Application to Present Research at PrivacyCon 2018

November 17, 2017

Contact information

Alexandra Wood
Fellow, Berkman Klein Center for Internet & Society at Harvard University

Kobbi Nissim
Professor, Department of Computer Science, Georgetown University

Presentation title

Understanding the Differential Privacy Guarantee

Abstract

Differential privacy is a formal mathematical standard for quantifying and managing privacy risk. Analyses satisfying differential privacy provide provable privacy protection against a wide range of potential attacks, including types of attacks currently unforeseen. Differential privacy is primarily studied in the context of the collection, analysis, and release of aggregate statistics. These range from simple statistical estimations, such as averages, to machine learning. Tools for differentially private analysis are now in early stages of use across a variety of industry and government settings, with practical implementations by companies such as Google, Apple, and Uber and federal agencies such as the US Census Bureau. Interest in the concept is continuing to grow within commercial settings, as it holds promise as a potential approach to satisfying legal requirements for privacy protection when handling personal information. Applications of differential privacy can help enable the collection, analysis, and sharing of personal data while protecting the privacy of individuals in accordance with existing legal requirements for de-identification or disclosure limitation. In particular, applications by Google and Apple have demonstrated that the use of differential privacy makes it possible for a companies to learn about the behavior of their user bases without learning (or even collecting) information specific to any individual user.

This presentation is based on research to develop intuitive explanations of differential privacy for non-technical audiences. The speakers will introduce differential privacy using simplified and informal, but mathematically accurate, illustrations. Drawing from simplified real-world scenarios and privacy risk concepts from law, policy, statistics, and social science, these illustrations are designed to help consumers, businesses, privacy practitioners, and policymakers conceptualize the differential privacy guarantee in terms that are relevant to the practical decisions they face. For instance, this research aims to

provide a rigorous, quantitative foundation for future decisions made by businesses when assessing privacy risk and selecting appropriate safeguards prior to collecting, analyzing, or sharing personal data. These intuitive explanations can also be used by businesses to provide consumers with a meaningful understanding of the privacy protection they will be afforded when differential privacy is used.

By integrating computer science, legal, and social science concepts and practices around privacy protection, this work contextualizes findings from recent scientific research on privacy and explains their real-world implications and relationship to other concepts in non-technical terms. It is intended to serve as a resource for businesses, government agencies, and research institutions that collect and analyze privacy-sensitive personal data, and are increasingly tasked with making decisions such as whether and how to share their data and statistical summaries of their data internally, with the research community, and with the public at large. It can also be used to the develop consumer-facing explanations of privacy risk and the protection provided by differential privacy. In addition, the findings of this research can inform legal scholars and policymakers as they consider how existing and future legal frameworks and instruments will apply to tools based on formal privacy models such as differential privacy. This work also provides intuitive illustrations of related concepts from the computer science literature, such as composition (the accumulation of risk across multiple analyses), privacy loss parameters, and privacy budgets. This work demonstrates that privacy loss can be formally quantified and bounded, modern notions of privacy are not limited to a binary view of information as "identifiable" or "de-identified," privacy risk is broader than traditional understandings of re-identification risk, privacy is a property of the computation rather than the outcome of an analysis, and formal definitions of privacy can be shown to be sufficient to satisfy legal and ethical requirements for privacy protection. These explanations are presented to guide future policy discussions regarding revisions to regulatory standards for privacy protection in light of the evolving technical understanding of privacy risk.

Publication details

A preliminary working paper summarizing this work is available at https://privacytools.seas.harvard.edu/publications/differential-privacy-primer-non-technical-audience-preliminary-version