The Promise of Differential Privacy
A Tutorial on Algorithmic Techniques

Cynthia Dwork

Abstract— *Differential privacy* describes a promise, made by a data curator to a data subject: you will not be affected, adversely or otherwise, by allowing your data to be used in any study, no matter what other studies, data sets, or information from other sources is available. At their best, differentially private database mechanisms can make confidential data widely available for accurate data analysis, without resorting to data clean rooms, institutional review boards, data usage agreements, restricted views, or data protection plans. To enjoy the fruits of the research described in this tutorial, the data analyst must accept that raw data can never be accessed directly and that eventually data utility is consumed: overly accurate answers to too many questions will destroy privacy. The goal of algorithmic research on differential privacy is to postpone this inevitability as long as possible.

Privacy is a charged term meaning different things to different people, and even different things to the same person, according to the context. In the digital information realm, loss of privacy is usually associated with failure to control access to information, to control the flow of information, or to control the purposes for which information is employed. *Differential privacy* arose in a context in which ensuring privacy is a challenge even if all these control problems are solved: privacy-preserving statistical analysis of data.

Privacy in data analysis is treated in the scholarly literature of many fields – statistics, databases, philosophy, law, cryptography, and theoretical computer science (see [4] and the references therein). The popularly known privacy breaks, such as the identification of the medical records of the then governor of Massachussetts in public "anonymized" medical encounter data [12], the identification of the search history of Thelma Arnold in public "anonymized" AOL query records [1], the identification of a homophobic individual in the public "anonymized" Netflix prize training data set [11], and the theorectical test for membership of the DNA of a given individual in a forensic mix or a genome-wide association study [10], were not breaks of the implementation of a privacy definition. Rather, the stated privacy goals were themselves inadequate: syntactic, *ad hoc*, and unable to cope with information from sources *other than* the database.

Differential privacy was inspired by the profound definitional work in modern cryptography [8], [9]. These cited works, and the tradition to which they gave rise, transformed the cycle of propose-break-propose-again to a path of progress in cryptography.

Microsoft Research, Silicon Valley, Mountain View, CA 94043. E-mail: dwork@microsoft.com.

Since the data analyst and the adversary are the same party, formulating a privacy goal in analogy to semantic security [8] cannot work [2], [7]. In the context of data analysis the goal is to protect the participants from harm. As the following parable shows, the challenge is to distentangle harm from utility.

> Analysis of a given data set teaches us that smoking causes cancer. Mary, a smoker, is harmed by this analysis: her insurance premiums rise. *Mary's premiums rise whether or not her data are in the data set.* In other words, Mary is harmed by the finding "smoking causes cancer," and not by her participation in the data set. Of course, Mary is also helped: having learned that smoking causes cancer, Mary enters a smoking cessation program.

Differential privacy aims to ensure that the *only* harms encounterd by Mary are the harms suffered from the conclusions of the analyses. These conclusions can also be helpful to her, which is the whole point of a medical study.

A database is modeled as a collection of *rows*, with each row containing the data of a different individual. Differential privacy will ensure that the ability of an adversary to inflict harm (or good, for that matter) – of any sort, to any set of people – should be essentially the same, independent of whether any individual opts in to, or opts out of, the dataset. This is done indirectly, simultaneously addressing all possible forms of harm and good, by focusing on the probability of any given output of a privacy mechanism and how this probability can change with the addition or deletion of any row. Thus, we concentrate on pairs of databases (D, D') differing only in one row, meaning one is a subset of the other and the larger database contains just one additional row. Finally, to handle worst case pairs of databases, the probabilities will be over the random choices made by the privacy mechanism.

Definition 0.1. [2], [6] *A randomized function $\mathcal{K}$ gives* ε-differential privacy *if for all data sets D and D' differing on at most one row, and all $S \subseteq Range(\mathcal{K})$,*

$$\Pr[\mathcal{K}(D) \in S] \quad \leq \quad \exp(\varepsilon) \times \Pr[\mathcal{K}(D') \in S],$$

where the probability space in each case is over the coin flips of $\mathcal{K}$.

The multiplicative nature of the guarantee implies that an output whose probability is zero on a given database must

also have probability zero on any neighboring database, and hence, by repeated application of the definition, on any other database. This rules out direct viewing of raw data.

Any data access mechanism satisfying this definition addresses all concerns one might have about the leakage of her personal information, regardless of any auxiliary information – other databases, newspapers, websites, and so on – known to an adversary: even if the participant removed her data from the data set, no outputs (and thus consequences of outputs) would become significantly more or less likely. For example, if the database were to be consulted by an insurance provider before deciding whether or not to insure a given individual, then the presence or absence of *any* individual's data in the database will not significantly affect her chance of receiving coverage.

In all differentially private analyses there is a tension between minimizing privacy loss and maximizing utility. The former is measured via the parameter ε; the latter may be measured in various ways, including, for example, L_1 loss, L_2 loss, predictive accuracy, and sample complexity. The choice of ε is a social question complicated by the ways in which privacy loss accumulates over a lifetime of exposure to multiple analyses, as well as membership in multiple databases. That is, the choice of ε also depends on the behavior of differentially private algorithms under composition.

A differential privacy overview appears in [4]. See [3], [5] for video presentations providing additional motivation for the definition of differential privacy. The current tutorial focuses on algorithmic techniques for achieving differential privacy and the behavior of differential privacy under composition. The tutorial closes with a discussion of directions for future research. Slides for the covered material are available at research.microsoft.com/en-us/projects/DatabasePrivacy/, which also contains several surveys, research papers, and links to relevant pages.

REFERENCES

[1] M. Barbaro and T. Z. Jr., "A face is exposed for aol searcher no. 441779," 8/9/2006.

[2] C. Dwork, "Differential privacy," in *Proceedings of the 33rd International Colloquium on Automata, Languages and Programming (ICALP)(2)*, 2006, pp. 1–12.

[3] ——, "I'm in the database, but nobody knows," 2010, cyber.law.harvard.edu/interactive/events/luncheon/2010/09/dwork.

[4] ——, "A firm foundation for private data analysis," *Communications of the ACM*, vol. 54, 2011.

[5] ——, "A firm foundation for private data analysis (video)," 2011, www.scivee.tv/node/26354.

[6] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proceedings of the 3rd Theory of Cryptography Conference*, 2006, pp. 265–284.

[7] C. Dwork and M. Naor, "On the difficulties of disclosure prevention in statistical databases or the case for differential privacy," *Journal of Privacy and Confidentiality*, vol. 2, 2010.

[8] S. Goldwasser and S. Micali, "Probabilistic encryption," *JCSS*, vol. 28, pp. 270–299, 1984.

[9] S. Goldwasser, S. Micali, and R. Rivest, "A digital signature scheme secure against adaptive chosen-message attacks," *SIAM J. Comput*, vol. 17, pp. 281–308, 1988.

[10] N. Homer, S. Szelinger, M. Redman, D. Duggan, W. Tembe, J. Muehling, J. Pearson, D. Stephan, S. Nelson, and D. Craig, "Resolving individuals contributing trace amounts of dna to highly complex mixtures using high-density snp genotyping microarrays," *PLoS Genet*, vol. 4, 2008.

[11] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets (how to break anonymity of the netflix prize dataset)," in *Proc. 29th IEEE Symposium on Security and Privacy*, 2008.

[12] L. Sweeney, "Weaving technology and policy together to maintain confidentiality," *J. Law Med. Ethics*, vol. 25, pp. 98–110, 1997.