CS208 Spring 2019 Annotated Bibliography

James Honaker and Salil Vadhan

April 5, 2019

• Background Material
  – Discrete math and proofs: Solow [2013], Rosen [2012]
  – Algorithms and complexity: Cormen et al. [2009], Mitzenmacher and Upfal [2005]

• General References
  – Many videos of talks on recent developments in the theory and applications of differential privacy: https://simons.berkeley.edu/programs/privacy2019
  – Tutorial on “DP in the Wild”: Machanavajjhala et al. [2017] (see also slides online)

• Reidentification Attacks
  – (required) Forbes article on Sweeney’s reidentification of Personal Genome Project participants: Tanner [2013]
  – (required) Narayanan-Shmatikov opinion piece on the concept of PII: Narayanan and Shmatikov [2010]
  – Sweeney’s original re-identification: Sweeney [1997]
  – Paper on the Personal Genome Project reidentification: Sweeney et al. [2013]
  – Paper introducing $k$-anonymity: Sweeney [2002]
  – Composition attack on $k$-anonymity: Ganta et al. [2008]
  – Biases introduced by de-identification of EdX data: Daries et al. [2014]
  – Netflix reidentification: Narayanan and Shmatikov [2008]
  – Cancellation of 2nd Netflix Challenge after Lawsuit: Singel [2010]

• Reconstruction Attacks
  – (required) Linear programming attack on Diffix: Cohen and Nissim [2018]
  – (required) SAT Solver attack on Census data: Garfinkel et al. [2018a]
  – Survey paper on attacks on aggregate statistics: Dwork et al. [2017, §1,2]
  – Paper introducing reconstruction attacks: Dinur and Nissim [2003]
  – Differencing attack on Israeli Census: Ziv [2013]

• Membership Attacks
– (required) P3G Consortium responses to membership attacks on genomic data: Consortium et al. [2009]
– (required) Privacy attacks on microtargeted ads: Korolova [2011, §1,4,6,8]
– Survey paper on attacks on aggregate statistics: Dwork et al. [2017, §3]
– Membership attack on means in genomic data: Homer et al. [2008]
– Membership attack on noisy means: Dwork et al. [2015]
– Membership attack on ML as a Service: Shokri et al. [2017]
– Attribute inference attacks on ML: Fredrikson et al. [2014]
– Blog post in response to inference attacks on ML: McSherry [2016]

• Foundations of Differential Privacy
  – (extracts required) Primer for non-technical audiences: Wood et al. [2018]
  – (extracts required) The standard textbook: Dwork and Roth [2013]
  – Survey on complexity-theoretic aspects of differential privacy: Vadhan [2017]
  – The papers leading up to and culminating in the definition of differential privacy and the first mechanisms (Laplace, histograms): Dinur and Nissim [2003], Dwork and Nissim [2004], Blum et al. [2005], Dwork et al. [2016].
  – Attacks on floating-point implementations of differential privacy and remedies: Mironov [2012], Balcer and Vadhan [2018]
  – The geometric mechanism: Ghosh et al. [2012]
  – A Bayesian interpretation of approximate DP: Kasiviswanathan and Smith [2014]
  – A survey on differential privacy for social networks: Raskhodnikova and Smith [2014]
  – The advanced and “optimal” composition theorems for approximate DP: Dwork et al. [2010], Kairouz et al. [2017], Murtagh and Vadhan [2018]
  – Other variants of DP that compose more cleanly than approximate DP: Dwork and Roth [2013], Bun and Steinke [2016], Mironov [2017], Bun et al. [2018]
  – Differential privacy and the Statistical Query model for machine learning: Blum et al. [2005], Kasiviswanathan et al. [2011]
  – The paper that introduced the exponential mechanism: McSherry and Talwar [2007]
  – Another mechanism for the median (via smooth sensitivity): Kasiviswanathan et al. [2013]
  – Survey of approaches to add noise closer to the local sensitivity: [Vadhan, 2017, Ch. 3]

• Implementing Differential Privacy: One-Shot Releases
  – The stability-based histogram and other histogram algorithms for large data universes: Korolova et al. [2009], Balcer and Vadhan [2018]
  – Early applications of DP synthetic data to commuting patterns and mobility data: Machanavajjhala et al. [2008], Mir et al. [2013]
  – (required or slides covered in class) Census Bureau’s adoption of DP: Garfinkel et al. [2018b], Garfinkel [2018]
  – Other papers and talks on the Census Bureau’s adoption of DP: Abowd [2018], Kifer [2019], Dajani et al. [2017]
  – Private Multiplicative Weights: Hardt and Rothblum [2010]. (See also sections of Dwork and Roth [2013], Vadhan [2017].)
– (excerpts required) DualQuery: Gaboardi et al. [2017]
– Another algorithm for synthetic data generation (MWEM): Hardt et al. [2012]
– Worst-case hardness of differentially private synthetic data generation: Dwork et al. [2009], Ullman and Vadhan [2011] (See also sections of Vadhan [2017].)
– (excerpts required) The Opportunity Atlas and the underlying privacy mechanism: Chetty et al. [2018], Chetty and Friedman [2019]
– The Matrix Mechanism: Li et al. [2015]
– The Hierarchical Mechanism for Range Queries: Hay et al. [2010]
– How to compare DP algorithms: Hay et al. [2016]
• Implementing Differential Privacy: Programming Frameworks and Query Systems
  – PinQ and its formal verification: McSherry [2010], Ebadi and Sands [2017]
  – εktelo: Zhang et al. [2018]
  – Differentially Private SQL: Johnson et al. [2018], Kotsogiannis et al. [2019]
  – Differentially Private MapReduce: Roy et al. [2010]
  – Side-channel attacks on implementations of DP: Haeberlen et al. [2011], Mironov [2012]
  – Survey on formal verification of DP and recent developments: Barthe et al. [2016], Zhang and Kifer [2017], Alibghouthi and Hsu [2017]
  – DP Query Systems that Budget via Accuracy: Mohan et al. [2012], Gaboardi et al. [2016]
• The Local and Multiparty Models of Differential Privacy, and Combining Cryptography and DP
  – Tutorial: Cormode et al. [2018], see also videos online
  – History of randomized response in the survey literature, and some current applications: Gingerich [2015, 2010], Blair et al. [2015]
  – Equivalence of local DP and the SQ model: Kasiviswanathan et al. [2011]
  – More on models for interactive and multiparty DP: Vadhan [2017, Chs. 9-10]
  – Composition when privacy parameters are chosen adaptively: Rogers et al. [2016]
  – Local DP with anonymous/shuffled data subjects: Bittau et al. [2017], Cheu et al. [2019], Erlingsson et al. [2019], Balle et al. [2019]
  – Recent papers on combining DP and secure multiparty computation: He et al. [2017], Archer et al. [2018]
  – Google’s RAPPOR: Erlingsson et al. [2014]
  – Tang et al. [2017]
  – Local DP for Evolving Data: Joseph et al. [2018]
• Machine Learning and Statistical Inference with DP
– Bibliography for Adam Smith’s Fall 2018 course CS 591 at BU: https://docs.google.com/document/d/1jsZLEd3ZM-ZwDNAjNRfUbgFsRUSQDHvy4Vkg7zJ8/edit#heading=h.6a7pxu1gz13i

• Software
– DualQuery: https://github.com/ejgallego/dualquery
– MWEM: https://github.com/mrtzh/PrivateMultiplicativeWeights.jl
– εktelo: https://ektelo.github.io/
– TensorFlow Privacy: https://github.com/tensorflow/privacy
– FLEX (SQL, deployed by Uber): http://www.uvm.edu/~jnear/elastic/
– PSI: http://psiprivacy.org/about/
– LightDP: https://github.com/RyanWangGit/lightdp
– RAPPOR: https://github.com/google/rappor
– Prochlo: https://github.com/google/prochlo
– DPComp (for comparing DP algorithms): https://www.dpcomp.org/

References


