Differential Privacy: a Study of Utility and Min-Entropy Leakage*

Mário S. Alvim, Miguel E. Andrés,
Konstantinos Chatzikokolakis, and Catuscia Palamidessi

INRIA and LIX, Ecole Polytechnique, France.

Abstract. Differential privacy is a notion that has emerged in the community of statistical databases, as a response to the problem of protecting the privacy of the database's participants when performing statistical queries. The idea is that a randomized query satisfies differential privacy if the likelihood of obtaining a certain answer for a database x is not too different from the likelihood of obtaining the same answer on adjacent databases, i.e. databases which differ from x for only one individual.
Information flow is an area of Security concerned with the problem of controlling the leakage of confidential information in programs and protocols. Nowadays, one of the most established approaches to quantify and to reason about leakage is based on the Rényi min-entropy version of information theory, also referred to as min-entropy leakage.
In this work we analyze critically the notion of differential privacy in light of the conceptual framework which has been developed for min-entropy leakage. We show that there is a close relation between differential privacy and leakage, due to the graph symmetries induced by the adjacency relation. Furthermore, we consider the utility of the randomized answer, which measures its expected degree of accuracy. We focus on certain kinds of utility functions called "binary", which have a close correspondence with the min-entropy leakage. Again, it turns out that there can be a tight correspondence between differential privacy and utility, depending on the symmetries induced by the adjacency relation and by the query. Depending on these symmetries we can also build an optimal-utility randomization mechanism while preserving the required level of differential privacy. Our main contribution is a study of the kind of structures that can be induced by the adjacency relation and the query, and how to use them to derive bounds on the leakage and achieve the optimal utility.

1 Summary

This summary is based on [1] and [2]. We refer to these works for more details.

Databases are commonly used for obtaining statistical information about their participants. Simple examples of statistical queries are, for instance, the

* This work has been partially supported by the project ANR-09-BLAN-0169-01 PANDA and by the INRIA DRI Equipe Associée PRINTEMPS. The work of Miguel E. Andrés has been supported by the LIX-Qualcomm postdoc fellowship 2010.

predominant disease of a certain population, or the average salary. The fact that the answer is publicly available, however, constitutes a threat for the privacy of the individuals.

In order to illustrate the problem, consider a set of individuals *Ind* whose attribute of interest[1] has values in *Val*. A particular database is formed by a subset of *Ind*, where a certain value in *Val* is associated to each participant. A query is a function $f : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$ is the set of all possible databases, and $\mathcal{Y}$ is the domain of the answers.

For example, let *Val* be the set of possible salaries and let f represent the query "what is the average salary of the participants in the database". In principle we would like to consider the *global information* relative to a database x as *public*, and the *individual information* about a participant i as *private*. Namely, we would like to be able to obtain $f(x)$ without being able to infer the salary of i. However, this is not always possible. In particular, if the number of participants in x is known (say n), then the removal of i from the database would allow to infer i's salary by querying again the new database x', and by applying the formula $n\,f(x) - (n-1)\,f(x')$. Using an analogous reasoning we can argue that not only the removal, but also the addition of an individual is a threat for his privacy.

Another kind of private information we may want to protect is whether an individual i is participating or not in a database. In this case, if we know for instance that i earns, say $5K$ Euros/month, and all the other individuals in *Ind* earn less than $4K$ Euros/month, then knowing that $f(x) > 4K$ Euros/month will reveal immediately that i is in the database x.

A common solution to the above problems is to introduce some output perturbation mechanism based on *randomization*: instead of the exact answer $f(x)$ we report a "noisy" answer. Namely, we use some randomized function $\mathcal{K}$ which produces values in some domain[2] $\mathcal{Z}$ according to some probability distribution that depends on the input $x \in \mathcal{X}$. The general scheme for a randomized function $\mathcal{K}$ is shown in Figure 1(a).

An important class of randomized functions are the *oblivious* ones. A randomized function $\mathcal{K}$ is oblivious if its probability distribution depends only on the answer to the query, and not on the database. In that case we can decompose $\mathcal{K}$ into the query f that maps databases in $\mathcal{X}$ to real answers in $\mathcal{Y}$, and the randomization mechanism $\mathcal{H}$ that maps real answers in $\mathcal{Y}$ to reported answers in $\mathcal{Z}$. Figure 1(b) shows this scheme.

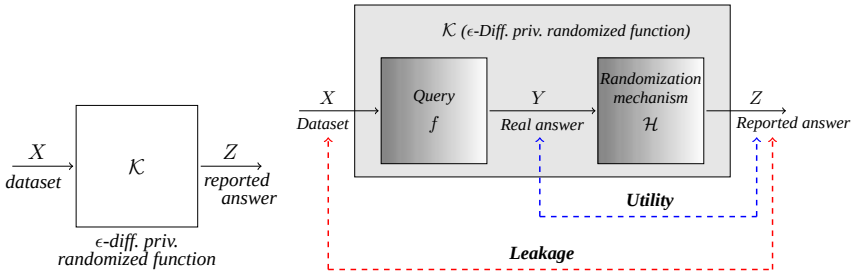
Of course for certain distributions it may still be possible to guess the value of an individual with a high probability of success. The notion of *differential privacy*, due to Dwork [8, 11, 9, 10], is a proposal to control the risk of violating privacy for both kinds of threats described above (value and participation). The idea is to say that $\mathcal{K}$ satisfies ϵ-differential privacy (for some $\epsilon > 0$) if the ratio

[1] In general we could be interested in several attributes simultaneously, and in this case *Val* would be a set of tuples.

[2] The new domain $\mathcal{Z}$ may coincide with $\mathcal{Y}$, but not necessarily. It depends on how the randomization mechanism is defined.

between the probabilities that two adjacent databases give the same answer is bound by e^{ϵ}, where by "adjacent" we mean that the databases differ for only one individual (either for the value of an individual or for the presence/absence of an individual). Often we will abbreviate "ϵ-differential privacy" as ϵ-d.p.

Obviously, the smaller is ϵ, the greater is the privacy protection. In particular, when ϵ is close to 0 the output of $\mathcal{K}$ is nearly independent from the input (all distributions are almost equal). Unfortunately, such $\mathcal{K}$ is practically useless. The *utility*, i.e. the capability to retrieve accurate answers from the reported ones, is the other important characteristic of $\mathcal{K}$, and it is clear that there is a trade-off between utility and privacy. On the other hand, these two notions are not the complete opposite of each other, because utility concerns the relation between the reported answer and the real answer, while privacy is concerns the relation between the reported answer and the information in the database, as shown in Figure 1(b). This asymmetry makes more interesting the problem of finding a good compromise between the two.



(a) Scheme of a general randomized function

(b) Scheme of an oblivious randomized function $\mathcal{K}$

Fig. 1. Different types of randomized functions $\mathcal{K}$

At this point, we would like to remark an intriguing analogy between the area of differential privacy and that of *quantitative information flow* (QIF), both in the motivations and in the basic conceptual framework. Information flow is concerned with the leakage of secret information through computer systems, and the attribute "quantitative" refers to the fact that we are interested in measuring the amount of leakage, not just its occurrence. One of the most established approaches to QIF is based on information theory: the idea is that a system is seen as a channel in the information-theoretic sense, where the secret is the input and the observables are the output. The entropy of the input represents its vulnerability, i.e. how easy it is for an attacher to guess the secret. We distinguish between the *a priori* entropy (before the observable) and the *a posteriori* entropy (given the observable). The difference between the two gives the *mutual*

information and represents, intuitively, the increase in vulnerability due to the observables produced by the system, so it is naturally considered as a measure of the leakage. The notion of entropy is related to the kind of attack we want to model, and in this work we focus on min-entropy [14], which represents the so-called *one-try attacks*. In recent years there has been a lot of research aimed at establishing the foundations of this framework [15, 6, 12, 3, 4]. It is worth pointing out that the a posteriori min-entropy corresponds to the concept of Bayes risk, which has also been proposed as a measure of the effectiveness of attacks [7, 5, 13].

The analogy hinted above between differential privacy and QIF is based on the following observations: at the motivational level, the concern about privacy is akin the concern about information leakage. At the conceptual level, the randomized function $\mathcal{K}$ can be seen as an information-theoretic channel, and the limit case of $\epsilon = 0$, for which the privacy protection is total, corresponds to a 0-capacity channel[3] (the rows of the channel matrix are all identical), which does not allow any leakage. Another promising similarity is that the notion of utility (in the binary case) corresponds closely to the Bayes risk.

In this work we investigate the notion of differential privacy, and its implications, in light of the min-entropy leakage theoretic framework developed for QIF. In particular, we wish to explore the following natural questions:

1. Does ϵ-d.p. induce a bound on the min-entropy leakage of $\mathcal{K}$?
2. Does ϵ-d.p. induce a bound on the min-entropy leakage *relative to an individual*?
3. Does ϵ-d.p. induce a bound on the utility?
4. Given f and ϵ, can we construct a $\mathcal{K}$ which satisfies ϵ-d.p. and maximum utility?

We will see that the answers to (1) and (2) are positive, and we provide bounds that are tight, in the sense that for every ϵ there is a $\mathcal{K}$ whose leakage reaches the bound. For (3) we are able to give a tight bound in some cases which depend on the structure of the query, and for the same cases, we are able to construct an oblivious $\mathcal{K}$ with maximum utility, as requested by (4).

The above results appear in [1], and are based on techniques which exploit the graph structure that the adjacency relation induces on the domain of all databases $\mathcal{X}$, and on the domain of the correct answers $\mathcal{Y}$. An an extension of those techniques, and a coherent graph-theoretic framework for reasoning about the symmetries of those domains can be found in [2]. More specifically, in the later work:

- We explore the graph-theoretic foundations of the adjacency relation, and point out various types of symmetries which allow us to establish a strict link between differential privacy and information leakage.
- We give a tight bound for the question (2) above, strictly smaller than the one in [1].

[3] The channel capacity is the maximum mutual information over all possible input distributions.

– We extend the structures for which we give a positive answer to the questions (3) and (4) above. In [1] the only case considered was the class of graphs with single-orbit automorphisms. Here we show that the results hold also for regular-distance graphs and a variant of vertex-transtive graphs.

In these works we focus on the case in which $\mathcal{X}$, $\mathcal{Y}$ and $\mathcal{Z}$ are finite, leaving the more general case for future work.

2 Future work

As future work, we plan to extend our result to other kinds of utility functions. In particular, we are interested in the case in which the the answer domain is provided with a metric, and we are interested in taking into account the degree of accuracy of the inferred answer.

References

1. Mário S. Alvim, Miguel E. Andrés, Konstantinos Chatzikokolakis, Pierpaolo Degano, and Catuscia Palamidessi. Differential privacy: on the trade-off between utility and information leakage. Technical report, 2011. `http://hal.inria.fr/inria-00580122/en/`.
2. Mário S. Alvim, Miguel E. Andrés, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. On the relation between differential privacy and quantitative information flow. In *Proceedings of ICALP*, 2011. To appear.
3. Miguel E. Andrés, Catuscia Palamidessi, Peter van Rossum, and Geoffrey Smith. Computing the leakage of information-hiding systems. In *Proc. of TACAS*, volume 6015 of *LNCS*, pages 373–389. Springer, 2010.
4. Michele Boreale, Francesca Pampaloni, and Michela Paolini. Asymptotic information leakage under one-try attacks. In *Proc. of FOSSACS*, volume 6604 of *LNCS*, pages 396–410. Springer, 2011.
5. Christelle Braun, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Compositional methods for information-hiding. In *Proc. of FOSSACS*, volume 4962 of *LNCS*, pages 443–457. Springer, 2008.
6. Christelle Braun, Konstantinos Chatzikokolakis, and Catuscia Palamidessi. Quantitative notions of leakage for one-try attacks. In *Proc. of MFPS*, volume 249 of *ENTCS*, pages 75–91. Elsevier, 2009.
7. Konstantinos Chatzikokolakis, Catuscia Palamidessi, and Prakash Panangaden. Probability of error in information-hiding protocols. In *Proc. of CSF*, pages 341–354. IEEE, 2007.
8. Cynthia Dwork. Differential privacy. In *Automata, Languages and Programming, 33rd Int. Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proc., Part II*, volume 4052 of *LNCS*, pages 1–12. Springer, 2006.
9. Cynthia Dwork. Differential privacy in new settings. In *Proc. of the Twenty-First Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2010, Austin, Texas, USA, January 17-19, 2010*, pages 174–183. SIAM, 2010.
10. Cynthia Dwork. A firm foundation for private data analysis. *Communications of the ACM*, 54(1):86–96, 2011.

11. Cynthia Dwork and Jing Lei. Differential privacy and robust statistics. In *Proc. of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*, pages 371–380. ACM, 2009.
12. Boris Köpf and Geoffrey Smith. Vulnerability bounds and leakage resilience of blinded cryptography under timing attacks. In *Proc. of CSF*, pages 44–56. IEEE, 2010.
13. Annabelle McIver, Larissa Meinicke, and Carroll Morgan. Compositional closure for bayes risk in probabilistic noninterference. In *Proc. of ICALP*, volume 6199 of *LNCS*, pages 223–235. Springer, 2010.
14. Alfréd Rényi. On Measures of Entropy and Information. In *Proc. of the 4th Berkeley Symposium on Mathematics, Statistics, and Probability*, pages 547–561, 1961.
15. Geoffrey Smith. On the foundations of quantitative information flow. In *Proc. of FOSSACS*, volume 5504 of *LNCS*, pages 288–302. Springer, 2009.