Differential privacy

Differential privacy is a system for publicly sharing information about a dataset by describing the patterns of groups within the dataset while withholding information about individuals in the dataset. The idea behind differential privacy is that if the effect of making an arbitrary single substitution in the database is small enough, the query result cannot be used to infer much about any single individual, and therefore provides privacy. Another way to describe differential privacy is as a constraint on the algorithms used to publish aggregate information about a statistical database which limits the disclosure of private information of records whose information is in the database. For example, differentially private algorithms are used by some government agencies to publish demographic information or other statistical aggregates while ensuring confidentiality of survey responses, and by companies to collect information about user behavior while controlling what is visible even to internal analysts.

Roughly, an algorithm is differentially private if an observer seeing its output cannot tell if a particular individual's information was used in the computation. Differential privacy is often discussed in the context of identifying individuals whose information may be in a database. Although it does not directly refer to identification and reidentification attacks, differentially private algorithms probably resist such attacks.[1]

Differential privacy was developed by cryptographers and thus is often associated with cryptography, and draws much of its language from cryptography.

Contents

History

Official statistics organizations are charged with collecting information from individuals or establishments and publishing aggregate data to serve the public interest. For example, the 1790 United States Census collected information about individuals living in the United States and published tabulations based on sex, age, race, and condition of servitude. Statistical organizations have long collected information under a promise of confidentiality that the information provided will be used for statistical purposes, but that the publications will not produce information that can be traced back to a specific individual or establishment. To accomplish this goal, statistical organizations have long suppressed information in their publications. For example, in a table presenting the sales of each business in a town grouped by business category, a cell that has information from only one company might be suppressed, in order to maintain the confidentiality of that company's specific sales.

The adoption of electronic information processing systems by statistical agencies in the 1950s and 1960s dramatically increased the number of tables that a statistical organization could produce and, in so doing, significantly increased the potential for an improper disclosure of confidential information. For example, if a business that had its sales numbers suppressed also had those numbers appear in the total sales of a region, then it might be possible to determine the suppressed value by subtracting the other sales from that total. But there might also be combinations of additions and subtractions that might cause the private information to be revealed. The number of combinations that needed to be checked increases exponentially with the number of publications, and it is potentially unbounded if data users are able to make queries of the statistical database using an interactive query system.

In 1977, Tore Dalenius formalized the mathematics of cell suppression.[2]

In 1979, Dorothy Denning, Peter J. Denning and Mayer D. Schwartz formalized the concept of a Tracker, an adversary that could learn the confidential contents of a statistical database by creating a series of targeted queries and remembering the results.[3] This and future research showed that privacy properties in a database could only be preserved by considering each new query in light of (possibly all) previous queries. This line of work is sometimes called *query privacy*, with the final result being that tracking the impact of a query on the privacy of individuals in the database was NP-hard.

In 2003, Kobbi Nissim and Irit Dinur demonstrated that it is impossible to publish arbitrary queries on a private statistical database without revealing some amount of private information, and that the entire information content of the database can be revealed by publishing the results of a surprisingly small number of random queries—far fewer than was implied by previous work.[4] The general phenomenon is known as the Fundamental Law of Information Recovery, and its key insight, namely that in the most general case, privacy cannot be protected without injecting some amount of noise, led to development of differential privacy.

In 2006, Cynthia Dwork, Frank McSherry, Kobbi Nissim and Adam D. Smith published an article formalizing the amount of noise that needed to be added and proposing a generalized mechanism for doing so.[1] Their work was a co-recipient of the 2016 TCC Test-of-Time Award[5] and the 2017 Gödel Prize.[6]

Since then, subsequent research has shown that there are many ways to produce very accurate statistics from the database while still ensuring high levels of privacy.[7][8]

ε-differential privacy

The 2006 Dwork, McSherry, Nissim and Smith article introduced the concept of ε-differential privacy, a mathematical definition for the privacy loss associated with any data release drawn from a statistical database. (Here, the term *statistical database* means a set of data that are collected under the pledge of confidentiality for the purpose of producing statistics that, by their production, do not compromise the privacy of those individuals who provided the data.)

The intuition for the 2006 definition of ε-differential privacy is that a person's privacy cannot be compromised by a statistical release if their data are not in the database. Therefore, with differential privacy, the goal is to give each individual roughly the same privacy that would result from having their data removed. That is, the statistical functions run on the database should not overly depend on the data of any one individual.

Of course, how much any individual contributes to the result of a database depends in part on how many people's data are involved in the query. If the database contains data from a single person, that person's data contributes 100%. If the database contains data from a hundred people, each person's data contributes just 1%. The key insight of differential privacy is that as the query is made on the data of fewer and fewer people, more noise needs to be added to the query result to produce the same amount of privacy. Hence the name of the 2006 paper, "Calibrating noise to sensitivity in private data analysis."

The 2006 paper presents both a mathematical definition of differential privacy and a mechanism based on the addition of Laplace noise (i.e. noise coming from the Laplace distribution) that satisfies the definition.

Definition of ε-differential privacy

Let ε be a positive real number and $\mathcal{A}$ be a randomized algorithm that takes a dataset as input (representing the actions of the trusted party holding the data). Let $\operatorname{im} \mathcal{A}$ denote the image of $\mathcal{A}$. The algorithm $\mathcal{A}$ is said to provide ϵ-differential privacy if, for all datasets D_1 and D_2 that differ on a single element (i.e., the data of one person), and all subsets S of $\operatorname{im} \mathcal{A}$:

$$\Pr[\mathcal{A}(D_1) \in S] \leq \exp(\epsilon) \cdot \Pr[\mathcal{A}(D_2) \in S],$$

where the probability is taken over the randomness used by the algorithm.[9]

Differential privacy offers strong and robust guarantees that facilitate modular design and analysis of differentially private mechanisms due to its composability, robustness to post-processing, and graceful degradation in the presence of correlated data.

Composability

(Self-)composability refers to the fact that the joint distribution of the outputs of (possibly adaptively chosen) differentially private mechanisms satisfies differential privacy.

Sequential composition. If we query an ε-differential privacy mechanism t times, and the randomization of the mechanism is independent for each query, then the result would be ϵt-differentially private. In the more general case, if there are n independent mechanisms: $\mathcal{M}_1, \ldots, \mathcal{M}_n$, whose privacy guarantees are $\epsilon_1, \ldots, \epsilon_n$ differential privacy, respectively, then any function g of them: $g(\mathcal{M}_1, \ldots, \mathcal{M}_n)$ is $\left(\sum_{i=1}^{n} \epsilon_i\right)$-differentially private.[10]

Parallel composition. If the previous mechanisms are computed on *disjoint* subsets of the private database then the function g would be $(\max_i \epsilon_i)$-differentially private instead.[10]

Robustness to post-processing

For any deterministic or randomized function F defined over the image of the mechanism $\mathcal{A}$, if $\mathcal{A}$ satisfies ε-differential privacy, so does $F(\mathcal{A})$.

Together, composability and robustness to post-processing permit modular construction and analysis of differentially private mechanisms and motivate the concept of the *privacy loss budget*. If all elements that access sensitive data of a complex mechanisms are separately differentially private, so will be their combination, followed by arbitrary post-processing.

Group privacy

In general, ε-differential privacy is designed to protect the privacy between neighboring databases which differ only in one row. This means that no adversary with arbitrary auxiliary information can know if **one** particular participant submitted his information. However this is also extendable if we want to protect databases differing in c rows, which amounts to adversary with arbitrary auxiliary information can know if c particular participants submitted their information. This can be achieved because if c items change, the probability dilation is bounded by $\exp(\epsilon c)$ instead of $\exp(\epsilon)$,[11] i.e., for D_1 and D_2 differing on c items:

$$\Pr[\mathcal{A}(D_1) \in S] \leq \exp(\epsilon c) \cdot \Pr[\mathcal{A}(D_2) \in S]$$

Thus setting ε instead to ϵ/c achieves the desired result (protection of c items). In other words, instead of having each item ε-differentially private protected, now every group of c items is ε-differentially private protected (and each item is (ϵ/c)-differentially private protected).

ε-differentially private mechanisms

Since differential privacy is a probabilistic concept, any differentially private mechanism is necessarily randomized. Some of these, like the Laplace mechanism, described below, rely on adding controlled noise to the function that we want to compute. Others, like the exponential mechanism[12] and posterior sampling[13] sample from a problem-dependent family of distributions instead.

Sensitivity

Let d be a positive integer, $\mathcal{D}$ be a collection of datasets, and $f\colon \mathcal{D} \to \mathbb{R}^d$ be a function. The *sensitivity* [1] of a function, denoted Δf, is defined by

$$\Delta f = \max \|f(D_1) - f(D_2)\|_1,$$

where the maximum is over all pairs of datasets D_1 and D_2 in $\mathcal{D}$ differing in at most one element and $\|\cdot\|_1$ denotes the ℓ_1 norm.

In the example of the medical database below, if we consider f to be the function Q_i, then the sensitivity of the function is one, since changing any one of the entries in the database causes the output of the function to change by either zero or one.

There are techniques (which are described below) using which we can create a differentially private algorithm for functions with low sensitivity.

The Laplace mechanism

The Laplace mechanism adds Laplace noise (i.e. noise from the Laplace distribution, which can be expressed by probability density function $\text{noise}(y) \propto \exp(-|y|/\lambda)$, which has mean zero and standard deviation $\sqrt{2}\lambda$). Now in our case we define the output function of $\mathcal{A}$ as a real valued function (called as the transcript output

by $\mathcal{A}$) as $\mathcal{T}_{\mathcal{A}}(x) = f(x) + Y$ where $Y \sim \mathbf{Lap}(\lambda)$ and f is the original real valued query/function we planned to execute on the database. Now clearly $\mathcal{T}_{\mathcal{A}}(x)$ can be considered to be a continuous random variable, where

$$\frac{\mathbf{pdf}(\mathcal{T}_{\mathcal{A},D_1}(x) = t)}{\mathbf{pdf}(\mathcal{T}_{\mathcal{A},D_2}(x) = t)} = \frac{\mathbf{noise}(t - f(D_1))}{\mathbf{noise}(t - f(D_2))}$$

which is at most $e^{\frac{|f(D_1) - f(D_2)|}{\lambda}} \leq e^{\frac{\Delta(f)}{\lambda}}$. We can consider $\frac{\Delta(f)}{\lambda}$ to be the privacy factor ϵ. Thus $\mathcal{T}$ follows a differentially private mechanism (as can be seen from the definition above). If we try to use this concept in our diabetes example then it follows from the above derived fact that in order to have $\mathcal{A}$ as the ϵ-differential private algorithm we need to have $\lambda = 1/\epsilon$. Though we have used Laplace noise here, other forms of noise, such as the Gaussian Noise, can be employed, but they may require a slight relaxation of the definition of differential privacy.[11]

According to this definition, differential privacy is a condition on the release mechanism (i.e., the trusted party releasing information *about* the dataset) and not on the dataset itself. Intuitively, this means that for any two datasets that are similar, a given differentially private algorithm will behave approximately the same on both datasets. The definition gives a strong guarantee that presence or absence of an individual will not affect the final output of the algorithm significantly.

For example, assume we have a database of medical records D_1 where each record is a pair (**Name**, **X**), where X is a Boolean denoting whether a person has diabetes or not. For example:

Name	Has Diabetes (X)
Ross	1
Monica	1
Joey	0
Phoebe	0
Chandler	1
Rachel	0

Now suppose a malicious user (often termed an *adversary*) wants to find whether Chandler has diabetes or not. Suppose he also knows in which row of the database Chandler resides. Now suppose the adversary is only allowed to use a particular form of query Q_i that returns the partial sum of the first i rows of column X in the database. In order to find Chandler's diabetes status the adversary executes $Q_5(D_1)$ and $Q_4(D_1)$, then computes their difference. In this example, $Q_5(D_1) = 3$ and $Q_4(D_1) = 2$, so their difference is 1. This indicates that the "Has Diabetes" field in Chandler's row must be 1. This example highlights how individual information can be compromised even without explicitly querying for the information of a specific individual.

Continuing this example, if we construct D_2 by replacing (Chandler, 1) with (Chandler, 0) then this malicious adversary will be able to distinguish D_2 from D_1 by computing $Q_5 - Q_4$ for each dataset. If the adversary were required to receive the values Q_i via an ϵ-differentially private algorithm, for a sufficiently small ϵ, then he or she would be unable to distinguish between the two datasets.

Randomized response

A simple example, especially developed in the social sciences,[14] is to ask a person to answer the question "Do you own the *attribute A*?", according to the following procedure:

1. Toss a coin.
2. If heads, then toss the coin again (ignoring the outcome), and answer the question honestly.
3. If tails, then toss the coin again and answer "Yes" if heads, "No" if tails.

(The seemingly redundant extra toss in the first case is needed in situations where just the *act* of tossing a coin may be observed by others, even if the actual result stays hidden.) The confidentiality then arises from the refutability of the individual responses.

But, overall, these data with many responses are significant, since positive responses are given to a quarter by people who do not have the *attribute A* and three-quarters by people who actually possess it. Thus, if p is the true proportion of people with A, then we expect to obtain $(1/4)(1-p) + (3/4)p = (1/4) + p/2$ positive responses. Hence it is possible to estimate p.

In particular, if the *attribute A* is synonymous with illegal behavior, then answering "Yes" is not incriminating, insofar as the person has a probability of a "Yes" response, whatever it may be.

Although this example, inspired by randomized response, might be applicable to microdata (i.e., releasing datasets with each individual response), by definition differential privacy excludes microdata release and is only applicable to queries (i.e., aggregating individual responses into one result) as this would violate the requirements, more specifically the plausible deniability that a subject participated or not.[15][16]

Stable transformations

A transformation T is c-stable if the hamming distance between $T(A)$ and $T(B)$ is at most c-times the hamming distance between A and B for any two databases A, B. Theorem 2 in [10] asserts that if there is a mechanism M that is ϵ-differentially private, then the composite mechanism $M \circ T$ is $(\epsilon \times c)$-differentially private.

This could be generalized to group privacy, as the group size could be thought of as the hamming distance h between A and B (where A contains the group and B doesn't). In this case $M \circ T$ is $(\epsilon \times c \times h)$-differentially private.

Other notions of differential privacy

Since differential privacy is considered to be too strong or weak for some applications, many versions of it have been proposed.[17] The most widespread relaxation is (ε, δ)-differential privacy,[18] which weakens the definition by allowing an additional small δ density of probability on which the upper bound ε does not hold.

Adoption of differential privacy in real-world applications

Several uses of differential privacy in practice are known to date:

- 2008: U.S. Census Bureau, for showing commuting patterns.[19]
- 2014: Google's RAPPOR, for telemetry such as learning statistics about unwanted software hijacking users' settings [20] (RAPPOR's open-source implementation (https://github.com/googl

e/rappor)).
- 2015: Google, for sharing historical traffic statistics.[21]
- 2016: Apple announced its intention to use differential privacy in iOS 10 to improve its Intelligent personal assistant technology.[22]
- 2017: Microsoft, for telemetry in Windows.[23]
- 2019: Privitar Lens is an API using differential privacy.[24]
- 2020: LinkedIn, for advertiser queries.[25]

See also

- Quasi-identifier
- Exponential mechanism (differential privacy) – a technique for designing differentially private algorithms
- k-anonymity
- Differentially private analysis of graphs

References

1. Calibrating Noise to Sensitivity in Private Data Analysis (https://link.springer.com/chapter/10.1007%2F11681878_14) by Cynthia Dwork, Frank McSherry, Kobbi Nissim, Adam Smith. In Theory of Cryptography Conference (TCC), Springer, 2006. doi:10.1007/11681878_14 (https://doi.org/10.1007%2F11681878_14). The full version (https://journalprivacyconfidentiality.org/index.php/jpc/article/view/405) appears in Journal of Privacy and Confidentiality, 7 (3), 17-51. doi:10.29012/jpc.v7i3.405 (https://doi.org/10.29012%2Fjpc.v7i3.405)
2. Tore Dalenius (1977). "Towards a methodology for statistical disclosure control". *Statistik Tidskrift*. **15**.
3. Dorothy E. Denning; Peter J. Denning; Mayer D. Schwartz (March 1978). "The Tracker: A Threat to Statistical Database Security" (http://www.dbis.informatik.hu-berlin.de/fileadmin/lectures/SS2011/VL_Privacy/Tracker1.pdf) (PDF). **4** (1): 76–96.
4. Irit Dinur and Kobbi Nissim. 2003. Revealing information while preserving privacy. In Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems (PODS '03). ACM, New York, NY, USA, 202–210. doi:10.1145/773153.773173 (https://doi.org/10.1145%2F773153.773173)
5. "TCC Test-of-Time Award" (https://www.iacr.org/workshops/tcc/awards.html).
6. "2017 Gödel Prize" (https://www.eatcs.org/index.php/component/content/article/1-news/2450-2017-godel-prize).
7. Hilton, Michael. "Differential Privacy: A Historical Survey". S2CID 16861132 (https://api.semanticscholar.org/CorpusID:16861132).
8. Dwork, Cynthia (2008-04-25). "Differential Privacy: A Survey of Results" (https://www.microsoft.com/en-us/research/publication/differential-privacy-a-survey-of-results/). In Agrawal, Manindra; Du, Dingzhu; Duan, Zhenhua; Li, Angsheng (eds.). *Theory and Applications of Models of Computation*. Lecture Notes in Computer Science. **4978**. Springer Berlin Heidelberg. pp. 1–19. doi:10.1007/978-3-540-79228-4_1 (https://doi.org/10.1007%2F978-3-540-79228-4_1). ISBN 9783540792277.
9. The Algorithmic Foundations of Differential Privacy (http://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf) by Cynthia Dwork and Aaron Roth. Foundations and Trends in Theoretical Computer Science. Vol. 9, no. 3–4, pp. 211-407, Aug. 2014. doi:10.1561/0400000042 (https://doi.org/10.1561%2F0400000042)

10. Privacy integrated queries: an extensible platform for privacy-preserving data analysis (http://research.microsoft.com/pubs/80218/sigmod115-mcsherry.pdf) by Frank D. McSherry. In Proceedings of the 35th SIGMOD International Conference on Management of Data (SIGMOD), 2009. doi:10.1145/1559845.1559850 (https://doi.org/10.1145%2F1559845.1559850)
11. Differential Privacy (http://research.microsoft.com/pubs/64346/dwork.pdf) by Cynthia Dwork, International Colloquium on Automata, Languages and Programming (ICALP) 2006, p. 1–12. doi:10.1007/11787006 1 (https://doi.org/10.1007%2F11787006+1)
12. F.McSherry and K.Talwar. Mechasim Design via Differential Privacy. Proceedings of the 48th Annual Symposium of Foundations of Computer Science, 2007. (http://research.microsoft.com/pubs/65075/mdviadp.pdf)
13. Christos Dimitrakakis, Blaine Nelson, Aikaterini Mitrokotsa, Benjamin Rubinstein. Robust and Private Bayesian Inference. Algorithmic Learning Theory 2014 (https://arxiv.org/abs/1306.1066)
14. Warner, S. L. (March 1965). "Randomised response: a survey technique for eliminating evasive answer bias". *Journal of the American Statistical Association*. Taylor & Francis. **60** (309): 63–69. doi:10.1080/01621459.1965.10480775 (https://doi.org/10.1080%2F01621459.1965.10480775). JSTOR 2283137 (https://www.jstor.org/stable/2283137). PMID 12261830 (https://pubmed.ncbi.nlm.nih.gov/12261830).
15. Dwork, Cynthia. "A firm foundation for private data analysis." Communications of the ACM 54.1 (2011): 86–95, supra note 19, page 91.
16. Bambauer, Jane, Krishnamurty Muralidhar, and Rathindra Sarathy. "Fool's gold: an illustrated critique of differential privacy." Vand. J. Ent. & Tech. L. 16 (2013): 701.
17. SoK: Differential Privacies (https://arxiv.org/abs/1906.01337) by Damien Desfontaines, Balázs Pejó. 2019.
18. Dwork, Cynthia, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. "Our data, ourselves: Privacy via distributed noise generation." In Advances in Cryptology-EUROCRYPT 2006, pp. 486–503. Springer Berlin Heidelberg, 2006.
19. Ashwin Machanavajjhala, Daniel Kifer, John M. Abowd, Johannes Gehrke, and Lars Vilhuber. "Privacy: Theory meets Practice on the Map". In Proceedings of the 24th International Conference on Data Engineering, ICDE) 2008.
20. Úlfar Erlingsson, Vasyl Pihur, Aleksandra Korolova. "RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response". (https://dl.acm.org/doi/10.1145/2660267.2660348) In Proceedings of the 21st ACM Conference on Computer and Communications Security (CCS), 2014. doi:10.1145/2660267.2660348 (https://doi.org/10.1145%2F2660267.2660348)
21. Tackling Urban Mobility with Technology (https://europe.googleblog.com/2015/11/tackling-urban-mobility-with-technology.html) by Andrew Eland. Google Policy Europe Blog, Nov 18, 2015.
22. "Apple - Press Info - Apple Previews iOS 10, the Biggest iOS Release Ever" (https://www.apple.com/pr/library/2016/06/13Apple-Previews-iOS-10-The-Biggest-iOS-Release-Ever.html). *Apple*. Retrieved 16 June 2016.
23. Collecting telemetry data privately (https://www.microsoft.com/en-us/research/publication/collecting-telemetry-data-privately/) by Bolin Ding, Jana Kulkarni, Sergey Yekhanin. NIPS 2017.
24. "Privitar Lens" (https://www.privitar.com/privitar-lens). Retrieved 20 February 2018.
25. LinkedIn's Audience Engagements API: A Privacy Preserving Data Analytics System at Scale (https://arxiv.org/abs/2002.05839) by Ryan Rogers, Subbu Subramaniam, Sean Peng, David Durfee, Seunghyun Lee, Santosh Kumar Kancha, Shraddha Sahay, Parvez Ahammad. arXiv:2002.05839.

Further reading

- A reading list on differential privacy (https://desfontain.es/privacy/index.html)
- Abowd, John. 2017. “How Will Statistical Agencies Operate When All Data Are Private?”. Journal of Privacy and Confidentiality 7 (3). (https://journalprivacyconfidentiality.org/index.php/j

pc/article/view/404) doi:10.29012/jpc.v7i3.404 (https://doi.org/10.29012%2Fjpc.v7i3.404) (slides (https://www2.census.gov/cac/sac/meetings/2017-09/role-statistical-agency.pdf))

- "Differential Privacy: A Primer for a Non-technical Audience" (http://www.jetlaw.org/wp-content/uploads/2018/12/4_Wood_Final.pdf), Kobbi Nissim, Thomas Steinke, Alexandra Wood, Micah Altman, Aaron Bembenek, Mark Bun, Marco Gaboardi, David R. O'Brien, and Salil Vadhan, Harvard Privacy Tools Project, February 14, 2018
- Dinur, Irit and Kobbi Nissim. 2003. Revealing information while preserving privacy. In Proceedings of the twenty-second ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems(PODS '03). ACM, New York, NY, USA, 202-210. doi:10.1145/773153.773173 (https://doi.org/10.1145%2F773153.773173).
- Dwork, Cynthia, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. in Halevi, S. & Rabin, T. (Eds.) Calibrating Noise to Sensitivity in Private Data Analysis Theory of Cryptography: Third Theory of Cryptography Conference, TCC 2006, New York, NY, USA, March 4–7, 2006. Proceedings, Springer Berlin Heidelberg, 265-284, doi:10.1007/11681878 14 (https://doi.org/10.1007%2F11681878+14).
- Dwork, Cynthia. 2006. Differential Privacy, 33rd International Colloquium on Automata, Languages and Programming, part II (ICALP 2006), Springer Verlag, 4052, 1-12, ISBN 3-540-35907-9.
- Dwork, Cynthia and Aaron Roth. 2014. The Algorithmic Foundations of Differential Privacy. Foundations and Trends in Theoretical Computer Science. Vol. 9, Nos. 3–4. 211–407, doi:10.1561/0400000042 (https://doi.org/10.1561%2F0400000042).
- Machanavajjhala, Ashwin, Daniel Kifer, John M. Abowd, Johannes Gehrke, and Lars Vilhuber. 2008. Privacy: Theory Meets Practice on the Map, International Conference on Data Engineering (ICDE) 2008: 277-286, doi:10.1109/ICDE.2008.4497436 (https://doi.org/10.1109%2FICDE.2008.4497436).
- Dwork, Cynthia and Moni Naor. 2010. On the Difficulties of Disclosure Prevention in Statistical Databases or The Case for Differential Privacy, Journal of Privacy and Confidentiality: Vol. 2: Iss. 1, Article 8. Available at: http://repository.cmu.edu/jpc/vol2/iss1/8.
- Kifer, Daniel and Ashwin Machanavajjhala. 2011. No free lunch in data privacy. In Proceedings of the 2011 ACM SIGMOD International Conference on Management of data (SIGMOD '11). ACM, New York, NY, USA, 193-204. doi:10.1145/1989323.1989345 (https://doi.org/10.1145%2F1989323.1989345).
- Erlingsson, Úlfar, Vasyl Pihur and Aleksandra Korolova. 2014. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. In Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS '14). ACM, New York, NY, USA, 1054-1067. doi:10.1145/2660267.2660348 (https://doi.org/10.1145%2F2660267.2660348).
- Abowd, John M. and Ian M. Schmutte. 2017 . Revisiting the economics of privacy: Population statistics and confidentiality protection as public goods. Labor Dynamics Institute, Cornell University, Labor Dynamics Institute, Cornell University, at https://digitalcommons.ilr.cornell.edu/ldi/37/
- Abowd, John M. and Ian M. Schmutte. Forthcoming. An Economic Analysis of Privacy Protection and Statistical Accuracy as Social Choices. American Economic Review, arXiv:1808.06303 (https://arxiv.org/abs/1808.06303)
- Apple, Inc. 2016. Apple previews iOS 10, the biggest iOS release ever. Press Release (June 13). https://www.apple.com/newsroom/2016/06/apple-previews-ios-10-biggest-ios-release-ever.html.
- Ding, Bolin, Janardhan Kulkarni, and Sergey Yekhanin 2017. Collecting Telemetry Data Privately, NIPS 2017.
- http://www.win-vector.com/blog/2015/10/a-simpler-explanation-of-differential-privacy/
- Ryffel, Theo, Andrew Trask, et. al. "A generic framework for privacy preserving deep learning"

External links

- Differential Privacy (https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/dwork.pdf) by Cynthia Dwork, ICALP July 2006.
- The Algorithmic Foundations of Differential Privacy (http://www.cis.upenn.edu/~aaroth/Papers/privacybook.pdf) by Cynthia Dwork and Aaron Roth, 2014.
- Differential Privacy: A Survey of Results (http://research.microsoft.com/apps/pubs/default.aspx?id=74339) by Cynthia Dwork, Microsoft Research, April 2008
- Privacy of Dynamic Data: Continual Observation and Pan Privacy (http://video.ias.edu/csdm/dynamicdata) by Moni Naor, Institute for Advanced Study, November 2009
- Tutorial on Differential Privacy (http://simons.berkeley.edu/talks/katrina-ligett-2013-12-11) by Katrina Ligett, California Institute of Technology, December 2013
- A Practical Beginner's Guide To Differential Privacy (http://www.cerias.purdue.edu/news_and_events/events/security_seminar/details/index/j9cvs3as2h1qds1jrdqfdc3hu8) by Christine Task, Purdue University, April 2012
- Private Map Maker v0.2 (http://blog.myplaceinthecrowd.org/2011/04/27/the-cdp-private-map-maker-v0-2/) on the Common Data Project blog
- Learning Statistics with Privacy, aided by the Flip of a Coin (https://research.googleblog.com/2014/10/learning-statistics-with-privacy-aided.html) by Úlfar Erlingsson, Google Research Blog, October 2014
