Differential Privacy for Databases

Other titles in Foundations and Trends® in Databases

Database Systems on GPUs
Johns Paul, Shengliang Lu and Bingsheng He
ISBN: 978-1-68083-848-0

Machine Knowledge: Creation and Curation of Comprehensive Knowledge Bases
Gerhard Weikum, Xin Luna Dong, Simon Razniewski and Fabian Suchanek
ISBN: 978-1-68083-836-7

Cloud Data Services: Workloads, Architectures and Multi-Tenancy
Vivek Narasayya and Surajit Chaudhuri
ISBN: 978-1-68083-774-2

Data Provenance
Boris Glavic
ISBN: 978-1-68083-828-2

FPGA-Accelerated Analytics: From Single Nodes to Clusters
Zsolt István, Kaan Kara and David Sidler
ISBN: 978-1-68083-734-6

Distributed Learning Systems with First-Order Methods
Ji Liu and Ce Zhango
ISBN: 978-1-68083-700-1

Differential Privacy for Databases

Joseph P. Near
University of Vermont
jnear@uvm.edu

Xi He
University of Waterloo
xi.he@uwaterloo.ca

the essence of knowledge
Boston — Delft

Foundations and Trends® in Databases

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
United States
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

Outside North America:
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is

J. Near and X. He. *Differential Privacy for Databases*. Foundations and Trends® in Databases, vol. 11, no. 2, pp. 109–225, 2021.

ISBN: 978-1-68083-850-0

Editorial Scope

Topics

Foundations and Trends® in Databases publishes survey and tutorial articles in the following topics:

- Data Models and Query Languages
- Query Processing and Optimization
- Storage, Access Methods, and Indexing
- Transaction Management, Concurrency Control and Recovery
- Deductive Databases
- Parallel and Distributed Database Systems
- Database Design and Tuning
- Metadata Management
- Object Management
- Trigger Processing and Active Databases
- Data Mining and OLAP
- Approximate and Interactive Query Processing
- Data Warehousing
- Adaptive Query Processing
- Data Stream Management
- Search and Query Integration
- XML and Semi-Structured Data
- Web Services and Middleware
- Data Integration and Exchange
- Private and Secure Data Management
- Peer-to-Peer, Sensornet and Mobile Data Management
- Scientific and Spatial Data Management
- Data Brokering and Publish/Subscribe
- Data Cleaning and Information Extraction
- Probabilistic Data Management

Information for Librarians

Foundations and Trends® in Databases, 2021, Volume 11, 4 issues. ISSN paper version 1931-7883. ISSN online version 1931-7891. Also available as a combined paper and online subscription.

Contents

Differential Privacy for Databases

Joseph P. Near[1] and Xi He[2]

[1] *University of Vermont; jnear@uvm.edu*
[2] *University of Waterloo; xi.he@uwaterloo.ca*

ABSTRACT

Differential privacy is a promising approach to *formalizing privacy*—that is, for writing down *what privacy means* as a mathematical equation. This book is provides overview of differential privacy techniques for answering database-style queries. Within this area, we describe useful algorithms and their applications, and systems and tools that implement them.

Joseph P. Near and Xi He (2021), "Differential Privacy for Databases", Foundations and Trends® in Databases: Vol. 11, No. 2, pp 109–225. DOI: 10.1561/1900000066.

1

Introduction

Differential privacy is a promising approach to *formalizing privacy*—that is, for writing down *what privacy means* as a mathematical equation. The definition of differential privacy acts as a bridge between societal notions of privacy and the mathematical properties of privacy-preserving algorithms—we can prove that a specific algorithm satisfies differential privacy, and then argue separately that the definition is a "good" approximation of society's *informal* notions of privacy. Differential privacy has been successful because it seems to serve particularly well in this role—it is the best mathematical model of privacy that we know of.

This book is intended to serve as an overview of the state-of-the-art in techniques for differential privacy. We focus in particular on techniques for answering database-style queries, on useful algorithms and their applications, and on systems and tools that implement them. While we do describe the formal properties of the techniques we cover, our focus is not on theoretical results.

What is privacy? In this book, we use the term *privacy* to refer to situations in which an adversary is **not able to learn too much about any one individual**. When the adversary learns too much about an individual, we say that privacy has been lost. One trivial solution for privacy is to prevent the adversary from learning *anything*—but this approach makes it pointless to collect and analyze data in the first place.

The techniques we explore in this book are ones that allow the adversary to learn *properties of the population* while hiding information specific to individuals. Such techniques allow us to learn useful information from sensitive data, while at the same time protecting the privacy of the individuals who contributed it.

What is privacy *not*? Privacy properties are often conflated with security properties. Though they are related, they are distinct in important ways. Common security properties include *confidentiality* (that an adversary learns *nothing* about the secret data) and *integrity* (that an adversary is not capable of corrupting the system's output).

Privacy-preserving algorithms do not necessarily satisfy either of these properties. Differentially private algorithms *intentionally* reveal some information to the adversary; the goal of differential privacy is to control *what can be learned* from that information.

Similarly, techniques for enforcing security properties do not necessarily ensure privacy. In particular, most techniques for security control *who* can view the data—not *what information* they can learn from it. Encrypting a dataset, for example, provides "all-or-nothing" access to its information—those without the key learn nothing, while those with the key learn everything, including information specific to individuals. Encryption, by itself, is not capable of making the distinction described above between properties of the population and properties of individuals.

However, security techniques can *complement* privacy techniques in important ways. In particular, such techniques allow us to target alternative *threat models* for differentially private algorithms. For example, many systems for differential privacy collect raw sensitive data on a central server, and assume the server will not be compromised. If the server is hacked, however, then the guarantee of differential privacy may be violated. Encrypting this data may help ensure that *only* differentially private results are ever made public—even if the server holding the data is compromised. Complementing differential privacy thus allows us to adjust the threat model to protect against a stronger adversary than before. We discuss combining differential privacy with security techniques in Chapter 9.

Why *differential* privacy? Differential privacy is the latest in a series of approaches for building privacy-preserving algorithms. The most common technique for releasing data while preserving privacy is *de-identification* (sometimes called anonymization), which involves removing *identifying information* from the data. De-identification appeals to our intuitions about privacy, but numerous results suggest that *re-identification* attacks on de-identified data are often possible (Sweeney, 2000; Dinur and Nissim, 2003).

More rigorous techniques, like k-Anonymity (Sweeney, 2002) and ℓ-Diversity (Machanavajjhala *et al.*, 2007), were developed to address this shortcoming by quantifying the "uniqueness" of an individual within a dataset. However, even these techniques are not *compositional*—releasing a single k-Anonymized dataset might provide strong privacy protection, but releasing *two* such datasets may enable an adversary to re-identify individuals in the data.

Differential privacy is attractive because in addition to closely approximating our informal notions of privacy, it is *compositional*. Compositionality means that if two data releases *individually* provide certain levels of differential privacy, then we can bound the *cumulative* privacy loss of both releases. Differential privacy is the first rigorous approach to privacy with this important property.

What does differential privacy protect? The goal of differential privacy is to make the following promise: *if you participate in a differentially private analysis of data, you will not suffer any additional harm as a result*. Roughly speaking, the mathematical definition of differential privacy achieves this goal by requiring that the outcome of any differentially private analysis is *the same whether or not you participate* (this notion is formalized in Chapter 2).

Importantly, this guarantee does not necessarily prevent an adversary from learning details about an individual—particularly when those details could have been learned *without the individual's participation in the analysis*. For example, if a differentially private study concludes that all people over age 50 enjoy playing tennis, then an adversary may infer that a specific 52-year-old enjoys the sport. Differential privacy does not prevent this situation, because it is possible *whether or not* the specific 52-year-old participates in the study.

What are the limits of differential privacy? A clear tension tension exists between revealing information about a dataset and protecting the privacy of its individuals—revealing too many properties of the data with too much accuracy must *necessarily* violate privacy. This idea—now often called the *database reconstruction theorem*—imposes upper bounds on what it is possible to learn before privacy is violated (Dinur and Nissim, 2003). Navigating this tension is a key part of designing differentially private algorithms, which typically have the goal of releasing the most accurate possible statistics while preserving privacy.

Why use differential privacy in database systems? Today's information systems collect and process vast amounts of data, and the majority of it flows into databases (relational or otherwise). These database systems are specifically designed to collect, store, and query data, and have been optimized for that task. If we would like to enable an analysis of sensitive data with differential privacy, it is logical to develop techniques that work for database systems, because *that's where the private data is.*

However, integrating differentially private techniques with database systems presents significant challenges—many of which are discussed later in this book. In particular, a primary goal of most database systems is to abstract away execution details, so that analysts may focus on the semantics of the queries they write instead of worrying about how they will be executed. But satisfying differential privacy requires careful control over the details of how a query is executed, which sometimes breaks this abstraction.

The techniques covered in this book represent significant progress towards building differentially private database systems. They differ in terms of their capabilities and the interfaces they present to the analyst, and none matches perfectly with the traditional abstractions used in relational databases. Indeed, significant challenges remain in achieving that goal—we discuss these in Chapter 10—and we may never get all the way there. On the other hand, the approaches described in this book have already resulted in useful, deployable systems, and we hope they will pave the way towards increasing adoption of differential privacy in practice.

Summary & Additional resources. This book focuses on techniques, algorithms, and systems for answering database-style queries with differential privacy. This area is just one part of the larger field of research in differential privacy. For an introduction to the theoretical foundations of differential privacy, we refer the reader to the excellent reference by Dwork & Roth (Dwork, Roth, *et al.*, 2014). We provide additional references to more detailed descriptions of smaller sub-areas of differential privacy throughout this book.

The rest of the book is organized into three parts. The first part defines our setting and provides background: Chapter 2 describes the basics of differential privacy, and Chapter 3 describes databases and queries. Section 3.6 summarizes the specific techniques covered in the book. The second part—Chapters 4, 5, 6, and 7—describes specific techniques, categorized by application area. The third part describes progress and challenges in building differentially-private systems: Chapter 8 describes frameworks for building such systems, Chapter 9 describes the use of security techniques to support privacy, and Chapter 10 discusses implementation issues and open challenges.

References

Abadi, M., A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. (2016). "Deep learning with differential privacy". In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM. 308–318.

Abowd, J. M. (2018). "The US Census Bureau adopts differential privacy". In: *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 2867–2867.

Agarwal, A., M. Herlihy, S. Kamara, and T. Moataz. (2018). "Encrypted Databases for Differential Privacy". In: *IACR Cryptology ePrint Archive*.

Agarwal, S., B. Mozafari, A. Panda, H. Milner, S. Madden, and I. Stoica. (2013). "BlinkDB: Queries with Bounded Errors and Bounded Response Times on Very Large Data". In: *EuroSys*.

Arapinis, M., D. Figueira, and M. Gaboardi. (2016). "Sensitivity of Counting Queries". In: *ICALP*. 120:1–120:13.

Balle, B., G. Barthe, and M. Gaboardi. (2018). "Privacy Amplification by Subsampling: Tight Analyses via Couplings and Divergences". In: *Advances in Neural Information Processing Systems 31*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Curran Associates, Inc. 6277–6287. URL: http://papers.nips.cc/paper/7865-privacy-amplification-by-subsampling-tight-analyses-via-couplings-and-divergences.pdf.

Balle, B., J. Bell, A. Gascon, and K. Nissim. (2020). "Private summation in the multi-message shuffle model". *arXiv preprint arXiv:2002.00817*.

Balle, B., J. Bell, A. Gascón, and K. Nissim. (2019). "The privacy blanket of the shuffle model". In: *Annual International Cryptology Conference*. Springer. 638–667.

Barthe, G., G. P. Farina, M. Gaboardi, E. J. G. Arias, A. Gordon, J. Hsu, and P. Strub. (2016). "Differentially Private Bayesian Programming". In: *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, October 24-28, 2016*. Ed. by E. R. Weippl, S. Katzenbeisser, C. Kruegel, A. C. Myers, and S. Halevi. ACM. 68–79. DOI: 10.1145/2976749.2978371.

Bater, J., G. Elliott, C. Eggen, S. Goel, A. Kho, and J. Rogers. (2017). "SMCQL: Secure Querying for Federated Databases". *pvldb*. 10(6): 673–684. DOI: 10.14778/3055330.3055334.

Bittau, A., Ú. Erlingsson, P. Maniatis, I. Mironov, A. Raghunathan, D. Lie, M. Rudominer, U. Kode, J. Tinnés, and B. Seefeld. (2017). "Prochlo: Strong Privacy for Analytics in the Crowd". In: *Proceedings of the 26th Symposium on Operating Systems Principles, Shanghai, China, October 28-31, 2017*. 441–459.

Blocki, J., A. Blum, A. Datta, and O. Sheffet. (2013). "Differentially private data analysis of social networks via restricted sensitivity". In: *Innovations in Theoretical Computer Science, ITCS '13, Berkeley, CA, USA, January 9-12, 2013*. Ed. by R. D. Kleinberg. ACM. 87–96. DOI: 10.1145/2422436.2422449.

Bun, M., C. Dwork, G. N. Rothblum, and T. Steinke. (2018). "Composable and versatile privacy via truncated CDP". In: *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*. ACM. 74–86.

Bun, M. and T. Steinke. (2016). "Concentrated differential privacy: Simplifications, extensions, and lower bounds". In: *Theory of Cryptography Conference*. Springer. 635–658.

Chan, T. H., E. Shi, and D. Song. (2010). "Private and continual release of statistics". In: *International Colloquium on Automata, Languages, and Programming*. Springer. 405–417.

Chatzikokolakis, K., M. E. Andrés, N. E. Bordenabe, and C. Palamidessi. (2013). "Broadening the Scope of Differential Privacy Using Metrics". In: *Privacy Enhancing Technologies - 13th International Symposium, PETS 2013, Bloomington, IN, USA, July 10-12, 2013. Proceedings*. Ed. by E. D. Cristofaro and M. K. Wright. Vol. 7981. *Lecture Notes in Computer Science*. Springer. 82–102. DOI: 10.1007/978-3-642-39077-7_5.

Chaum, D. L. (1981). "Untraceable Electronic Mail, Return Addresses, and Digital Pseudonyms". *Commun. ACM*. 24(2): 84–90. DOI: 10.1145/358549.358563.

Chen, S. and S. Zhou. (2013). "Recursive Mechanism: Towards Node Differential Privacy and Unrestricted Joins". In: *ACM SIGMOD*.

Chen, Y., A. Machanavajjhala, M. Hay, and G. Miklau. (2017). "Pegasus: Data-adaptive differentially private stream processing". In: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. 1375–1388.

Cummings, R., S. Krehbiel, K. A. Lai, and U. T. Tantipongpipat. (2018). "Differential Privacy for Growing Databases". In: *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*. Ed. by S. Bengio, H. M. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. 8878–8887. URL: https://proceedings.neurips.cc/paper/2018/hash/ac27b77292582bc293a51055bfc994ee-Abstract.html.

Dinur, I. and K. Nissim. (2003). "Revealing information while preserving privacy". In: *Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*. 202–210.

Dwork, C. (2006). "Differential Privacy". In: *33rd International Colloquium on Automata, Languages and Programming, part II (ICALP 2006)*. Vol. 4052. *Lecture Notes in Computer Science*. Springer Verlag. 1–12. URL: https://www.microsoft.com/en-us/research/publication/differential-privacy/.

Dwork, C. (2011). "A firm foundation for private data analysis". *Commun. ACM*. 54(1): 86–95. DOI: 10.1145/1866739.1866758.

Dwork, C., K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor. (2006a). "Our Data, Ourselves: Privacy Via Distributed Noise Generation." In: *EUROCRYPT*. Ed. by S. Vaudenay. Vol. 4004. *Lecture Notes in Computer Science*. Springer. 486–503. URL: http://dblp.uni-trier.de/db/conf/eurocrypt/eurocrypt2006.html#DworkKMMN06.

Dwork, C. and J. Lei. (2009). "Differential privacy and robust statistics". In: *Proceedings of the 41st Annual ACM Symposium on Theory of Computing, STOC 2009, Bethesda, MD, USA, May 31 - June 2, 2009*. Ed. by M. Mitzenmacher. ACM. 371–380. DOI: 10.1145/1536414.1536466.

Dwork, C., F. McSherry, K. Nissim, and A. Smith. (2006b). "Calibrating Noise to Sensitivity in Private Data Analysis". In: *Theory of Cryptography*. Ed. by S. Halevi and T. Rabin. Berlin, Heidelberg: Springer Berlin Heidelberg. 265–284.

Dwork, C., M. Naor, T. Pitassi, and G. N. Rothblum. (2010). "Differential Privacy Under Continual Observation". In: *Proceedings of the Forty-second ACM Symposium on Theory of Computing*. *STOC '10*.

Dwork, C., M. Naor, O. Reingold, G. N. Rothblum, and S. Vadhan. (2009). "On the Complexity of Differentially Private Data Release: Efficient Algorithms and Hardness Results". In: *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*. *STOC '09*. Bethesda, MD, USA: Association for Computing Machinery. 381–390. DOI: 10.1145/1536414.1536467.

Dwork, C., A. Roth, *et al.* (2014). "The algorithmic foundations of differential privacy." *Foundations and Trends in Theoretical Computer Science*. 9(3-4): 211–407.

Ebadi, H., D. Sands, and G. Schneider. (2015). "Differential Privacy: Now it's Getting Personal". In: *Proceedings of the 42nd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2015, Mumbai, India, January 15-17, 2015*. Ed. by S. K. Rajamani and D. Walker. ACM. 69–81. DOI: 10.1145/2676726.2677005.

Erlingsson, Ú., V. Feldman, I. Mironov, A. Raghunathan, K. Talwar, and A. Thakurta. (2019). "Amplification by Shuffling: From Local to Central Differential Privacy via Anonymity". *Annual ACM-SIAM Symposium on Discrete Algorithms*: 2468–2479.

Erlingsson, Ú., V. Pihur, and A. Korolova. (2014). "RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response". In: *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security, Scottsdale, AZ, USA, November 3-7, 2014*. 1054–1067. DOI: 10.1145/2660267.2660348.

Gaboardi, M., E. J. G. Arias, J. Hsu, A. Roth, and Z. S. Wu. (2014). "Dual query: Practical private query release for high dimensional data". In: *International Conference on Machine Learning*. 1170–1178.

Garfinkel, S. L. and P. Leclerc. (2020). "Randomness Concerns when Deploying Differential Privacy". In: *WPES'20: Proceedings of the 19th Workshop on Privacy in the Electronic Society, Virtual Event, USA, November 9, 2020*. Ed. by J. Ligatti, X. Ou, W. Lueks, and P. Syverson. ACM. 73–86. DOI: 10.1145/3411497.3420211.

Ge, C., X. He, I. F. Ilyas, and A. Machanavajjhala. (2019). "APEx: Accuracy-Aware Differentially Private Data Exploration". In: *Proceedings of the 2019 International Conference on Management of Data, SIGMOD Conference 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019*. Ed. by P. A. Boncz, S. Manegold, A. Ailamaki, A. Deshpande, and T. Kraska. ACM. 177–194.

Ghazi, B., P. Manurangsi, R. Pagh, and A. Velingker. (2020). "Private aggregation from fewer anonymous messages". In: *Annual International Conference on the Theory and Applications of Cryptographic Techniques*. Springer. 798–827.

Haeberlen, A., B. C. Pierce, and A. Narayan. (2011). "Differential Privacy Under Fire". In: *20th USENIX Security Symposium, San Francisco, CA, USA, August 8-12, 2011, Proceedings*. USENIX Association. URL: http://static.usenix.org/events/sec11/tech/full%5C_papers/Haeberlen.pdf.

Hardt, M., K. Ligett, and F. McSherry. (2012). "A simple and practical algorithm for differentially private data release". In: *Advances in Neural Information Processing Systems*. 2339–2347.

Hay, M., A. Machanavajjhala, G. Miklau, Y. Chen, and D. Zhang. (2016a). "Principled evaluation of differentially private algorithms using dpbench". In: *Proceedings of the 2016 International Conference on Management of Data*. 139–154.

Hay, M., A. Machanavajjhala, G. Miklau, Y. Chen, D. Zhang, and G. Bissias. (2016b). "Exploring Privacy-Accuracy Tradeoffs using DPComp". In: *Proceedings of the 2016 International Conference on Management of Data, SIGMOD Conference 2016, San Francisco, CA, USA, June 26 - July 01, 2016*. Ed. by F. Özcan, G. Koutrika, and S. Madden. ACM. 2101–2104. DOI: 10.1145/2882903.2899387.

He, X., A. Machanavajjhala, and B. Ding. (2014). "Blowfish privacy: tuning privacy-utility trade-offs using policies". In: *International Conference on Management of Data, SIGMOD 2014, Snowbird, UT, USA, June 22-27, 2014*. Ed. by C. E. Dyreson, F. Li, and M. T. Özsu. ACM. 1447–1458. DOI: 10.1145/2588555.2588581.

Hsu, J., M. Gaboardi, A. Haeberlen, S. Khanna, A. Narayan, B. C. Pierce, and A. Roth. (2014). "Differential Privacy: An Economic Method for Choosing Epsilon". In: *IEEE 27th Computer Security Foundations Symposium, CSF 2014, Vienna, Austria, 19-22 July, 2014*. IEEE Computer Society. 398–410. DOI: 10.1109/CSF.2014.35.

"IBM Differential Privacy Library". https://github.com/IBM/differential-privacy-library.

Ilvento, C. (2020). "Implementing the Exponential Mechanism with Base-2 Differential Privacy". In: *CCS '20: 2020 ACM SIGSAC Conference on Computer and Communications Security, Virtual Event, USA, November 9-13, 2020*. Ed. by J. Ligatti, X. Ou, J. Katz, and G. Vigna.

Johnson, N. M., J. P. Near, J. M. Hellerstein, and D. Song. (2020). "Chorus: a Programming Framework for Building Scalable Differential Privacy Mechanisms". In: *IEEE European Symposium on Security and Privacy, EuroS&P 2020, Genoa, Italy, September 7-11, 2020*. IEEE. 535–551. DOI: 10.1109/EuroSP48549.2020.00041.

Johnson, N. M., J. P. Near, and D. Song. (2018). "Towards Practical Differential Privacy for SQL Queries". *Proc. VLDB Endow.* 11(5): 526–539. DOI: 10.1145/3187009.3177733.

Karwa, V., S. Raskhodnikova, A. Smith, and G. Yaroslavtsev. (2011). "Private Analysis of Graph Structure". In: *PVLDB*.

Kasiviswanathan, S. P., H. K. Lee, K. Nissim, S. Raskhodnikova, and A. D. Smith. (2011). "What Can We Learn Privately?" *SIAM J. Comput.* 40(3): 793–826. DOI: 10.1137/090756090.

Kasiviswanathan, S. P., K. Nissim, S. Raskhodnikova, and A. D. Smith. (2013). "Analyzing Graphs with Node Differential Privacy". In: *Theory of Cryptography - 10th Theory of Cryptography Conference, TCC 2013, Tokyo, Japan, March 3-6, 2013. Proceedings*. Ed. by A. Sahai. Vol. 7785. *Lecture Notes in Computer Science*. Springer. 457–476. DOI: 10.1007/978-3-642-36594-2_26.

Kifer, D. and A. Machanavajjhala. (2014). "Pufferfish: A framework for mathematical privacy definitions". *ACM Trans. Database Syst.* 39(1): 3:1–3:36. DOI: 10.1145/2514689.

Korolova, A., K. Kenthapadi, N. Mishra, and A. Ntoulas. (2009). "Releasing search queries and clicks privately". In: *Proceedings of the 18th International Conference on World Wide Web, WWW 2009, Madrid, Spain, April 20-24, 2009*. Ed. by J. Quemada, G. León, Y. S. Maarek, and W. Nejdl. ACM. 171–180. DOI: 10.1145/1526709.1526733.

Kotsogiannis, I., Y. Tao, X. He, M. Fanaeepour, A. Machanavajjhala, M. Hay, and G. Miklau. (2019). "PrivateSQL: A Differentially Private SQL Query Engine". *Proc. VLDB Endow.* 12(11): 1371–1384. DOI: 10.14778/3342263.3342274.

Lee, J. and C. Clifton. (2011). "How Much Is Enough? Choosing ϵ for Differential Privacy". In: *Information Security, 14th International Conference, ISC 2011, Xi'an, China, October 26-29, 2011. Proceedings*. Ed. by X. Lai, J. Zhou, and H. Li. Vol. 7001. *Lecture Notes in Computer Science*. Springer. 325–340. DOI: 10.1007/978-3-642-24861-0_22.

Lee, J. and C. W. Clifton. (2014). "Top-k Frequent Itemsets via Differentially Private FP-Trees". In: *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '14*. New York, New York, USA: Association for Computing Machinery. 931–940. DOI: 10.1145/2623330.2623723.

Li, C., G. Miklau, M. Hay, A. McGregor, and V. Rastogi. (2015). "The matrix mechanism: optimizing linear counting queries under differential privacy". *VLDB J.* 24(6): 757–781. DOI: 10.1007/s00778-015-0398-x.

Liu, C., S. Chakraborty, and P. Mittal. (2016). "Dependence Makes You Vulnerable: Differential Privacy Under Dependent Tuples". In: *23rd Annual Network and Distributed System Security Symposium, NDSS 2016, San Diego, California, USA, February 21-24, 2016*. The Internet Society. URL: http://wp.internetsociety.org/ndss/wp-content/uploads/sites/25/2017/09/dependence-makes-you-vulnerable-differential-privacy-under-dependent-tuples.pdf.

Lu, W., G. Miklau, and V. Gupta. (2014). "Generating Private Synthetic Databases for Untrusted System Evaluation". In: *ICDE*.

Lyu, M., D. Su, and N. Li. (2017). "Understanding the Sparse Vector Technique for Differential Privacy". *Proc. VLDB Endow.*

Machanavajjhala, A., D. Kifer, J. Gehrke, and M. Venkitasubramaniam. (2007). "l-diversity: Privacy beyond k-anonymity". *ACM Transactions on Knowledge Discovery from Data (TKDD)*. 1(1): 3–es.

McKenna, R., G. Miklau, M. Hay, and A. Machanavajjhala. (2018). "Optimizing error of high-dimensional statistical queries under differential privacy". *PVLDB*. 11(10): 1206–1219. DOI: 10.14778/3231751.3231769.

McGregor, A., I. Mironov, T. Pitassi, O. Reingold, K. Talwar, and S. Vadhan. (2010). "The Limits of Two-Party Differential Privacy". In: *Annual Symposium on Foundations of Computer Science*. IEEE.

McKenna, R., R. K. Maity, A. Mazumdar, and G. Miklau. (2020). "A workload-adaptive mechanism for linear queries under local differential privacy". *Proc. VLDB Endow.* 13(11): 1905–1918. URL: http://www.vldb.org/pvldb/vol13/p1905-mckenna.pdf.

McKenna, R., D. Sheldon, and G. Miklau. (2019). "Graphical-model based estimation and inference for differential privacy". In: *Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA*. Ed. by K. Chaudhuri and R. Salakhutdinov. Vol. 97. *Proceedings of Machine Learning Research*. PMLR. 4435–4444. URL: http://proceedings.mlr.press/v97/mckenna19a.html.

McSherry, F. and K. Talwar. (2007). “Mechanism Design via Differential Privacy”. In: *Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE. URL: https://www.microsoft.com/en-us/research/publication/mechanism-design-via-differential-privacy/.

McSherry, F. D. (2009). “Privacy Integrated Queries: An Extensible Platform for Privacy-Preserving Data Analysis”. In: *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data. SIGMOD ’09*. Providence, Rhode Island, USA. 19–30.

Mironov, I. (2012). “On significance of the least significant bits for differential privacy”. In: *Proceedings of the 2012 ACM conference on Computer and communications security*. 650–661.

Mironov, I. (2017). “Renyi differential privacy”. In: *Computer Security Foundations Symposium (CSF), 2017 IEEE 30th*. IEEE. 263–275.

Mironov, I., O. Pandey, O. Reingold, and S. Vadhan. (2009). “Computational Differential Privacy”. In: *Proceedings of the 29th Annual International Cryptology Conference on Advances in Cryptology. CRYPTO ’09*. Santa Barbara, CA: Springer-Verlag. 126–142. DOI: 10.1007/978-3-642-03356-8_8.

Mohan, P., A. Thakurta, E. Shi, D. Song, and D. E. Culler. (2012). “GUPT: privacy preserving data analysis made easy”. In: *Proceedings of the ACM SIGMOD International Conference on Management of Data, SIGMOD 2012, Scottsdale, AZ, USA, May 20-24, 2012*. 349–360. DOI: 10.1145/2213836.2213876.

Murtagh, J., K. Taylor, G. Kellaris, and S. Vadhan. (2018). “Usable Differential Privacy: A Case Study with PSI”. arXiv: 1809.04103 [cs.HC].

Narayan, A. and A. Haeberlen. (2012). “DJoin: Differentially Private Join Queries over Distributed Databases”. In: *10th USENIX Symposium on Operating Systems Design and Implementation*. 149–162. URL: https://www.usenix.org/conference/osdi12/technical-sessions/presentation/narayan.

Nissim, K., S. Raskhodnikova, and A. D. Smith. (2007). “Smooth sensitivity and sampling in private data analysis”. In: *Proceedings of the 39th Annual ACM Symposium on Theory of Computing, San Diego, California, USA, June 11-13, 2007*. Ed. by D. S. Johnson and U. Feige. ACM. 75–84. DOI: 10.1145/1250790.1250803.

Ohm, P. (2009). "Broken promises of privacy: Responding to the surprising failure of anonymization". *UCLA l. Rev.* 57: 1701.

"OpenDP". https://privacytools.seas.harvard.edu/opendp.

Raskhodnikova, S. and A. D. Smith. (2015). "Efficient Lipschitz Extensions for High-Dimensional Graph Statistics and Node Private Degree Distributions". *CoRR*. abs/1504.07912. arXiv: 1504.07912. URL: http://arxiv.org/abs/1504.07912.

Roth, E., D. Noble, B. H. Falk, and A. Haeberlen. (2019). "Honeycrisp: large-scale differentially private aggregation without a trusted core". In: *Proceedings of the 27th ACM Symposium on Operating Systems Principles, SOSP 2019, Huntsville, ON, Canada, October 27-30, 2019*. Ed. by T. Brecht and C. Williamson. ACM. 196–210. DOI: 10.1145/3341301.3359660.

Roy, I., S. T. V. Setty, A. Kilzer, V. Shmatikov, and E. Witchel. (2010). "Airavat: Security and Privacy for MapReduce". In: *Proceedings of the 7th USENIX Symposium on Networked Systems Design and Implementation, NSDI 2010, April 28-30, 2010, San Jose, CA, USA*. USENIX Association. 297–312. URL: http://www.usenix.org/events/nsdi10/tech/full%5C_papers/roy.pdf.

"Smart Noise". https://github.com/opendp/smartnoise-samples.

Smith, A. D. (2011). "Privacy-preserving statistical estimation with optimal convergence rates". In: *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*. Ed. by L. Fortnow and S. P. Vadhan. ACM. 813–822. DOI: 10.1145/1993636.1993743.

Song, S., Y. Wang, and K. Chaudhuri. (2017). "Pufferfish Privacy Mechanisms for Correlated Data". In: *Proceedings of the 2017 ACM International Conference on Management of Data, SIGMOD Conference 2017, Chicago, IL, USA, May 14-19, 2017*. Ed. by S. Salihoglu, W. Zhou, R. Chirkova, J. Yang, and D. Suciu. ACM. 1291–1306. DOI: 10.1145/3035918.3064025.

Sweeney, L. (2000). "Simple demographics often identify people uniquely". *Health (San Francisco)*. 671(2000): 1–34.

Sweeney, L. (2002). "k-anonymity: A model for protecting privacy". *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*. 10(05): 557–570.

Tang, J., A. Korolova, X. Bai, X. Wang, and X. Wang. (2017). "Privacy Loss in Apple's Implementation of Differential Privacy on MacOS 10.12". *CoRR*. abs/1709.02753. arXiv: 1709.02753. URL: http://arxiv.org/abs/1709.02753.

Tao, Y., X. He, A. Machanavajjhala, and S. Roy. (2020). "Computing Local Sensitivities of Counting Queries with Joins". In: *Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA], June 14-19, 2020*. Ed. by D. Maier, R. Pottinger, A. Doan, W. Tan, A. Alawini, and H. Q. Ngo. ACM. 479–494. DOI: 10.1145/3318464.3389762.

Vesga, E. L., A. Russo, and M. Gaboardi. (2020). "A Programming Framework for Differential Privacy with Accuracy Concentration Bounds". In: *2020 IEEE Symposium on Security and Privacy, SP 2020, San Francisco, CA, USA, May 18-21, 2020*. IEEE. 411–428. DOI: 10.1109/SP40000.2020.00086.

Wang, T., J. Blocki, N. Li, and S. Jha. (2017). "Locally Differentially Private Protocols for Frequency Estimation". In: *26th USENIX Security Symposium, USENIX Security 2017, Vancouver, BC, Canada, August 16-18, 2017*. Ed. by E. Kirda and T. Ristenpart. USENIX Association. 729–745. URL: https://www.usenix.org/conference/usenixsecurity17/technical-sessions/presentation/wang-tianhao.

Wang, T., B. Ding, J. Zhou, C. Hong, Z. Huang, N. Li, and S. Jha. (2019). "Answering Multi-Dimensional Analytical Queries under Local Differential Privacy". In: *Proceedings of the 2019 International Conference on Management of Data, SIGMOD Conference 2019, Amsterdam, The Netherlands, June 30 - July 5, 2019*. 159–176. DOI: 10.1145/3299869.3319891.

Warner, S. L. (1965). "Randomized response: A survey technique for eliminating evasive answer bias". *Journal of the American Statistical Association*. 60(309): 63–69.

Wilson, R. J., C. Y. Zhang, W. Lam, D. Desfontaines, D. Simmons-Marengo, and B. Gipson. (2020). "Differentially Private SQL with Bounded User Contribution". *Proceedings on Privacy Enhancing Technologies*. 2020(2): 230–250.

Yang, J., T. Wang, N. Li, X. Cheng, and S. Su. (2020). "Answering Multi-Dimensional Range Queries under Local Differential Privacy". *CoRR*. abs/2009.06538. arXiv: 2009.06538. URL: https://arxiv.org/abs/2009.06538.

Zhang, D., R. McKenna, I. Kotsogiannis, G. Bissias, M. Hay, A. Machanavajjhala, and G. Miklau. (2020). "ϵKTELO: A Framework for Defining Differentially Private Computations". *ACM Trans. Database Syst.*

Zhang, D. and D. Kifer. (2017). "LightDP: towards automating differential privacy proofs". In: *Proceedings of the 44th ACM SIGPLAN Symposium on Principles of Programming Languages, POPL 2017, Paris, France, January 18-20, 2017*. Ed. by G. Castagna and A. D. Gordon. ACM. 888–901. URL: http://dl.acm.org/citation.cfm?id=3009884.

Zhang, J., G. Cormode, C. M. Procopiuc, D. Srivastava, and X. Xiao. (2014). "PrivBayes: private data release via bayesian networks". In: *International Conference on Management of Data, SIGMOD 2014, Snowbird, UT, USA, June 22-27, 2014*. 1423–1434. DOI: 10.1145/2588555.2588573.

Zhang, Z., T. Wang, N. Li, S. He, and J. Chen. (2018). "CALM: Consistent Adaptive Local Marginal for Marginal Release under Local Differential Privacy". In: *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security, CCS 2018, Toronto, ON, Canada, October 15-19, 2018*. Ed. by D. Lie, M. Mannan, M. Backes, and X. Wang. ACM. 212–229. DOI: 10.1145/3243734.3243742.