Theoretical Computer Science 597 (2015) 40–61

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

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

Michele Boreale [a,*], Michela Paolini [b]

[a] Università di Firenze, Dipartimento di Statistica, Informatica, Applicazioni (DiSIA), Italy
[b] Gung s.r.l., Firenze, Italy

ARTICLE INFO

Article history:
Received 10 July 2013
Received in revised form 15 April 2015
Accepted 8 July 2015
Available online 15 July 2015
Communicated by B.P.F. Jacobs

Keywords:
Foundations of security
Quantitative information flow
Differential privacy
Utility
Information theory

ABSTRACT

In a variety of contexts, randomization is regarded as an effective technique to conceal sensitive information. Viewing randomization mechanisms as information-theoretic channels, we start from a semantic notion of security, which 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 factorized 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.

1. Introduction

In a variety of contexts, randomization is regarded as an effective means to conceal sensitive information. For example, anonymity protocols like Crowds [31] or the Dining Cryptographers [12] 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. [18] and references therein).

☆ Extended and revised version of [8]. Work partially supported by the Italian PRIN project CINA. Work done while the second author was at IMT – Institute for Advanced Studies – Piazza S. Ponziano 6, I-55100 Lucca.

* Corresponding author at: Università di Firenze, Dipartimento di Statistica, Informatica, Applicazioni (DiSIA), Viale Morgagni 65, I-50134 Firenze, Italy.
E-mail address: michele.boreale@unifi.it (M. Boreale).

http://dx.doi.org/10.1016/j.tcs.2015.07.010
0304-3975/

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) [9,25,6,7,10,11,33] and *Differential Privacy* (DP) [15,16,28,29,19–23]. 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 recently researchers have begun investigating the relations between these two notions [1–4].

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. [18]. 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 following terms [15]: *Regardless of external knowledge, an adversary with access to the sanitized database draws the same conclusions, whether or not my data is included.* A formalization along these lines is in [20]. We investigate how this kind of resistance 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 example, in a sensor networks scenario, an attacker can easily gather multiple observations related to an individual [27]. The same is true in a scenario of anonymity networks comprising corrupted nodes. For another example, in a de-anonymization scenario similar to [30], [7] shows that gathering information about a target individual can be modeled as collecting multiple observations from a certain randomization mechanism. Again, an online, randomized data-releasing mechanism might offer users the possibility of asking the same query a number of times, thus potentially allowing an adversary to remove enough noise to learn valuable information about the secret. This is an instance of the *composition* attacks 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. [28,19]. 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 the related work paragraph. Here, we are interested in studying the growth of expected utility in the model of Ghosh et al. [21] 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 factorizes 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 [21].