Foundations and Trends® in
Theoretical Computer Science
Vol. 9, Nos. 3–4 (2014) 211–407

DOI: 10.1561/0400000042

now
the essence of knowledge

The Algorithmic Foundations of Differential Privacy

Cynthia Dwork
Microsoft Research, USA
dwork@microsoft.com

Aaron Roth
University of Pennsylvania, USA
aaroth@cis.upenn.edu

Contents

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.

C. Dwork and A. Roth. *The Algorithmic Foundations of Differential Privacy.* Foundations and Trends® in Theoretical Computer Science, vol. 9, nos. 3–4, pp. 211–407, 2014.
DOI: 10.1561/0400000042.

Preface

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.