Differential Privacy: A Survey of Results

Cynthia Dwork

Microsoft Research
dwork@microsoft.com

Abstract. Over the past five years a new approach to privacy-preserving data analysis has born fruit [13, 18, 7, 19, 5, 37, 35, 8, 32]. This approach differs from much (but not all!) of the related literature in the statistics, databases, theory, and cryptography communities, in that a formal and *ad omnia* privacy guarantee is defined, and the data analysis techniques presented are rigorously proved to satisfy the guarantee. The key privacy guarantee that has emerged is *differential privacy*. Roughly speaking, this ensures that (almost, and quantifiably) no risk is incurred by joining a statistical database.

In this survey, we recall the definition of differential privacy and two basic techniques for achieving it. We then show some interesting applications of these techniques, presenting algorithms for three specific tasks and three general results on differentially private learning.

1 Introduction

Privacy-preserving data analysis is also known as statistical disclosure control, inference control, privacy-preserving datamining, and private data analysis. Our principal motivating scenario is a *statistical database*. A statistic is a quantity computed from a sample. Suppose a trusted and trustworthy curator gathers sensitive information from a large number of respondents (the sample), with the goal of learning (and releasing to the public) statistical facts about the underlying population. The problem is to release statistical information without compromising the privacy of the individual respondents. There are two settings: in the *noninteractive* setting the curator computes and publishes some statistics, and the data are not used further. Privacy concerns may affect the precise answers released by the curator, or even the set of statistics released. Note that since the data will never be used again the curator can destroy the data (and himself) once the statistics have been published.

In the *interactive* setting the curator sits between the users and the database. Queries posed by the users, and/or the responses to these queries, may be modified by the curator in order to protect the privacy of the respondents. The data cannot be destroyed, and the curator must remain present throughout the lifetime of the database. Of course, any interactive solution yields a non-interactive solution, provided the queries are known in advance: the curator can simulate an interaction in which these known queries are posed, and publish the resulting transcript.

There is a rich literature on this problem, principally from the satistics community (see, *e.g.*, [10, 14, 27, 28, 29, 38, 40, 26, 39] and the literature on controlled

M. Agrawal et al. (Eds.): TAMC 2008, LNCS 4978, pp. 1–19, 2008.

release of tabular data, contingency tables, and cell suppression), and from such diverse branches of computer science as algorithms, database theory, and cryptography, for example as in [3, 4, 21, 22, 23, 33, 34, 41, 48], [1, 24, 25, 31], and [6, 9, 11, 12, 13, 18, 7, 19]; see also the survey [2] for a summary of the field prior to 1989.

This survey is about *differential privacy*. Roughly speaking, differential privacy ensures that the removal or addition of a single database item does not (substantially) affect the outcome of any analysis. It follows that no risk is incurred by joining the database, providing a mathematically rigorous means of coping with the fact that distributional information may be disclosive.

We will first describe three differentially private algorithms for specific, unrelated, data analysis tasks. We then present three general results about computational learning when privacy of individual data items is to be protected. This is not usually a concern in the learning theory literature, and signals the emergence of a new line of research.

2 Differential Privacy

In the sequel, the randomized function $\mathcal{K}$ is the algorithm applied by the curator when releasing information. So the input is the data set, and the output is the released information, or *transcript*. We do not need to distinguish between the interactive and non-interactive settings.

Think of a database as a set of rows. We say databases D_1 and D_2 *differ in at most one element* if one is a proper subset of the other and the larger database contains just one additional row.

Definition 1. *A randomized function $\mathcal{K}$ gives* ϵ-differential privacy *if for all data sets D_1 and D_2 differing on at most one element, and all $S \subseteq Range(\mathcal{K})$,*

$$\Pr[\mathcal{K}(D_1) \in S] \leq \exp(\epsilon) \times \Pr[\mathcal{K}(D_2) \in S] \tag{1}$$

The probability is taken is over the coin tosses of $\mathcal{K}$.

A mechanism $\mathcal{K}$ satisfying this definition addresses concerns that any participant might have about the leakage of her personal information: 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 that individual's data in the database will not significantly affect her chance of receiving coverage.

Differential privacy is therefore an *ad omnia* guarantee. It is also a very strong guarantee, since it is a statistical property about the behavior of the mechanism and therefore is independent of the computational power and auxiliary information available to the adversary/user.

Differential privacy is not an absolute guarantee of privacy. In fact, Dwork and Naor have shown that any statistical database with any non-trivial utility