breach of level ϵ*) for $\mathcal{R}$ is a subset $Q \subseteq \mathcal{X}$ such that for some a priori probability distribution $p(x)$ on $\mathcal{X}$, we have $p(Q) > 0$ and*

$$\max_{p(y)>0} |\log \frac{p(Q|y)}{p(Q)}| > \epsilon\,.$$

[3] Note that $C(p,q) = 0$ iff $p = q$ and that $C(p,q) = C(q,p)$. However $C(\cdot,\cdot)$ fails to satisfy the triangle inequality.

$\mathcal{R}$ *is* ϵ-secure *if it has no breach of level* ϵ*. The* security level *of* $\mathcal{R}$ *is defined as* $\epsilon_{\mathcal{R}} \triangleq \inf\{\epsilon \geq 0 : \mathcal{R}$ *is* ϵ*-secure*$\}$.

If $|\log \frac{p(Q|y)}{p(Q)}| > \epsilon$, we say *y causes a Q-breach of level* ϵ.

Remark 1. Note that the condition $\max_y |\log \frac{p(Q|y)}{p(Q)}| > \epsilon$ can be equivalently reformulated as $\max_y \max\{\frac{p(Q|y)}{p(Q)}, \frac{p(Q)}{p(Q|y)}\} > 2^{\epsilon}$.

For each $y \in \mathcal{Y}$, let $\pi_{M,y}$ and $\pi_{m,y}$ be the maximum and the minimum in the column y of the matrix $p(\cdot|\cdot)$, respectively. We give the following operational characterization of ϵ-security. A similar property (*amplification*) was considered as a sufficient condition for the absence of breaches in [15]. In the theorem below, we stipulate that $\frac{\pi_{M,y}}{\pi_{m,y}} = +\infty$ if $\pi_{M,y} > 0$ and $\pi_{m,y} = 0$.

Theorem 1. $\mathcal{R}$ *is* ϵ*-secure iff* $\log \max_y \frac{\pi_{M,y}}{\pi_{m,y}} \leq \epsilon$.

Example 1. The following example is inspired by [15]. The private information is represented by the set of integers $\mathcal{X} = \{0, ..., 5\}$, and $\mathcal{Y} = \mathcal{X}$. We consider a mechanism that replaces any $x \in \mathcal{X}$ by a number y that retains some information about the original x. More precisely, we let $Y = \lfloor X + \xi \rfloor \mod 6$, where with probability 0.5 ξ is a chosen uniformly at random in $\{-\frac{1}{2}, \frac{1}{2}\}$, and with probability 0.5 it is chosen uniformly at random in $\mathcal{X}$. We can easily compute the resulting conditional probability matrix.

$$\begin{pmatrix} 0.2500 & 0.2500 & 0.0833 & 0.0833 & 0.0833 & 0.2500 \\ 0.2500 & 0.2500 & 0.2500 & 0.0833 & 0.0833 & 0.0833 \\ 0.0833 & 0.2500 & 0.2500 & 0.2500 & 0.0833 & 0.0833 \\ 0.0833 & 0.0833 & 0.2500 & 0.2500 & 0.2500 & 0.0833 \\ 0.0833 & 0.0833 & 0.0833 & 0.2500 & 0.2500 & 0.2500 \\ 0.2500 & 0.0833 & 0.0833 & 0.0833 & 0.2500 & 0.2500 \end{pmatrix}.$$

The security level of this mechanism is $\epsilon_R = \log \frac{0.25}{0.083} = 3.0012$.

3.2 The average-case scenario

We want to asses the security of $\mathcal{R}$ by comparing the prior and posterior success probability for an adversary wanting to infer whether the secret is in Q or not after observing Y. This will give us an *average* measure of the seriousness of the breach induced by Q.

Fix a prior probability distribution $p(x)$ on $\mathcal{X}$. For every nonempty $Q \subseteq \mathcal{X}$, we shall denote by $\hat{Q}$ the binary random variable $I_Q(X)$, where $I_Q : \mathcal{X} \to \{0, 1\}$ is the indicator function of Q – in this notation, the dependence from $p(x)$ is left implicit, as $p(x)$ will always be clear from the context. An adversary, after observing Y, wants to determine whether it holds $\hat{Q}$ or $\hat{Q}^c$. This is a binary Bayesian hypothesis testing problem, which, as seen in Section 2, can be formulated in terms of min-entropy.

Definition 3 (Average-case breach). *Let* $\epsilon \geq 0$*. A* ϵ-A-breach (average case breach of level ϵ) *of* $\mathcal{R}$ *is a* $Q \subseteq \mathcal{X}$ *s.t. for some a priori distribution* $p(x)$ *on* $\mathcal{X}$*, we have that* $p(Q) > 0$ *and* $\mathcal{L}(\hat{Q}; Y) = H_\infty(\hat{Q}) - H_\infty(\hat{Q}|Y) > \epsilon$*.* $\mathcal{R}$ *is* ϵ-A-secure *if it has no average case breach of level* ϵ*. The* A-security level *of* $\mathcal{R}$ *is defined as* $\epsilon_{\mathcal{R}}^{A} \triangleq \inf\{\epsilon \geq 0 :$ $\mathcal{R}$ *is* ϵ*-A-secure*$\}$.

Of course, Y leaks at most one bit about the truth of Q: $0 \leq \mathcal{L}(\hat{Q}; Y) \leq 1$. In the next theorem, recall that for each $x \in \mathcal{X}$, $p_x(\cdot)$ denotes the distribution $p(\cdot|x)$.

Theorem 2. *Let* $l \triangleq \max_{x,x'} \|p_x - p_{x'}\|_1$ *and* $\epsilon \geq 0$. *Then* $\mathcal{R}$ *is* ϵ*-A-secure iff* $\log(\frac{l}{2} + 1) \leq \epsilon$.

4 Worst-case security vs. differential privacy

We first introduce DP, then discuss its relation to worst-case security. The definition of differential privacy relies on a notion of "neighborhood" between inputs of an underlying randomization mechanism. In the original formulation, two neighbors x and x' are two database instances that only differ by one entry. More generally, one can rely upon a notion of *adjacency*. An undirected graph is a pair (V, E) where V is a set of nodes and E is a set of unordered pairs $\{u, v\}$ with $u, v \in V$ and $u \neq v$. We also say that E is an adjacency relation on V and if $v \sim v'$ say v and v' are adjacent.

Definition 4 (differential privacy). *A* differentially private mechanism $\mathcal{D}$ *is a pair* $(\mathcal{R}, \sim)$ *where* $\mathcal{R} = (\mathcal{X}, \mathcal{Y}, p(\cdot|\cdot))$ *is a randomization mechanism and* $\sim$ *is an adjacency relation on* $\mathcal{X}$*, that is,* $(\mathcal{X}, \sim)$ *forms an undirected graph.*

Let $\epsilon > 0$. *We say* $\mathcal{D}$ *provides* ϵ-differential privacy *if for each* $x, x' \in \mathcal{X}$ *s.t.* $x \sim x'$*, it holds that for each* $y \in \mathcal{Y}$*:*

$$\max_y |\log \frac{p(y|x)}{p(y|x')}| \leq \epsilon \,. \tag{9}$$

Note that condition (9) is exactly that given in Theorem 1 to characterize worst-case privacy breaches, but limited to pairs of adjacent rows x and x'. This prompts the question of the exact relationship between the two notions. In the rest of the section, we will consider the standard domain $\mathcal{X} = \{0, 1\}^n$ of *databases*, corresponding to subsets of a given set of individuals $\{1, ..., n\}$. We deem two databases x, x' adjacent if they differ for the value of exactly one individual, that is if their Hamming distance is 1 [14,3,1]. Throughout the section, we let $\mathcal{D} = (\mathcal{R}, \sim)$ be a generic mechanism equipped with this $\mathcal{X}$ and this adjacency relation. Moreover, we will denote by Q_i ($i \in \{1, ..., n\}$) the set of databases $\{x \in \mathcal{X} \,|\, x_i = 1\}$, that is databases containing individual i.

The following theorem provides a precise characterization of (worst-case) ϵ-security in terms of privacy of individuals: interaction with the mechanism does not significantly change the belief about the participation of any individual to the database.

Theorem 3. $\mathcal{R}$ *satisfies* ϵ*-security iff for each* $i \in \{1, ..., n\}$ *and prior* $p(\cdot)$*,* Q_i *is not an* ϵ*-breach.*

Remark 2. The above theorem is of course still valid if one strengthens the "only if" part by requiring that *both* Q_i *and* Q_i^c are not ϵ-breach.

We proceed now by linking (worst-case) ϵ-security to ϵ-DP. The next result sets limits to the "arbitrariness" of background information against which DP offers guarantees: for example, it fails in some cases where an adversary has sufficient background information to rule out all possible databases but two, which are substantially different from each other.

Theorem 4. *If $\mathcal{R}$ satisfies ϵ-security then $\mathcal{D} = (\mathcal{R}, \sim)$ provides ϵ-DP. On the contrary, for each n there exist mechanisms providing ϵ-DP but not ϵ-security; in particular, these mechanisms exhibit Q_i-breaches ($i \in \{1, ..., n\}$) of level arbitrarily close to $n\epsilon > \epsilon$.*

Example 2. Let us consider the mechanism with input domain $\mathcal{X} = \{0, 1\}^2$, corresponding to the following matrix:

$$\begin{pmatrix} \frac{2}{3} & \frac{1}{6} & \frac{1}{12} & \frac{1}{64} & \frac{1}{48} & \frac{1}{48} \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{6} & \frac{1}{12} & \frac{1}{24} & \frac{1}{24} \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{12} & \frac{1}{12} \\ \frac{1}{12} & \frac{1}{12} & \frac{1}{6} & \frac{1}{3} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}.$$

This mechanism provides ϵ-DP with $\epsilon = 1$. However, it is not ϵ-secure, as e.g. $\frac{2/3}{1/12} = 8 > 2^{\epsilon}$.

We recover coincidence between ϵ-security and ϵ-DP if we confine ourselves to background knowledge that can be factorised as the product of independent measures over individuals. This provides another characterization of ϵ-DP. In what follows, for any $x \in \mathcal{X}$, we denote by $x_{\setminus i}$ the element of $\{0, 1\}^{n-1}$ obtained by removing the i-th component from x.

Theorem 5. *The following statements are equivalent:*

1. *$\mathcal{D}$ satisfies ϵ-DP;*
2. *for each $i \in \{1, ..., n\}$ and $p(x)$ of the form $p_i(x_i)q(x_{\setminus i})$, Q_i is not an ϵ-breach;*
3. *for each $p(x)$ of the form $\prod_{j=1}^{n} p_j(x_j)$ and for each $i \in \{1, ..., n\}$, Q_i is not an ϵ-breach.*

5 Asymptotic security

We assume that the attacker collects a tuple $y^n = (y_1, y_2, \ldots, y_n) \in \mathcal{Y}^n$ of observations generated i.i.d from the mechanism $\mathcal{R}$. We expect that, given Q, as n grows, the breach level approaches a threshold value. In order to the characterize synthetically the security of the randomization mechanism, though, it is important to characterize *how fast* this threshold is approached. Again, we distinguish a worst- from an average-case scenario and, for the rest of the section, fix a generic randomization mechanism $\mathcal{R}$.

5.1 Worst-case scenario

We begin with an obvious generalization of the notion of breach.

Definition 5 (n-breach of level ϵ). *A (n, ϵ)-privacy breach is a subset $Q \subseteq \mathcal{X}$ s.t. for some prior distribution $p(x)$ on $\mathcal{X}$, we have that $p(Q) > 0$ and*

$$\max_{p(y^n)>0} |\log \frac{p(Q|y^n)}{p(Q)}| > \epsilon\,.$$

The next proposition says that a notion of security based on bounding the level of n-breaches is not achievable. For the sake of simplicity, we shall discuss some of the following results in the case $\mathcal{R}$ is non-degenerate (all rows of the matrix are distinct).

Proposition 1. *Assume $\mathcal{R}$ is non-degenerate. For n large enough, $\mathcal{R}$ has n-breaches of arbitrary level. More explicitly, for any nonempty $Q \subseteq \mathcal{X}$ and any $\epsilon \geq 0$ there is a prior distribution $p(x)$ s.t. for any n large enough there is y^n ($p(y^n) > 0$) such that* $\log \frac{p(Q|y^n)}{p(Q)} > \epsilon$.

The above proposition suggests that, in the case of a large number of observations, worst-case analysis should focus on how fast $p(Q|y^n)$ can grow, rather than looking at the maximum level of a breach.

Definition 6 (rate of a breach). *Let $\rho \geq 0$. A* breach of rate ρ *is a subset $Q \subseteq \mathcal{X}$ such that there exist a prior distribution $p(x)$ on $\mathcal{X}$ with $p(Q) > 0$ and a sequence of n-tuples, $\{y^n\}_{n\geq 0}$, with $p(y^n) > 0$, such that $p(Q|y^n) \doteq 1 - 2^{-n\rho'}$ with $\rho' > \rho$. A randomization mechanism is* ρ-rate secure *if it has no privacy breach of rate ρ. The* rate security level *is defined as $\rho_{\mathcal{R}} \triangleq \inf\{\rho \geq 0 : \mathcal{R}$ is ρ-rate secure$\}$.*

Theorem 6. *$\mathcal{R}$ is ρ-rate secure iff $\rho \geq \log \max_y \frac{\pi_{M,y}}{\pi_{m,y}}$.*

The above theorem says that, for large n, the seriousness of the breach, for certain y^n, can be as bad as $\approx \log \dfrac{1 - (\frac{\pi_m}{\pi_M})^n}{p(Q)}$. The result, however, does not tell us *how likely* a serious breach is depending on n. The next result shows that the probability that *some* observable y^n causes a Q-breach grows exponentially fast. We premise some notation.

Fix a prior $p(x)$ over $\mathcal{X}$. Recall that we let $X \sim p(x)$ denote a random variable representing the secret information, and $Y^n = (Y_1, ..., Y_n)$ be the corresponding random vector of n observations, which are i.i.d. given X. Let us fix $Q \subseteq \mathcal{X}$ s.t. $p(Q) > 0$. Then $p(Q|Y^n)$ is a random variable. For any fixed $\epsilon > 0$, let us consider the two events

$$Breach_n^\epsilon \triangleq \{\frac{p(Q|Y^n)}{p(Q)} > 2^\epsilon\} \quad \text{and} \quad \overline{Breach}_n^\epsilon \triangleq \{\frac{p(Q)}{p(Q|Y^n)} > 2^\epsilon\}.$$

Clearly, the event $Breach_n^\epsilon \cup \overline{Breach}_n^\epsilon$ is the event that Y^n causes a Q-breach of level ϵ. As n grows, we expect that the probability of this event approaches 1 quite fast. The next theorem tells us exactly how fast.

Theorem 7. *Assume $\mathcal{R}$ is non-degenerate and strictly positive. Then, with the notation introduced above*

$$\Pr(Breach_n^\epsilon | X \in Q) \doteq 1 - 2^{-nC} \quad \textit{and} \quad \Pr(\overline{Breach}_n^\epsilon | X \in Q^c) \doteq 1 - 2^{-nC}$$

where $C = \min_{x\in Q, x'\in Q^c} C(p_x, p_{x'})$, with the understanding that x and x' in the min *are taken of positive probability. As a consequence, the probability that Y^n causes a Q-breach reaches* 1 *at rate at least C.*

5.2 Average-case scenario

It is straightforward to extend the definition of average-case breach to the case with multiple observations. For any nonempty subset $Q \subseteq \mathcal{X}$, and random variable $X \sim p(x)$,

s.t. $p(Q) > 0$, we consider $\hat{Q} = I_Q(X)$ and define the leakage imputable to Q after n observations as

$$\mathcal{L}^n(\hat{Q}; Y^n) \triangleq H_\infty(\hat{Q}) - H_\infty(\hat{Q}|Y^n).$$

An n-breach of level $\epsilon \geq 0$ is a Q such that $\mathcal{L}^n(\hat{Q}; Y^n) > \epsilon$. Recall from (4) that $P^n_{succ} = 2^{-H_\infty(\hat{Q}|Y^n)}$ is the success probability of guessing between $p(\cdot|Q)$ and $p(\cdot|Q^c)$ after observing Y^n. Provided $p(\cdot|Q) \neq p(\cdot|Q^c)$, (7) implies that, as $n \to +\infty$ we have $P^n_{succ} \to 1$, hence $\mathcal{L}^n(\hat{Q}; Y^n) \to -\log \max\{p(Q), 1 - p(Q)\}$. If $p(\cdot|Q) = p(\cdot|Q^c)$ then P^n_{succ} is constantly $\max\{p(Q), 1 - p(Q)\}$, so that the observations give no advantage to the attacker. These remarks suggest that, in the case of repeated observations, it is again important to characterize how fast $P^n_{succ} \to 1$.

Definition 7 (rate of a breach - average case). *Let $\rho \geq 0$. An* A-breach of rate ρ *is a subset $Q \subseteq \mathcal{X}$ such that for some prior distribution $p(x)$ on $\mathcal{X}$ with $p(Q) > 0$ one has that $P^n_{succ} \doteq 1 - 2^{-n\rho'}$, for some $\rho' > \rho$. A randomization mechanism is* ρ-rate A-secure *if it has no privacy breach of rate ρ. The* rate A-security level *is defined as* $\rho^A_{\mathcal{R}} \triangleq \inf\{\rho \geq 0 : \mathcal{R}$ *is ρ-rate A-secure*$\}$.

Now we can proceed with the following theorem.

Theorem 8. *$\mathcal{R}$ is ρ-rate A-secure iff* $\max_{x,x'} C(p_x, p_{x'}) \leq \rho$.

6 Utility

We next turn to the study of utility. In the rest of the section, we fix a mechanism $\mathcal{R}$ and a prior distribution $p(\cdot)$. Without any significant loss of generality, we shall assume that $\mathcal{R}$ is strictly positive and that $\mathrm{supp}(p) = \mathcal{X}$. Moreover, in this section, we shall work under the more general assumption that $\mathcal{Y}$ is *finite or denumerable*.

For any $n \geq 1$, we are now going to define the expected utility of $\mathcal{R}$, depending on user-specific belief, modeled as a prior $p(\cdot)$ on $\mathcal{X}$, and on function $loss : \mathcal{X} \times \mathcal{X} \to \mathbb{R}^+$. Here, $loss(x, x')$ represents the loss of a user who interprets the result of an observation of $\mathcal{R}$ as x', given that the real answer is x. For the sake of simplicity, we shall assume that $loss$ achieves a proper minimum when $x = x'$: for each $x \neq x'$, $loss(x, x) < loss(x, x')$. We also presuppose a *guessing function* $g : \mathcal{Y}^n \to \mathcal{X}$. The *expected utility* of $\mathcal{D}$ – relative to g – after n observations is in fact defined as an expected loss (the lower the better), thus

$$U_n \triangleq \sum_x p(x) \sum_{y^n} p(y^n|x) loss(x, g(y^n)). \tag{10}$$

Note that this definition coincides with that of Ghosh et al. [18] when one interprets our guessing function g as the *remap* considered in [18]. This is also the utility model of Alvim et al. [3], modulo the fact they only consider the 0/1-loss, or better, the complementary gain.

Example 3. When $\mathcal{Y}$ is a subset of the reals, legal $loss$ functions include the absolute value error $loss(x, x') = |x' - x|$ and the squared error $loss(x, x') = (x' - x)^2$. The binary loss function defined as 0 if $x = x'$ and 1 otherwise is another example: the resulting expected loss is just error probability, $U_n = P^n_e$.

It is quite easy to argue that, since a proper minimum in $loss(x, x')$ is reached when $x = x'$, the utility is maximized asymptotically when g respects the MAP criterion: $p(g(y^n)|y^n) \geq p(x|y^n)$ for each $x \in \mathcal{X}$. In what follows, we just assume that g is a fixed MAP function. Below, we study the behavior of utility in relation to differential privacy. The crucial quantity is

$$\rho_{\mathcal{R}} \stackrel{\triangle}{=} \min_{x,x' \in \mathcal{X}, p_x \neq p_{x'}} C(p_x, p_{x'}) \,. \tag{11}$$

We will show that the the asymptotic rate of utility is determined solely by $\rho_{\mathcal{R}}$. Note that this quantity does not depend on the user-defined *loss* function, nor on the prior $p(\cdot)$. For the sake of simplicity, below we discuss the result only in the case when $\mathcal{R}$ is non-degenerate[4].

Remark 3. We note that the formula (6) for Chernoff Information extends to the case when $p(\cdot)$ and $q(\cdot)$ have the same *denumerable* support.

Theorem 9. *Assume $\mathcal{R}$ is non-degenerate. Then $U_n \doteq U_{\mathcal{R}} + 2^{-n\rho_{\mathcal{R}}}$, where $U_{\mathcal{R}} \stackrel{\triangle}{=} \sum_x p(x) loss(x, x)$.*

Having established the centrality of $\rho_{\mathcal{R}}$ in the asymptotic behavior of utility, we now discuss the relationship of this quantity with the worst-case security level ϵ provided by the mechanism. The first result provides us with a simple, general bound relating $\rho_{\mathcal{R}}$ and ϵ.

Theorem 10. *Assume $\mathcal{R}$ is worst-case ϵ-secure. Then $\rho_{\mathcal{R}} \leq \epsilon$. The same conclusion holds if $\mathcal{D} = (\mathcal{R}, \sim)$ provides ϵ-DP.*

In what follows, we will obtain more precise results relating ϵ to the utility rate $\rho_{\mathcal{R}}$ in the case of a class of mechanisms providing ϵ-DP. Specifically, we will consider mechanisms with a finite input domain $\mathcal{X} = \{0, 1, \ldots, N\}$, a denumerable $\mathcal{Y} = \mathbb{Z}$ and a conditional probability matrix of the form $p_i(j) = Mc^{|i-j|}$, for some positive $c < 1$. This class of mechanisms includes the geometric mechanism (a discrete variant of the Laplacian mechanism, see [18]) and also a version extended to $\mathbb{Z}$ of the optimal mechanism considered by Alvim et al. [3].

Theorem 11. *Let $\mathcal{R}$ be a mechanism as described above. Then* $\rho_{\mathcal{R}} = \log(1+c) - \frac{1}{2}\log c - 1$.

Remark 4. The geometric mechanism is obtained by equipping the above described mechanism with the the line topology over $\mathcal{X} = \{0, ..., N\}$: $i \sim j$ iff $d_{ij} \stackrel{\triangle}{=} |i - j| = 1$. This is the topology for counting queries in "oblivious" mechanisms, for example. If we set $c = 2^{-\epsilon}$, then this mechanism provides ϵ-DP. The above theorem tells us that in this case $\rho_{\mathcal{R}} = \frac{\epsilon}{2} + \log \frac{1+2^{-\epsilon}}{2}$. By setting e.g. $\epsilon = 1$, one gets $\rho_{\mathcal{R}} \approx 0.085$.

For any mechanism $\mathcal{R}$ with input $\mathcal{X} = \{0, ..., N\}$ and output $\mathcal{Y} = \mathbb{Z}$, we can consider the corresponding *truncated* mechanism $\mathcal{R}'$: it has $\mathcal{X} = \mathcal{Y} = \{0, 1, \ldots, N\}$ and its matrix is obtained from $\mathcal{R}$'s by summing all the columns $y < 0$ to column $y = 0$, and all the columns $y > N$ to column $y = N$.

[4] The result carries over to the general case, at the cost of some notational burden: one has to replace $U_{\mathcal{R}}$ with a more complicated expression.

Corollary 1. *Assume $\mathcal{R}'$ is the truncated version of a mechanism $\mathcal{R}$. Then $\rho_{\mathcal{R}'} < \rho_{\mathcal{R}}$.*

In the case of a single observation case, treated by Ghosh et al. [18], there is no substantial difference between the geometric mechanism and the truncated geometric one. Corollary 1 shows that the situation is different in the case with repeated observations.

7 Conclusion and further work

We have analyzed security of randomization mechanisms against privacy breaches with respect to various dimensions (worst vs. average case, single vs. repeated observations, utility). Whenever appropriate, we have characterized the resulting security measures in terms of simple row-distance properties of the underlying channel matrix. We have clarified the relation our worst-case measure with DP.

A problem left open by our study is the exact relationship between our average-case security notion and the maximum leakage considered in QIF – see e.g. [19]. We would also like to apply and possibly extend the results of the present paper to the setting of de-anonymization attacks on dataset containing micro-data. [23] has shown that the effectiveness of these attacks depends on certain features of sparsity and similarity of the dataset, which roughly quantify how difficult it is to find two rows of the dataset that are similar. The problem can be formalized in terms of randomization mechanisms with repeated observations – see [6] for some preliminary results on this aspect. Then the row-distance measures considered in the present paper appear to be strongly related to the notion of similarity, and might play a crucial in the formulation of a robust definition of dataset security.

References

1. M. S. Alvim, M. E. Andrés, K. Chatzikokolakis, P. Degano, C. Palamidessi. Differential Privacy: on the trade-off between Utility and Information Leakage. In: *FAST* 2011, *LNCS*, 7140,2011.
2. M. S. Alvim, M. E. Andrés, K. Chatzikokolakis, C. Palamidessi. Quantitative Information Flow and Applications to Differential Privacy. *FOSAD* VI, *LNCS* 6858: 211-230, 2011.
3. M. S. Alvim, M. E. Andrés, K. Chatzikokolakis, C. Palamidessi. On the relation between Differential Privacy and Quantitative Information Flow. *ICALP* (2) 2011: 60-76, 2011.
4. G. Barthe, B. Köpf. Information-theoretic Bounds for Differentially Private Mechanisms. *In 24rd IEEE Computer Security Foundations Symposium, CSF* 2011, 191-204, 2011. *IEEE Computer Society.*
5. M. Boreale, F. Pampaloni, M. Paolini. Asymptotic information leakage under one-try attacks. *FoSSaCS* 2011, *LNCS* 6604:396-410, 2011. Full version to appear on *MSCS* available at `http://rap.dsi.unifi.it/~boreale/Asympt.pdf`.
6. M. Boreale, F. Pampaloni, M. Paolini. Quantitative Information Flow, with a View. *ESORICS 2011*, *LNCS* 6879:588-604, 2011.
7. M. Boreale, M. Paolini. Worst- and average-case privacy breaches in randomization mechanisms. Full version of the present paper. Available at `http://rap.dsi.unifi.it/~boreale/FullBreach.pdf`.
8. C. Braun, K. Chatzikokolakis, C. Palamidessi. Quantitative Notions of Leakage for One-try Attacks. *Proc. of MFPS 2009*, *Electr. Notes Theor. Comput. Sci.* 249: 75-91, 2009.

9. K. Chatzikokolakis, C. Palamidessi, P. Panangaden. Anonymity protocols as noisy channels. *Information and Computation* 206(2-4): 378-401, 2008.
10. K. Chatzikokolakis, C. Palamidessi, P. Panangaden. On the Bayes risk in information-hiding protocols. *Journal of Computer Security* 16(5): 531-571, 2008.
11. D. Chaum. The Dining Cryptographers Problem: Unconditional Sender and Recipient Untraceability. *Journal of Cryptology* 1 (1): 65-75, 1988.
12. T. M. Cover, J. A. Thomas. *Elements of Information Theory, 2/e*. John Wiley Sons, 2006.
13. C. Dwork. Differential Privacy. *ICALP* 2006. *LNCS*, 4052: 1-12, 2006.
14. C. Dwork, F. McSherry, K. Nissim, A. Smith. Calibrating Noise to Sensitivity in Private Data Analysis. *Proc. of the 3rd IACR Theory of Cryptography Conference*, 2006.
15. A. Evfimievski, J. Gehrke, R. Srikant. Limiting Privacy Breaches in Privacy Preserving Data Mining. *Proc. of the ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, 2003.
16. A. Friedman, A. Shuster. Data Mining with Differential Privacy. Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD), 2010.
17. S. R. Ganta, S. P. Kasiviswanathan, A. Smith. Composition Attacks and Auxiliary Information in Data Privacy. *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD)*, 2008.
18. A. Ghosh, T. Roughgarden, M. Sundararajan. Universally utility-maximizing privacy mechanisms. *In STOC* 2009, 351-360, 2009.
19. B. Köpf, G. Smith. Vulnerability Bounds and Leakage Resilience of Blinded Cryptography under Timing Attacks. *CSF 2010*: 44-56, 2010.
20. C.C. Leang, D.H. Johnson. On the asymptotics of M-hypothesis Bayesian detection. *IEEE Transactions on Information Theory* 43: 280-282, 1997.
21. F. McSherry. Privacy Integrated Queries. *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data (SIGMOD)* 2009.
22. F. McSherry, K. Talwar. Mechanism Design via Differential Privacy. *Proceedings Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, 2007.
23. A. Narayanan, V. Shmatikov. Robust De-anonymization of Large Sparse Datasets. *Proc. of IEEE Symposium on Security and Privacy*, 2008.
24. M. K. Reiter, A. D. Rubin. Crowds: Anonymity for Web Transactions. *ACM Trans. Inf. Syst. Secur.* 1(1): 66-92, 1998.
25. A. Rényi. On Measures of Entropy and Information. In *Proc. of the 4th Berkeley Symposium on Mathematics, Statistics, and Probability.* (1961) 547-561.
26. G. Smith. On the Foundations of Quantitative Information Flow. *FoSSaCS 2009, LNCS* 5504: 288-302, 2009.