The Algorithmic Foundations of Differential Privacy

Cynthia Dwork
Microsoft Research, USA
dwork@microsoft.com

Aaron Roth
University of Pennsylvania, USA
aaroth@gmail.com
Editorial Scope

Topics

Foundations and Trends® in Theoretical Computer Science publishes surveys and tutorials on the foundations of computer science. The scope of the series is broad. Articles in this series focus on mathematical approaches to topics revolving around the theme of efficiency in computing. The list of topics below is meant to illustrate some of the coverage, and is not intended to be an exhaustive list.

- Algorithmic game theory
- Computational algebra
- Computational aspects of combinatorics and graph theory
- Computational aspects of communication
- Computational biology
- Computational complexity
- Computational geometry
- Computational learning
- Computational Models and Complexity
- Computational Number Theory
- Cryptography and information security
- Data structures
- Database theory
- Design and analysis of algorithms
- Distributed computing
- Information retrieval
- Operations research
- Parallel algorithms
- Quantum computation
- Randomness in computation

Information for Librarians

Foundations and Trends® in Theoretical Computer Science, 2013, Volume 9, 4 issues. ISSN paper version 1551-305X. ISSN online version 1551-3068. Also available as a combined paper and online subscription.
The Algorithmic Foundations of Differential Privacy

Cynthia Dwork
Microsoft Research, USA
dwork@microsoft.com

Aaron Roth
University of Pennsylvania, USA
aaroth@gmail.com
### Contents

**Preface**  
3

1 **The Promise of Differential Privacy**  
   1.1 Privacy-preserving data analysis  ................. 6  
   1.2 Bibliographic notes  ............................. 10

2 **Basic Terms**  
   2.1 The model of computation  .......................... 11  
   2.2 Towards defining private data analysis .......... 12  
   2.3 Formalizing differential privacy  ................. 15  
   2.4 Bibliographic notes  ............................. 26

3 **Basic Techniques and Composition Theorems**  
   3.1 Useful probabilistic tools  ....................... 28  
   3.2 Randomized response  ............................. 29  
   3.3 The Laplace mechanism  ........................... 30  
   3.4 The exponential mechanism  ....................... 37  
   3.5 Composition theorems  ............................ 41  
   3.6 The sparse vector technique  ..................... 55  
   3.7 Bibliographic notes  ............................. 64
Abstract

The problem of privacy-preserving data analysis has a long history spanning multiple disciplines. As electronic data about individuals becomes increasingly detailed, and as technology enables ever more powerful collection and curation of these data, the need increases for a robust, meaningful, and mathematically rigorous definition of privacy, together with a computationally rich class of algorithms that satisfy this definition. Differential Privacy is such a definition.

After motivating and discussing the meaning of differential privacy, the preponderance of this monograph is devoted to fundamental techniques for achieving differential privacy, and application of these techniques in creative combinations, using the query-release problem as an ongoing example. A key point is that, by rethinking the computational goal, one can often obtain far better results than would be achieved by methodically replacing each step of a non-private computation with a differentially private implementation. Despite some astonishingly powerful computational results, there are still fundamental limitations — not just on what can be achieved with differential privacy but on what can be achieved with any method that protects against a complete breakdown in privacy. Virtually all the algorithms discussed herein maintain differential privacy against adversaries of arbitrary computational power. Certain algorithms are computationally intensive, others are efficient. Computational complexity for the adversary and the algorithm are both discussed.

We then turn from fundamentals to applications other than query-release, discussing differentially private methods for mechanism design and machine learning. The vast majority of the literature on differentially private algorithms considers a single, static, database that is subject to many analyses. Differential privacy in other models, including distributed databases and computations on data streams is discussed.
Finally, we note that this work is meant as a thorough introduction to the problems and techniques of differential privacy, but is not intended to be an exhaustive survey — there is by now a vast amount of work in differential privacy, and we can cover only a small portion of it.

DOI: 10.1561/0400000042.
The problem of privacy-preserving data analysis has a long history spanning multiple disciplines. As electronic data about individuals becomes increasingly detailed, and as technology enables ever more powerful collection and curation of these data, the need increases for a robust, meaningful, and mathematically rigorous definition of privacy, together with a computationally rich class of algorithms that satisfy this definition. *Differential Privacy* is such a definition.

After motivating and discussing the meaning of differential privacy, the preponderance of the book is devoted to fundamental techniques for achieving differential privacy, and application of these techniques in creative combinations (Sections 3–7), using the *query-release* problem as an ongoing example. A key point is that, by rethinking the computational goal, one can often obtain far better results than would be achieved by methodically replacing each step of a non-private computation with a differentially private implementation.

Despite some astonishingly powerful computational results, there are still fundamental limitations — not just on what can be achieved with differential privacy but on what can be achieved with *any* method that protects against a complete breakdown in privacy (Section 8).

Virtually all the algorithms discussed in this book maintain differential privacy against adversaries of arbitrary computational power. Certain algorithms are computationally intensive, others are
efficient. Computational complexity for the adversary and the algorithm are both discussed in Section 9.

In Sections 10 and 11 we turn from fundamentals to applications other than query-release, discussing differentially private methods for mechanism design and machine learning. The vast majority of the literature on differentially private algorithms considers a single, static, database that is subject to many analyses. Differential privacy in other models, including distributed databases and computations on data streams is discussed in Section 12.

Finally, we note that this book is meant as a thorough introduction to the problems and techniques of differential privacy, but is not intended to be an exhaustive survey — there is by now a vast amount of work in differential privacy, and we can cover only a small portion of it.
“Differential privacy” describes a promise, made by a data holder, or curator, to a data subject: “You will not be affected, adversely or otherwise, by allowing your data to be used in any study or analysis, no matter what other studies, data sets, or information sources, are available.” At their best, differentially private database mechanisms can make confidential data widely available for accurate data analysis, without resorting to data clean rooms, data usage agreements, data protection plans, or restricted views. Nonetheless, data utility will eventually be consumed: the Fundamental Law of Information Recovery states that overly accurate answers to too many questions will destroy privacy in a spectacular way. The goal of algorithmic research on differential privacy is to postpone this inevitability as long as possible.

Differential privacy addresses the paradox of learning nothing about an individual while learning useful information about a population. A medical database may teach us that smoking causes cancer, affecting an insurance company’s view of a smoker’s long-term medical costs. Has the smoker been harmed by the analysis? Perhaps — his insurance

---

1This result, proved in Section 8.1, applies to all techniques for privacy-preserving data analysis, and not just to differential privacy.
premiums may rise, if the insurer knows he smokes. He may also be
helped — learning of his health risks, he enters a smoking cessation
program. Has the smoker’s privacy been compromised? It is certainly
the case that more is known about him after the study than was known
before, but was his information “leaked”? Differential privacy will take
the view that it was not, with the rationale that the impact on the
smoker is the same independent of whether or not he was in the study.
It is the conclusions reached in the study that affect the smoker, not
his presence or absence in the data set.

Differential privacy ensures that the same conclusions, for example,
smoking causes cancer, will be reached, independent of whether any
individual opts into or opts out of the data set. Specifically, it ensures
that any sequence of outputs (responses to queries) is “essentially”
equally likely to occur, independent of the presence or absence of any
individual. Here, the probabilities are taken over random choices made
by the privacy mechanism (something controlled by the data curator),
and the term “essentially” is captured by a parameter, $\varepsilon$. A smaller $\varepsilon$
will yield better privacy (and less accurate responses).

Differential privacy is a definition, not an algorithm. For a given
computational task $T$ and a given value of $\varepsilon$ there will be many differ-
entially private algorithms for achieving $T$ in an $\varepsilon$-differentially private
manner. Some will have better accuracy than others. When $\varepsilon$ is small,
finding a highly accurate $\varepsilon$-differentially private algorithm for $T$ can be
difficult, much as finding a numerically stable algorithm for a specific
computational task can require effort.

1.1 Privacy-preserving data analysis

Differential privacy is a definition of privacy tailored to the problem
of privacy-preserving data analysis. We briefly address some concerns
with other approaches to this problem.

Data Cannot be Fully Anonymized and Remain Useful. Generally
speaking, the richer the data, the more interesting and useful it is.
This has led to notions of “anonymization” and “removal of person-
ally identifiable information,” where the hope is that portions of the
data records can be suppressed and the remainder published and used for analysis. However, the richness of the data enables “naming” an individual by a sometimes surprising collection of fields, or attributes, such as the combination of zip code, date of birth, and sex, or even the names of three movies and the approximate dates on which an individual watched these movies. This “naming” capability can be used in a \textit{linkage attack} to match “anonymized” records with non-anonymized records in a different dataset. Thus, the medical records of the governor of Massachusetts were identified by matching anonymized medical encounter data with (publicly available) voter registration records, and Netflix subscribers whose viewing histories were contained in a collection of anonymized movie records published by Netflix as training data for a competition on recommendation were identified by linkage with the Internet Movie Database (IMDb).

Differential privacy neutralizes linkage attacks: since being differentially private is a property of the data access mechanism, and is unrelated to the presence or absence of auxiliary information available to the adversary, access to the IMDb would no more permit a linkage attack to someone whose history is in the Netflix training set than to someone not in the training set.

\textit{Re-Identification of “Anonymized” Records is Not the Only Risk.} Re-identification of “anonymized” data records is clearly undesirable, not only because of the re-identification \textit{per se}, which certainly reveals membership in the data set, but also because the record may contain compromising information that, were it tied to an individual, could cause harm. A collection of medical encounter records from a specific urgent care center on a given date may list only a small number of distinct complaints or diagnoses. The additional information that a neighbor visited the facility on the date in question gives a fairly narrow range of possible diagnoses for the neighbor’s condition. The fact that it may not be possible to match a specific record to the neighbor provides minimal privacy protection to the neighbor.

\textit{Queries Over Large Sets are Not Protective.} Questions about specific individuals cannot be safely answered with accuracy, and indeed one
might wish to reject them out of hand (were it computationally feasible to recognize them). Forcing queries to be over large sets is not a panacea, as shown by the following differencing attack. Suppose it is known that Mr. X is in a certain medical database. Taken together, the answers to the two large queries “How many people in the database have the sickle cell trait?” and “How many people, not named X, in the database have the sickle cell trait?” yield the sickle cell status of Mr. X.

Query Auditing Is Problematic. One might be tempted to audit the sequence of queries and responses, with the goal of interdicting any response if, in light of the history, answering the current query would compromise privacy. For example, the auditor may be on the lookout for pairs of queries that would constitute a differencing attack. There are two difficulties with this approach. First, it is possible that refusing to answer a query is itself disclosive. Second, query auditing can be computationally infeasible; indeed if the query language is sufficiently rich there may not even exist an algorithmic procedure for deciding if a pair of queries constitutes a differencing attack.

Summary Statistics are Not “Safe.” In some sense, the failure of summary statistics as a privacy solution concept is immediate from the differencing attack just described. Other problems with summary statistics include a variety of reconstruction attacks against a database in which each individual has a “secret bit” to be protected. The utility goal may be to permit, for example, questions of the form “How many people satisfying property $P$ have secret bit value 1?” The goal of the adversary, on the other hand, is to significantly increase his chance of guessing the secret bits of individuals. The reconstruction attacks described in Section 8.1 show the difficulty of protecting against even a linear number of queries of this type: unless sufficient inaccuracy is introduced almost all the secret bits can be reconstructed.

A striking illustration of the risks of releasing summary statistics is in an application of a statistical technique, originally intended for confirming or refuting the presence of an individual’s DNA in a forensic mix, to ruling an individual in or out of a genome-wide association study. According to a Web site of the Human Genome Project, “Single nucleotide polymorphisms, or SNPs (pronounced “snips”), are DNA
sequence variations that occur when a single nucleotide (A,T,C, or G) in the genome sequence is altered. For example a SNP might change the DNA sequence AAGGCTAA to ATGGCTAA. In this case we say there are two alleles: A and T. For such a SNP we can ask, given a particular reference population, what are the frequencies of each of the two possible alleles? Given the allele frequencies for SNPs in the reference population, we can examine how these frequencies may differ for a subpopulation that has a particular disease (the “case” group), looking for alleles that are associated with the disease. For this reason, genome-wide association studies may contain the allele frequencies of the case group for large numbers of SNPs. By definition, these allele frequencies are only aggregated statistics, and the (erroneous) assumption has been that, by virtue of this aggregation, they preserve privacy. However, given the genomic data of an individual, it is theoretically possible to determine if the individual is in the case group (and, therefore, has the disease). In response, the National Institutes of Health and Wellcome Trust terminated public access to aggregate frequency data from the studies they fund.

This is a challenging problem even for differential privacy, due to the large number — hundreds of thousands or even one million — of measurements involved and the relatively small number of individuals in any case group.

“Ordinary” Facts are Not “OK.” Revealing “ordinary” facts, such as purchasing bread, may be problematic if a data subject is followed over time. For example, consider Mr. T, who regularly buys bread, year after year, until suddenly switching to rarely buying bread. An analyst might conclude Mr. T most likely has been diagnosed with Type 2 diabetes. The analyst might be correct, or might be incorrect; either way Mr. T is harmed.

“Just a Few.” In some cases a particular technique may in fact provide privacy protection for “typical” members of a data set, or more generally, “most” members. In such cases one often hears the argument that the technique is adequate, as it compromises the privacy of “just a few” participants. Setting aside the concern that outliers may be precisely those people for whom privacy is most important, the “just a few”
philosophy is not intrinsically without merit: there is a social judgment, a weighing of costs and benefits, to be made. A well-articulated definition of privacy consistent with the “just a few” philosophy has yet to be developed; however, for a single data set, “just a few” privacy can be achieved by randomly selecting a subset of rows and releasing them in their entirety (Lemma 4.3, Section 3). Sampling bounds describing the quality of statistical analysis that can be carried out on random subsamples govern the number of rows to be released. Differential privacy provides an alternative when the “just a few” philosophy is rejected.

1.2 Bibliographic notes

Sweeney [81] linked voter registration records to “anonymized” medical encounter data; Narayanan and Shmatikov carried out a linkage attack against anonymized ranking data published by Netflix [65]. The work on presence in a forensic mix is due to Homer et al. [46]. The first reconstruction attacks were due to Dinur and Nissim [18].
References


References


References


References


