Utility-Enhancing Flexible Mechanisms for Differential Privacy

Vaikkunth Mugunthan(✉), Wanyi Xiao, and Lalana Kagal

Massachusetts Institute of Technology, Cambridge, MA 02139, USA
vaik@mit.edu

Abstract. Differential privacy is a mathematical technique that provides strong theoretical privacy guarantees by ensuring the statistical indistinguishability of individuals in a dataset. It has become the de facto framework for providing privacy-preserving data analysis over statistical datasets. Differential privacy has garnered significant attention from researchers and privacy experts due to its strong privacy guarantees. However, the accuracy loss caused by the noise added has been an issue. First, we propose a new noise adding mechanism that preserves (ϵ, δ)-differential privacy. The distribution pertaining to this mechanism can be observed as a generalized truncated Laplacian distribution. We show that the proposed mechanism adds optimal noise in a global context, conditional upon technical lemmas. We also show that the generalized truncated Laplacian mechanism performs better than the optimal Gaussian mechanism. In addition, we also propose an (ϵ)-differentially private mechanism to improve the utility of differential privacy by fusing multiple Laplace distributions. We also derive the closed-form expressions for absolute expectation and variance of noise for the proposed mechanisms. Finally, we empirically evaluate the performance of the proposed mechanisms and show an increase in all utility measures considered, while preserving privacy.

Keywords: Differential privacy · Generalized truncated Laplacian distribution · Merging of Laplacian distributions

1 Introduction

In the modern era, there has been a rapid increase in the amount of digital information collected by governments, social media, hospitals, etc. Though this data holds great utility for business and research purposes, inappropriate use of data can lead to a myriad of issues pertaining to privacy. For example, Target inferred that a teen girl was pregnant before her family knew and started sending her coupons related to baby products [8]. Few years ago, Uber's poor privacy practices caught news' attention: their employees misused customer data to track their customers, including politicians and celebrities, in real time, and blogged about "Rides of Glory", where Uber was able to track one night stands [13]. [11]

J. Domingo-Ferrer and K. Muralidhar (Eds.): PSD 2020, LNCS 12276, pp. 74–90, 2020.
https://doi.org/10.1007/978-3-030-57521-2_6

used the Internet Movie Database as a source of prior knowledge to re-identify the anonymized Netflix records of users, unveiling their alleged political preferences and other sensitive information. Due to these and other similar incidents, governments and policymakers start to recognize the importance of protecting personal data. The European Union (EU) recently proposed the General Data Protection Regulation (GDPR) to protect all EU citizens from privacy breaches in today's data-driven world [17] and other countries are contemplating similar regulation meanwhile. Unfortunately, the gaps in current privacy preserving techniques make it difficult for data collectors to support this kind of privacy regulation. However, differential privacy helps organizations to comply with these regulations. The key idea of differential privacy is to reduce the privacy impact on individuals whose information is available in the dataset. Hence, it is not possible to identify individual records and sensitive information pertaining to a particular user.

Many possible approaches can be taken to preserve the privacy of datasets. Early techniques included simple mechanisms for anonymizing datasets by redacting or removing certain fields from datasets and operating on them normally. However, it quickly became apparent that an adversary with auxiliary information could learn significant information from these anonymized datasets. This led to the development of k-anonymity, which generalizes quasi-identifiers (pieces of data that by themselves are not unique identifiers but can be combined with others to act like one) and ensures that a particular user's data is indistinguishable from that of at least $(k-1)$ other users [16]. Though k-anonymity can protect against identity disclosure, it is susceptible against homogeneity and background-knowledge based attacks [14]. l-diversity overcomes this problem and protects against inference-based attacks [10]. However, the semantic relationship between the sensitive attributes makes l-diversity prone to skewness and similarity-based attacks as it is inadequate to avoid attribute exposure [14]. Differential privacy, which provides strong theoretical privacy guarantees, was proposed to provide statistical indistinguishability of datasets.

Differentially private mechanisms are used to release statistics of a dataset as a whole while protecting the sensitive information of individuals in the dataset. Basically, differential privacy guarantees that the released results reveal little or no new information about an individual in the dataset. As an individual sample cannot affect the output significantly, the attackers thus cannot infer the private information corresponding to a particular individual.

Though there has been a myriad of significant contributions in the field of differential privacy, the reasons that it has not yet been adopted by many in the industry are: first, lack of flexibility in the existing mechanisms due to dearth of configurable parameters, second, concerns over reduced utility and privacy. In this paper, we address these issues and offer solutions. Our contributions are as follows:

1. First, we propose the generalized truncated Laplacian mechanism. We also derive the upper bounds on noise amplitude and noise power for the proposed mechanism. We also show that the generalized truncated Laplacian

mechanism offers better flexibility than existing (ϵ, δ)-differentially private mechanisms and performs better than the optimal Gaussian mechanism by reducing the noise amplitude and noise power in all valid privacy regimes [1].
2. Second, we propose how different Laplacian distributions can be merged based on different breakpoints and we also prove that the resulting distribution is differentially private. We also show how it can enhance the utility while guaranteeing privacy.

The proposed mechanisms enable data controllers to fine-tune the perturbation that is necessary to protect privacy for use case specific distortion requirements. This also mitigates the problems pertaining to inaccuracy and provides better utility in bounding noise.

The paper is organized as follows. Section 2 compares and contrasts our work with related work in the field of differential privacy. Section 3 provides background on differential privacy. In Sect. 4 and Sect. 5, we present the generalized truncated Laplacian mechanism and the merging of Laplacian distribution mechanism, respectively. In Sect. 6 we conclude with a summary and a discussion of our future work.

2 Related Work

For numeric queries, ϵ-differential privacy [3] is achieved by adding Laplacian noise to the query result. It has been the de facto approach in a number of works pertaining to differential privacy [4,9] and [5]. [2] proposed (ϵ, δ)-differential privacy, which can be interpreted as ϵ-differential privacy "with probability $1-\delta$". In spite of its near-ubiquitous use, the Laplacian mechanism has no single substantiation of its optimality. [6] proposes a truncated Laplacian mechanism which draw noises from truncated Laplacian distribution. They have shown that the mechanism is more optimal than the optimal Gaussian mechanism as it significantly reduces the noise amplitude and noise power in a myriad of privacy regimes. [6] offers approximate differential privacy and is defined for the symmetric truncated region, that is, $[-A, A]$. [15] propose piecewise mixture distributions that preserve differential privacy and elucidate the importance of flexibility. Most mechanisms and algorithms in differential privacy uses probability distribution with density functions where ϵ is the only variable, a predefined and fixed sensitivity, and minimal amount of additional flexibility for the query-mechanism designer. In this paper, we propose other mechanisms that offer greater flexibility and provide better privacy guarantees than the existing mechanisms. In order to make use of the perturbed query outputs, we have to understand the trade-off between accuracy and privacy. ϵ plays a significant role in determining this trade-off. It is inversely proportional to the scale parameter in the Laplacian distribution. If the value of ϵ is close to zero, the response to two queries made on neighboring datasets is virtually indistinguishable. However, this makes the queries useless as a large amount of noise would have been added to the result and make it futile. In prior literature pertaining to the accuracy and privacy of differentially private mechanisms, the metric of accuracy is in terms of the

amount of noise added to the output of a query or in terms of variance. [7] studied the trade-off between privacy and error for answering a group of linear queries in a differentially private manner, where the error is defined as the lowest expectation of the ℓ^2-norm of the noise among the query outputs. They also derived the boundary conditions on the error given the differential privacy constraint. [12] were able to extend the result on the trade-off between privacy and error to the case of (ϵ, δ)-differential privacy.

3 Background

In this section, we will provide an overview of differential privacy, describe the privacy-accuracy trade-off under (ϵ)-differential privacy and (ϵ, δ)-differential privacy, and provide the cost functions that are commonly used in evaluating the utility and privacy trade-off of mechanisms that satisfy differential privacy.

3.1 Differential Privacy

Consider a query function,

$$q : \mathcal{D} \to \mathbb{R},$$

where $\mathcal{D}$ denotes the set of all possible datasets. The query function q is applied to a dataset or subsets of datasets and returns a real number. Any two datasets $\mathcal{D}_1 \in \mathcal{D}$ and $\mathcal{D}_2 \in \mathcal{D}$ are called *neighboring datasets* if they differ by at most one element. In other words, one dataset is a subset of the other and $|\mathcal{D}_1 - \mathcal{D}_2| \leq 1$. We denote two neighboring datasets $\mathcal{D}_1$, $\mathcal{D}_2$ as $\mathcal{D}_1 \sim \mathcal{D}_2$. A randomized query-answering mechanism $\mathcal{A}$ is a function of the query function q, and will randomly output a real number with certain probability distribution $\mathcal{P}$ depending on $q(\mathcal{D})$, where $\mathcal{D}$ is the dataset.

A more relaxed notion of ϵ-differential privacy is (ϵ, δ)-differential privacy, which can be interpreted as the algorithm that is mostly ϵ-differentially private with the factor δ denoting the probability that it fails to be. Formally, we have the following definition.

Definition 1 *((ϵ, δ)-Differential Privacy). A randomized mechanism $\mathcal{A} : \mathcal{D} \to \mathcal{O}$ preserves (ϵ, δ)-differential privacy ((ϵ, δ)-DP) when there exists $\epsilon > 0$, $\delta > 0$ such that,*

$$Pr\,[\mathcal{A}(\mathcal{D}_1) \in \mathcal{T}] \leq e^{\epsilon} Pr\,[\mathcal{A}(\mathcal{D}_2) \in \mathcal{T}] + \delta$$

holds for every subset $\mathcal{T} \subseteq \mathcal{O}$ and for any two neighboring datasets $\mathcal{D}_1 \sim \mathcal{D}_2$.

Definition 2 *(Global Sensitivity). For a real-valued query function $q : \mathcal{D} \to \mathbb{R}$, where $\mathcal{D}$ denotes the set of all possible datasets, the global sensitivity of q, denoted by Δ, is defined as*

$$\Delta = \max_{\mathcal{D}_1 \sim \mathcal{D}_2} |q(\mathcal{D}_1) - q(\mathcal{D}_2)|,$$

for all $\mathcal{D}_1 \in \mathcal{D}$ and $\mathcal{D}_2 \in \mathcal{D}$.

Note when the query function q is a counting query or a histogram query, the global sensitivity $\Delta = 1$ because removing one user from the dataset $\mathcal{D}$ only affects the output of the query by at most 1.

3.2 Utility Model

In this section, we discuss the way that we will be using to evaluate the utility and privacy of a differentially private mechanism. Consider a cost function $\mathcal{L} : \mathbb{R} \rightarrow \mathbb{R}$, which is a function of the random additive noise in the mechanism $\mathcal{A}$. Given a random additive noise x, the cost function for it is defined as $\mathcal{L}(x)$. Therefore, we can derive the expectation of the cost over the probability distribution $\mathcal{P}$ by solving:

$$\int_{x \in \mathbb{R}} \mathcal{L}(x)\mathcal{P}(x)dx$$

Upper bounds on the minimum noise amplitude and noise power, correspond to the l^1 cost function $\mathcal{L}(x) = |x|$ and l^2 cost function $\mathcal{L}(x) = x^2$, respectively. Our objective is to minimize such expectation of the cost over the probability distribution for preserving differential privacy.

3.3 Differentially Private Mechanisms

For the case of real output, introducing noise in an additive manner is a standard technique to preserve differential privacy. Thus, we will be discussing mechanisms $\mathcal{A}$ that preserves ϵ or (ϵ, δ)- differential privacy by adding a random noise X drawn from a probability distribution $\mathcal{P}$. So we will reserve the notation $\mathcal{A}$ for mechanisms that take the standard formula:

$$\mathcal{A}(\mathcal{D}) = q(\mathcal{D}) + X.$$

We will also reserve the variable X for the additive random noise drawn from the probability distribution $\mathcal{P}$ from now on unless stated otherwise.

One of the most well-known differentially private mechanism is the Laplacian mechanism, which uses random noise X drawn from the symmetric Laplacian distribution. The zero-mean Laplacian distribution has a symmetric probability density function $f(x)$ with a scale parameter λ defined as:

$$f(x) = \frac{1}{2\lambda} e^{-\frac{|x|}{\lambda}}.$$

Given the global sensitivity, Δ, of the query function q, and the privacy parameter ϵ, the *Laplacian mechanism* $\mathcal{A}$ uses random noise X drawn from the Laplacian distribution with scale $\lambda = \frac{\Delta}{\epsilon}$. The Laplacian mechanism preserves ϵ-differential privacy [2].

A variant of the symmetric Laplacian mechanism is the truncated Laplacian mechanism, which uses a random noise generated from the truncated Laplace distribution. The zero-mean truncated Laplace distribution has a symmetric-bounded probability density function $f(x)$ with scale λ defined as:

$$f(x) = \begin{cases} Be^{-\frac{|x|}{\lambda}}, & \text{for } x \in [-A, A] \\ 0, & \text{otherwise} \end{cases}$$

where

$$A = \frac{\Delta}{\epsilon} \ln(1 + \frac{e^{\epsilon} - 1}{2\delta}) \text{ and } B = \frac{1}{2\frac{\Delta}{\epsilon}(1 - \frac{1}{1+\frac{e^{\epsilon}-1}{2\delta}})}.$$

Given the global sensitivity Δ of the query function q, and the privacy parameters ϵ, δ, the *truncated Laplacian mechanism* $\mathcal{A}$ uses random noise X drawn from the truncated Laplacian distribution with scale $\lambda = \frac{\Delta}{\epsilon}$. It has been proven to be (ϵ, δ)-differentially private for $\delta < \frac{1}{2}$ [6].

Remark 1. Note that an ϵ or (ϵ, δ)-differential private mechanism $\mathcal{A}$ with additive noise X drawn from probability distribution $\mathcal{P}$ will still be ϵ or (ϵ, δ)-differential private when the mean μ of $\mathcal{P}$ is any finite real number instead of 0. Therefore, we will just be discussing and proving the $\mu = 0$ case in this paper. However, the proof for any real number μ is similar.

4 Generalized Truncated Laplacian Mechanism

In this section, we propose an (ϵ, δ)-differentially private mechanism that offers better flexibility than the symmetrically bounded truncated Laplacian mechanism [6] and better accuracy than the optimal Gaussian mechanism [1]. First, we state the probability density function and the cumulative distribution function of the generalized truncated Laplacian distribution. Then, we elucidate the (ϵ, δ)-differentially private mechanism. Finally, we evaluate the upper bound on noise amplitude and noise power.

4.1 Generalized Truncated Laplace Distribution

The probability distribution can be viewed as a generalized truncated Laplace distribution. Such a probability distribution is motivated by the symmetrically bounded Laplace distribution proposed by [6]. The proposed distribution in this paper is a more general version as it is asymmetrically bounded.

To construct such a distribution, we set the privacy parameter ϵ and δ. In contrast to most of the existing (ϵ, δ)-differential private mechanisms, where ϵ and δ are the only two variables in the algorithm design, the generalized truncated Laplacian distribution allows another parameter to specify the upper or lower bound of the probability density function. Therefore, with the additional bounding parameter, not depending on the value of ϵ or δ, the proposed generalized truncated Laplace distribution provides more flexibility.

Definition 3. *The zero-mean generalized truncated Laplace distribution has a probability density function $f(x)$ with scale λ, and is asymmetrically bounded by A and B where $A < 0 < B$, defined as:*

$$f(x) = \begin{cases} Me^{-\frac{|x|}{\lambda}} \text{ for } x \in [A, B] \\ 0 \text{ otherwise} \end{cases} \quad \text{where, } M = \frac{1}{\lambda(2 - e^{\frac{A}{\lambda}} - e^{-\frac{B}{\lambda}})}$$

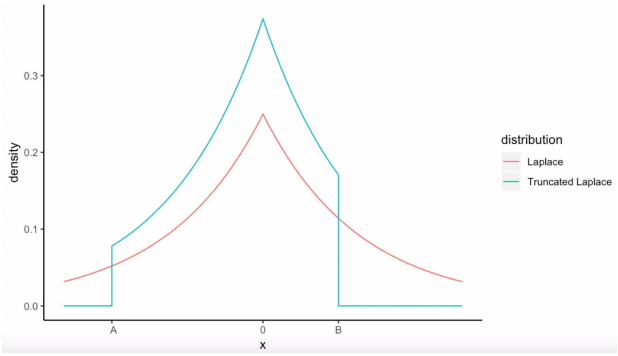

Fig. 1. Laplacian mechanism vs Generalized truncated Laplacian mechanism

Figure 1 depicts a zero-mean Laplace distribution and generalized truncated Laplacian distribution with a scale factor of 2.

The proposed probability distribution is valid, as its probability density function $f(x)$ is greater than 0 for x in the sample space and $\int_{\infty}^{\infty} f(x)dx = 1$

Then we present the closed form of the cumulative distribution function, $F(x)$, for the generalized truncated Laplacian distribution.

The cumulative distribution function is defined as,

$$F(x) = \begin{cases} \frac{e^{\frac{x}{\lambda}} - e^{\frac{A}{\lambda}}}{2 - e^{\frac{A}{\lambda}} - e^{-\frac{B}{\lambda}}} & \text{if } x < 0 \\ \frac{2 - e^{\frac{A}{\lambda}} - e^{-\frac{x}{\lambda}}}{2 - e^{\frac{A}{\lambda}} - e^{-\frac{B}{\lambda}}} & \text{if } x \geq 0 \end{cases}$$

4.2 Mechanism

Given the global sensitivity Δ of the query function q, and the privacy parameters ϵ, δ, the *Generalized Truncated Laplacian mechanism* $\mathcal{A}$ uses random noise X drawn from the generalized truncated Laplacian distribution in Definition 3 with the following parameters:

$$\lambda = \frac{\Delta}{\epsilon}, A + \Delta \leq 0 \leq B - \Delta$$

If $|A| \geq |B|$,

$$\begin{cases} A = \lambda \ln \left[2 + (\frac{1-\delta}{\delta}) e^{-\frac{B}{\lambda}} - (\frac{1}{\delta}) e^{-\frac{B-\Delta}{\lambda}} \right] \\ B = \text{any positive real number satisfying} |A| \geq |B| \end{cases}$$

If $|A| < |B|$,

$$\begin{cases} A = \text{any negative real number satisfying } |A| < |B| \\ B = -\lambda \ln \left[2 + (\frac{1-\delta}{\delta}) e^{\frac{A}{\lambda}} - (\frac{1}{\delta}) e^{\frac{A+\Delta}{\lambda}} \right] \end{cases}$$

Theorem 1. *The generalized truncated Laplacian mechanism preserves* (ϵ, δ)*-differential privacy.*

Proof. The proof for Theorem 1 relies on the following two lemmas, and the proof for those lemmas can be found in Appendix A.

Lemma 1.

$$\max\Big(\int_{A}^{A+\Delta} f(x)dx, \int_{B-\Delta}^{B} f(x)dx\Big) = \delta$$

for the probability density function $f(x)$*,* λ, A *and* B *of the generalized truncated Laplace distribution.*

Lemma 2. *A mechanism* $\mathcal{A}(\mathcal{D}) = q(\mathcal{D}) + X$ *that adds a random noise* X *drawn from probability distribution* $\mathcal{P}$ *with probability density function* $f(x)$*, satisfies* (ϵ, δ)*-differential privacy when*

$$\mathcal{P}(\mathcal{S}) - e^{\epsilon}\mathcal{P}(\mathcal{S} + d) \leq \delta$$

holds for any $|d| \leq \Delta$*, and any measurable set* $\mathcal{S} \subseteq \mathbb{R}$*, where* Δ *is the global sensitivity for the query function* q*.*

Using Lemma 2, in order to prove that our mechanism is (ϵ, δ)-differential private, we need to show that for the global sensitivity, Δ, of our query function q,

$$\mathcal{P}(\mathcal{S}) - e^{\epsilon}\mathcal{P}(\mathcal{S} + d) \leq \delta$$

holds for any $|d| \leq \Delta$, and any measurable set $\mathcal{S} \subseteq \mathbb{R}$. Equivalently, it is sufficient to show that

$$\max(\mathcal{P}(\mathcal{S}) - e^{\epsilon}\mathcal{P}(\mathcal{S} + d)) \leq \delta$$

If $x \in (A + \Delta, B - \Delta)$ then,

$$\frac{f(x)}{f(x+d)} = \frac{Me^{-\frac{|x|}{\lambda}}}{Me^{-\frac{|x+d|}{\lambda}}} = e^{\frac{|x+d|-|x|}{\lambda}} \leq e^{\frac{|d|}{\lambda}} \leq e^{\epsilon},$$

which implies $\forall |d| \leq \Delta, f(x) - e^{\epsilon} f(x+d) \leq 0$ when $x \in (A + \Delta, B - \Delta)$.

Thus, for measurable set $\mathcal{S}$,

$$\mathcal{P}(\mathcal{S}) - e^{\epsilon}\mathcal{P}(\mathcal{S} + d) \leq \mathcal{P}(\mathcal{S}') - e^{\epsilon}\mathcal{P}(\mathcal{S}' + d)$$

for $\mathcal{S} \subseteq \mathbb{R}$ and $\mathcal{S}' = \mathcal{S} \backslash (A + \Delta, B - \Delta)$. Therefore, $\mathcal{P}(\mathcal{S}) - e^{\epsilon}\mathcal{P}(\mathcal{S} + d)$ is maximized for some set $\mathcal{S} \subseteq (-\infty, A + \Delta]$ or $\mathcal{S} \subseteq [B - \Delta, \infty)$. Since the distribution changes exponentially with rate $\frac{1}{\lambda} = \frac{\epsilon}{\Delta}$, multiplying the probability distribution $\mathcal{P}$ by e^{ϵ} will result in shifting the probability distribution by Δ. Therefore,

$$\sup_{\mathcal{S} \subseteq \mathbb{R}} \mathcal{P}(\mathcal{S}) - e^{\epsilon}\mathcal{P}(\mathcal{S} + d) \leq \max\Big(\int_{-\infty}^{A+\Delta} f(x)dx, \int_{B-\Delta}^{\infty} f(x)dx\Big) = \max\Big(\int_{A}^{A+\Delta} f(x)dx, \int_{B-\Delta}^{B} f(x)dx\Big)$$

From Lemma 1, we have the desired inequality

$$\mathcal{P}(S) - e^{\epsilon}\mathcal{P}(S + d) \leq \delta.$$

Remark 2. We claim that

$$0 < \delta \leq \min\left(\int_A^0 f(x)dx, \int_0^B f(x)dx\right).$$

Proof. If $|A| \geq |B|$, then

$$\int_A^0 f(x)dx > \int_0^B f(x)dx$$
$$\Rightarrow \min\left(\int_A^0 f(x)dx, \int_0^B f(x)dx\right) = \int_0^B f(x)dx$$

Additionally,

$$\max\left(\int_A^{A+\Delta} f(x)dx, \int_{B-\Delta}^B f(x)dx\right) = \int_{B-\Delta}^B f(x)dx = \delta$$

Since $0 \leq B - \Delta$,

$$\delta = \int_{B-\Delta}^B f(x)dx \leq \int_0^B f(x)dx.$$

If $|A| < |B|$, the proof is similar.

4.3 Upper Bound on Noise Amplitude and Noise Power

We apply the generalized truncated Laplacian mechanism to derive upper bounds on the minimum noise amplitude and noise power, corresponding to the l^1 cost function $\mathcal{L}(x) = |x|$ and l^2 cost function $\mathcal{L}(x) = x^2$, respectively.

When $\mathcal{L}(x) = |x|$, the upper bound on minimum noise amplitude is

$$\frac{2\lambda - (\lambda - A)e^{\frac{A}{\lambda}} - (\lambda + B)e^{-\frac{B}{\lambda}}}{2 - e^{\frac{A}{\lambda}} - e^{-\frac{B}{\lambda}}}, \text{where } \lambda = \frac{\Delta}{\epsilon}, \text{ A and B are specified in Theorem 1.}$$

This result is obtained by evaluating

$$inf_{\mathcal{P}\in\mathcal{P}_{\epsilon,\delta}} \int_{x\in\mathbb{R}} |x|\mathcal{P}(dx) = \int_A^B |x|f(x)dx = M\left(\int_A^0 -xe^{\frac{x}{\lambda}}dx + \int_0^B xe^{-\frac{x}{\lambda}}dx\right)$$

As the noise with probability density function $f(x)$ satisfies (ϵ, δ)-differential privacy, this provides an upper bound on $\inf_{P\in P_{\epsilon,\delta}} \int_{x\in\mathbb{R}} |x|\mathcal{P}(dx)$.

Similarly, we derive the upper bound on the minimum noise power by having $\mathcal{L}(x) = x^2$, and we get

$$\frac{4\lambda^2 - (2\lambda^2 - 2\lambda A + A^2)e^{\frac{A}{\lambda}} - (2\lambda^2 + 2\lambda B + B^2)e^{-\frac{B}{\lambda}}}{2 - e^{\frac{A}{\lambda}} - e^{-\frac{B}{\lambda}}}$$

where $\lambda = \frac{\Delta}{\epsilon}$, and A and B are specified in Sect. 4.2.
As the noise with probability density function $f(x)$ satisfies (ϵ, δ)-differential privacy, this provides an upper bound on $\inf_{P \in P_{\epsilon,\delta}} \int_{x \in \mathbb{R}} x^2 \mathcal{P}(dx)$.

We performed experiments to compare the performance of the generalized truncated Laplacian mechanism with the optimal Gaussian mechanism [1]. [1] calculate the variance of the optimal Gaussian mechanism using the cumulative density function instead of a tail bound approximation. The ratio of the noise amplitude and noise power of generalized truncated Laplacian mechanism and the optimal Gaussian mechanism is always less than 1 for appropriate values of δ, A and B as shown in Appendix B. Compared to the optimal Gaussian mechanism, the generalized truncated Laplacian mechanism reduces the noise power and noise amplitude across all privacy regimes.

5 Merged Laplacian Mechanism

In this section, we propose an ϵ-differentially private mechanism that merges different Laplacian distributions on different breakpoints. We also evaluate the l^1 cost function $\mathcal{L}(x) = |x|$ and l^2 cost function $\mathcal{L}(x) = x^2$ for the proposed mechanism, compare it with the Laplacian mechanism and show that our proposed mechanism achieves better utility.

Definition 4. *The zero-mean merged Laplacian distribution has a probability density function $f(x)$ with n break points $0 < c_1 < c_2 < \cdots < c_n = \infty$ and n scale parameters $\lambda_1, \lambda_2, ..., \lambda_n$ defined as:*

$$f(x) = f_m(x) \text{ for } x \in (-c_m, -c_{m-1}] \cup [c_{m-1}, c_m)$$
$$\text{Let } c_0 = 0, \forall m \in \{1, 2, ..., n\} \text{ where } f_m(x) = a_m e^{-\frac{|x|}{\lambda_m}},$$

and all $a_m > 0$ computed by

$$\sum_{m=1}^{n} \int_{c_{m-1}}^{c_m} a_m e^{-\frac{|x|}{\lambda_m}} dx = \frac{1}{2}, \tag{1}$$

$$\text{and } f_m(c_m) = f_{m+1}(c_m). \tag{2}$$

Remark 3. Note that (1) and (2) gives sufficient inputs to calculate a_m for $m \in \{1, 2, .., n\}$ as we can write a_m's in terms of $\lambda_1, \lambda_2, ..., \lambda_n$ and a_1 and by inductively applying

$$f_m(c_m) = f_{m+1}(c_m) \implies a_m e^{-\frac{c_m}{\lambda_m}} = a_{m+1} e^{-\frac{c_m}{\lambda_{m+1}}} \implies a_{m+1} = a_m e^{\frac{c_m}{\lambda_{m+1}} - \frac{c_m}{\lambda_m}}.$$

Then, we can rewrite (1) with a_1 as the only variable to solve for the value of a_1. Hence, we can get the values for the rest of the a_m's.

Now we will prove that the probability distributions that we proposed is a valid one. To do this, we need to show that its probability density function $f(x)$ is continuous and greater than 0 for x in the domain and the cumulative probability from $-\infty$ to ∞ is 1.

Proof. First, it is easy to see that $f(x) > 0$ for x in the domain as $e^{-\frac{|x|}{\lambda_m}} > 0$ for $m \in \{1, 2, ..., n\}$, and all $a_m > 0$. Thus, $f_m(x) > 0 \Rightarrow f(x) > 0$.

Additionally, $f(x)$ is continuous on $\cup_{m=0}^{n}(c_{m-1}, c_m)$ as $e^{-\frac{|x|}{\lambda_m}}$ is continuous. At each break point c_m, the continuity is ensured by $f_m(c_m) = f_{m+1}(c_m)$. Now we will show that the cumulative probability from $-\infty$ to ∞ is 1.

$$\int_{-\infty}^{\infty} f(x)dx = \sum_{m=1}^{n}\left(\int_{-c_m}^{-c_{m-1}} f_m(x)dx + \int_{c_{m-1}}^{c_m} f_m(x)dx\right) = 2\sum_{m=1}^{n}\int_{c_{m-1}}^{c_m} f_m(x)dx = 2 \cdot \frac{1}{2} = 1$$

Now we propose a differentially private mechanism which adds noise drawn from the Merged Laplacian distribution as defined in Definition 4.

Theorem 2. *Given the global sensitivity, Δ, of the query function q, and the privacy parameter $\epsilon = \epsilon_1$, the Merged Laplacian mechanism $\mathcal{A}$ uses random noise X drawn from the merged Laplacian distribution with scale parameter $\lambda_m = \frac{\epsilon_m}{\Delta}$ where $\lambda_1 > \lambda_2 > \cdots > \lambda_n$ and preserves ϵ - differential privacy.*

Proof. To prove that our mechanism preserves ϵ - differential privacy, we need to show that for $\mathcal{D}_1 \sim \mathcal{D}_2$,

$$Pr[\mathcal{A}(\mathcal{D}_1) \in \mathcal{T}] \leq e^{\epsilon} Pr[\mathcal{A}(\mathcal{D}_2) \in \mathcal{T}]$$

for any subset $\mathcal{T} \subseteq \mathcal{O}$, where $\mathcal{O}$ is the set of all outputs of the mechanism. And the above inequality is equivalent to

$$\frac{Pr[\mathcal{A}(\mathcal{D}_1) = t]}{Pr[\mathcal{A}(\mathcal{D}_2) = t]} \leq e^{k\epsilon}, \forall t \in \mathcal{T} \Leftrightarrow \qquad \frac{Pr[X = t - q(\mathcal{D}_1)]}{Pr[X = t - q(\mathcal{D}_2)]} \leq e^{k\epsilon}.$$

We will prove this inductively. Our base case is when $n = 1$, then the mechanism becomes the well-known Laplacian mechanism, which is ϵ - differentially private as $\epsilon = \max(\epsilon_1)$. Now, notice that since $\lambda_m = \frac{\epsilon_m}{\Delta}$ and $\lambda_1 > \lambda_2 > \cdots > \lambda_n$, then $\max(\epsilon_1, \epsilon_2, ..., \epsilon_n) = \epsilon_1 = \epsilon$.

Now, assume with the same break points $c_1, c_2, ..., c_{k-1}$ where $0 < c_1 < c_2 < \cdots < c_k = \infty$, the merged Laplacian mechanism is $\epsilon = \epsilon_1$ - differentially private. We want to prove that adding one more break point $c_k < \infty$ to the new merged mechanism satisfies ϵ - differential privacy. We will prove the case where $t - q(\mathcal{D}_1)$ and $t - q(\mathcal{D}_2)$ are negative, as the other cases follows the similar proof with a few sign changes. For $m \in \{1, 2, ..., k-1\}$, we have

$$\frac{Pr[X = t - q(\mathcal{D}_1)]}{Pr[X = t - q(\mathcal{D}_2)]} = \frac{a_m e^{\frac{t - q(\mathcal{D}_1)}{\lambda_m}}}{a_k e^{\frac{t - q(\mathcal{D}_2)}{\lambda_k}}} = \frac{a_m}{a_k} \cdot e^{\frac{t - q(\mathcal{D}_1)}{\lambda_m} - \frac{t - q(\mathcal{D}_2)}{\lambda_k}}$$

We also know that,

$$a_k = a_{k-1} e^{\frac{c_{k-1}}{\lambda_k} - \frac{c_{k-1}}{\lambda_{k-1}}} = a_{k-2} e^{\frac{c_{k-2}}{\lambda_{k-1}} - \frac{c_{k-2}}{\lambda_{k-2}} + \frac{c_{k-1}}{\lambda_k} - \frac{c_{k-1}}{\lambda_{k-1}}} = a_m e^{\sum_{i=m-1}^{k-1} \left(\frac{c_i}{\lambda_{i+1}} - \frac{c_i}{\lambda_i} \right)}$$

Hence,

$$\frac{a_m}{a_k} = e^{\sum_{i=m-1}^{k-1} \left(\frac{c_i}{\lambda_i} - \frac{c_i}{\lambda_{i+1}} \right)}.$$

Notice that

$$\sum_{i=m-1}^{k-1} \left(\frac{c_i}{\lambda_i} - \frac{c_i}{\lambda_{i+1}} \right) = \sum_{i=m-1}^{k-1} \left(\frac{c_i \left(\lambda_{i+1} - \lambda_i \right)}{\lambda_i \lambda_{i+1}} \right) < 0 \text{ since } \lambda_1 > \lambda_2 > \cdots > \lambda_n.$$

Thus,

$$\frac{Pr[X = t - q(\mathcal{D}_1)]}{Pr[X = t - q(\mathcal{D}_2)]} = \frac{a_m}{a_k} \cdot e^{\frac{t - q(\mathcal{D}_1)}{\lambda_m} - \frac{t - q(\mathcal{D}_2)}{\lambda_k}} < e^{\frac{t - q(\mathcal{D}_1)}{\lambda_m} - \frac{t - q(\mathcal{D}_2)}{\lambda_k}} = e^{\frac{\lambda_k (t - q(\mathcal{D}_1)) - \lambda_m (t - q(\mathcal{D}_2))}{\lambda_m \lambda_k}}$$
$$< e^{\frac{\lambda_m (t - q(\mathcal{D}_1)) - \lambda_m (t - q(\mathcal{D}_2))}{\lambda_m \lambda_k}} = e^{\frac{t - q(\mathcal{D}_1) - t + q(\mathcal{D}_2)}{\lambda_k}} = e^{\frac{q(\mathcal{D}_2) - q(\mathcal{D}_1)}{\lambda_k}} = e^{\epsilon_k} < e^{\epsilon}.$$

Hence, we have proved that the proposed mechanism is ϵ - deferentially private.

We evaluate the l^1 cost function $\mathcal{L}(x) = |x|$ and l^2 cost function $\mathcal{L}(x) = x^2$, for the Laplacian, Merged Laplacian with 1 break point and Merged Laplacian with 2 break points as shown in Appendix C. We show that the cost for the Merged Laplacian with 2 break points is lower than that of the Laplacian mechanism and hence we achieve better utility for the same privacy loss.

6 Conclusion and Future Work

In this paper, we presented two novel differentially private mechanisms that provide better accuracy guarantees compared to existing mechanisms. Firstly, we presented a new noise adding mechanism that preserves (ϵ, δ)-differential privacy. The proposed mechanisms provide more scope for customization as they have more parameters to tune. Due to this customizable and flexible nature, appropriate values for different parameters in the mechanisms can be set. We also show that the generalized truncated Laplacian mechanism performs better than the optimal Gaussian mechanism. Next, we show that the proposed merging of Laplacian mechanisms demonstrates better in performance in terms of various metrics for l^1 and l^2 loss without sacrificing additional privacy. As a part of future work, we plan to perform an in-depth comparison of all (ϵ, δ)-differentially private and ϵ-differentially private mechanisms and highlight the pros and cons of every mechanism.

A Proof for Lemma 1 and Lemma 2

Here, we present the proof for Lemma 1 and Lemma 2 used in Sect. 4.2.

Proof (Proof for Lemma 1). Since the probability density function $f(x)$ is monotonically increasing when $x \geq 0$ and is monotonically decreasing when $x < 0$,

$$\max\Big(\int_{A}^{A+\Delta} f(x)dx, \int_{B-\Delta}^{B} f(x)dx\Big) = \begin{cases} \int_{B-\Delta}^{B} f(x)dx \text{ when } |A| \geq |B| \\ \int_{A}^{A+\Delta} f(x)dx \text{ when } |A| < |B|. \end{cases}$$

We will first discuss the case when $|A| \geq |B|$,

$$\max\Big(\int_{A}^{A+\Delta} f(x)dx, \int_{B-\Delta}^{B} f(x)dx\Big) = \int_{B-\Delta}^{B} f(x)dx = \int_{B-\Delta}^{B} Me^{-\frac{x}{\lambda}}dx$$
$$= M\lambda\Big(e^{-\frac{B-\Delta}{\lambda}} - e^{-\frac{B}{\lambda}}\Big)\Big)$$

Plugging in $M = \frac{1}{\lambda(2-e^{\frac{A}{\lambda}}-e^{-\frac{B}{\lambda}})}$ as in our definition for the generalized truncated Laplacian distribution, $A = \lambda \ln\left[2 + (\frac{1-\delta}{\delta})e^{-\frac{B}{\lambda}} - (\frac{1}{\delta})e^{-\frac{B-\Delta}{\lambda}}\right]$ as specified in Theorem 1, we have

$$\int_{B-\Delta}^{B} f(x)dx = \frac{e^{-\frac{B-\Delta}{\lambda}} - e^{-\frac{B}{\lambda}}}{2 - e^{\frac{A}{\lambda}} - e^{-\frac{B}{\lambda}}} = \frac{e^{-\frac{B-\Delta}{\lambda}} - e^{-\frac{B}{\lambda}}}{(\frac{1}{\delta})\left(e^{-\frac{B-\Delta}{\lambda}} - e^{-\frac{B}{\lambda}}\right)} = \delta$$

We omit showing the computation for the case when $|A| \leq |B|$ as the derivation is very similar to that of the above mentioned case.

Now, we will proceed to prove Lemma 2.

Proof (Proof for Lemma 2). Given two neighboring datasets $\mathcal{D}_1 \sim \mathcal{D}_2$, we know that $|q(\mathcal{D}_1) - q(\mathcal{D}_2)| \leq \Delta$, thus the condition $\mathcal{P}(\mathcal{S}) - e^{\epsilon}\mathcal{P}(\mathcal{S} + d) \leq \delta$ for any $|d| \leq \Delta$ is equivalent to

$$\mathcal{P}(\mathcal{S}) - e^{\epsilon}\mathcal{P}(\mathcal{S} + q(\mathcal{D}_1) - q(\mathcal{D}_2)) \leq \delta \Leftrightarrow \mathcal{P}(\mathcal{S} - q(\mathcal{D}_1)) \leq e^{\epsilon}\mathcal{P}(\mathcal{S} - q(\mathcal{D}_2)) + \delta$$

Hence, for any $t \in \mathcal{S}$, the condition is equivalent to

$$\Pr(X = t - q(\mathcal{D}_1)) \leq e^{\epsilon}\Pr(X = t - q(\mathcal{D}_2)) + \delta$$
$$\Leftrightarrow \Pr(q(\mathcal{D}_1) + X = t) \leq e^{\epsilon}\Pr(q(\mathcal{D}_2) + X = t) + \delta$$
$$\Leftrightarrow \Pr\left[\mathcal{A}(\mathcal{D}_1) \in \mathcal{T}\right] \leq e^{\epsilon}\Pr\left[\mathcal{A}(\mathcal{D}_2) \in \mathcal{T}\right] + \delta,$$

which is the necessary condition for mechanism $\mathcal{A}$ to preserve (ϵ, δ)-differential privacy.

B Generalized Truncated Laplacian - Evaluation

We empirically show that the ratio of the noise amplitude L_1^* and noise power L_2^* of generalized truncated Laplacian mechanism and the optimal Gaussian mechanism is always less than 1 for appropriate values of δ, A and B as described in Sect. 4. Compared to the optimal Gaussian mechanism, the generalized truncated Laplacian mechanism reduces the noise power and noise amplitude across all privacy regimes. The implementation can be found in https://github.com/vaikkunth/DPMechanisms.

ϵ	δ	A	L_1^*	L_2^*
0.7	2.5e−06	−17.46	0.25	0.13
0.4	4.0e−06	−27.57	0.27	0.15
0.4	2.5e−06	−28.74	0.27	0.14
0.4	9.5e−06	−25.4	0.29	0.17
0.4	3.0e−06	−28.29	0.27	0.14
0.7	6.0e−06	−16.21	0.27	0.14
0.4	8.5e−06	−25.68	0.29	0.16
0.7	3.5e−06	−16.98	0.26	0.13
0.7	4.0e−06	−16.79	0.26	0.14
0.4	2.0e−06	−29.3	0.26	0.14
0.1	4.5e−06	−93.66	0.31	0.19
0.7	3.0e−06	−17.2	0.26	0.13
0.4	5.5e−06	−26.77	0.28	0.15
0.7	9.5e−06	−15.55	0.28	0.15
0.4	6.0e−06	−26.55	0.28	0.16
0.7	1.0e−06	−18.77	0.24	0.12
0.4	6.5e−06	−26.35	0.28	0.16
0.4	7.0e−06	−26.17	0.28	0.16
0.4	4.5e−06	−27.27	0.27	0.15
0.4	9.0e−06	−25.54	0.29	0.17
0.4	3.5e−06	−27.9	0.27	0.15
0.1	9.5e−06	−86.19	0.32	0.21
0.1	5.5e−06	−91.66	0.31	0.19
0.1	5.0e−06	−92.61	0.31	0.19
0.4	1.5e−06	−30.02	0.26	0.13
0.4	8.0e−06	−25.83	0.29	0.16

ϵ	δ	A	L_1^*	L_2^*
0.7	7.0e−06	−15.99	0.27	0.15
0.7	7.5e−06	−15.89	0.27	0.15
0.4	1.0e−06	−31.03	0.25	0.13
0.1	7.0e−06	−89.24	0.32	0.2
0.4	5.0e−06	−27.01	0.28	0.15

C Merging Laplacian Distributions - Evaluation

We evaluate the l^1 and l^2 cost for the Laplacian, Merged Laplacian with 1 break point and Merged Laplacian with 2 break points. We show that the cost for the Merged Laplacian with 2 break points is lower than that of the Laplacian mechanism and hence we achieve better utility for the same privacy loss. The implementation can be found in https://github.com/vaikkunth/DPMechanisms.

$(\epsilon_1, \epsilon_2, \epsilon_3)$	(c_1, c_2)	L_1^*	L_2^*
$(0.2, 0.25, 0.33)$	$(1, 3)$	$(3.0, 3.03, 1.12)$	$(18.0, 18.22, 3.94)$
$(0.17, 0.25, 0.33)$	$(1, 3)$	$(3.0, 3.03, 1.12)$	$(18.0, 18.22, 3.96)$
$(0.17, 0.2, 0.33)$	$(1, 3)$	$(3.0, 3.05, 1.13)$	$(18.0, 18.35, 4.07)$
$(0.14, 0.25, 0.33)$	$(1, 3)$	$(3.0, 3.03, 1.12)$	$(18.0, 18.22, 3.97)$
$(0.14, 0.2, 0.33)$	$(1, 3)$	$(3.0, 3.05, 1.13)$	$(18.0, 18.35, 4.08)$
$(0.14, 0.17, 0.33)$	$(1, 3)$	$(3.0, 3.06, 1.13)$	$(18.0, 18.43, 4.15)$
$(0.14, 0.17, 0.2)$	$(1, 3)$	$(5.0, 5.01, 1.96)$	$(50.0, 50.15, 16.26)$
$(0.12, 0.25, 0.33)$	$(1, 3)$	$(3.0, 3.03, 1.13)$	$(18.0, 18.22, 3.98)$
$(0.12, 0.2, 0.33)$	$(1, 3)$	$(3.0, 3.05, 1.13)$	$(18.0, 18.35, 4.08)$
$(0.12, 0.17, 0.33)$	$(1, 3)$	$(3.0, 3.06, 1.13)$	$(18.0, 18.43, 4.16)$
$(0.12, 0.17, 0.2)$	$(1, 3)$	$(5.0, 5.01, 1.96)$	$(50.0, 50.15, 16.28)$
$(0.12, 0.14, 0.33)$	$(1, 3)$	$(3.0, 3.07, 1.14)$	$(18.0, 18.49, 4.21)$
$(0.12, 0.14, 0.2)$	$(1, 3)$	$(5.0, 5.02, 1.98)$	$(50.0, 50.26, 16.5)$
$(0.11, 0.25, 0.33)$	$(1, 3)$	$(3.0, 3.03, 1.13)$	$(18.0, 18.22, 3.98)$
$(0.11, 0.2, 0.33)$	$(1, 3)$	$(3.0, 3.05, 1.13)$	$(18.0, 18.35, 4.09)$
$(0.11, 0.17, 0.33)$	$(1, 3)$	$(3.0, 3.06, 1.14)$	$(18.0, 18.43, 4.16)$
$(0.11, 0.17, 0.2)$	$(1, 3)$	$(5.0, 5.01, 1.96)$	$(50.0, 50.15, 16.3)$
$(0.11, 0.14, 0.33)$	$(1, 3)$	$(3.0, 3.07, 1.14)$	$(18.0, 18.49, 4.21)$
$(0.11, 0.14, 0.2)$	$(1, 3)$	$(5.0, 5.02, 1.98)$	$(50.0, 50.26, 16.52)$
$(0.11, 0.12, 0.33)$	$(1, 3)$	$(3.0, 3.08, 1.14)$	$(18.0, 18.53, 4.25)$
$(0.11, 0.12, 0.2)$	$(1, 3)$	$(5.0, 5.03, 1.99)$	$(50.0, 50.34, 16.68)$
$(0.11, 0.12, 0.14)$	$(1, 3)$	$(7.0, 7.01, 3.28)$	$(98.0, 98.12, 42.57)$
$(0.2, 0.25, 0.33)$	$(1, 5)$	$(3.0, 3.03, 1.61)$	$(18.0, 18.22, 5.17)$

$(\epsilon_1, \epsilon_2, \epsilon_3)$	(c_1, c_2)	L_1^*	L_2^*
$(0.17, 0.25, 0.33)$	$(1, 5)$	$(3.0, 3.03, 1.61)$	$(18.0, 18.22, 5.19)$
$(0.17, 0.2, 0.33)$	$(1, 5)$	$(3.0, 3.05, 1.63)$	$(18.0, 18.35, 5.4)$
$(0.14, 0.25, 0.33)$	$(1, 5)$	$(3.0, 3.03, 1.61)$	$(18.0, 18.22, 5.2)$
$(0.14, 0.2, 0.33)$	$(1, 5)$	$(3.0, 3.05, 1.63)$	$(18.0, 18.35, 5.41)$
$(0.14, 0.17, 0.33)$	$(1, 5)$	$(3.0, 3.06, 1.64)$	$(18.0, 18.43, 5.55)$
$(0.14, 0.17, 0.2)$	$(1, 5)$	$(5.0, 5.01, 1.85)$	$(50.0, 50.15, 10.69)$
$(0.12, 0.25, 0.33)$	$(1, 5)$	$(3.0, 3.03, 1.62)$	$(18.0, 18.22, 5.21)$
$(0.12, 0.2, 0.33)$	$(1, 5)$	$(3.0, 3.05, 1.63)$	$(18.0, 18.35, 5.42)$
$(0.12, 0.17, 0.33)$	$(1, 5)$	$(3.0, 3.06, 1.64)$	$(18.0, 18.43, 5.56)$
$(0.12, 0.17, 0.2)$	$(1, 5)$	$(5.0, 5.01, 1.85)$	$(50.0, 50.15, 10.7)$
$(0.12, 0.14, 0.33)$	$(1, 5)$	$(3.0, 3.07, 1.65)$	$(18.0, 18.49, 5.65)$
$(0.12, 0.14, 0.2)$	$(1, 5)$	$(5.0, 5.02, 1.86)$	$(50.0, 50.26, 10.98)$

References

1. Balle, B., Wang, Y.: Improving the Gaussian mechanism for differential privacy: analytical calibration and optimal denoising. CoRR abs/1805.06530 (2018). http://arxiv.org/abs/1805.06530
2. Dwork, C., Kenthapadi, K., McSherry, F., Mironov, I., Naor, M.: Our data, ourselves: privacy via distributed noise generation. In: Vaudenay, S. (ed.) EUROCRYPT 2006. LNCS, vol. 4004, pp. 486–503. Springer, Heidelberg (2006). https://doi.org/10.1007/11761679_29
3. Dwork, C., McSherry, F., Nissim, K., Smith, A.: Calibrating noise to sensitivity in private data analysis. In: Halevi, S., Rabin, T. (eds.) TCC 2006. LNCS, vol. 3876, pp. 265–284. Springer, Heidelberg (2006). https://doi.org/10.1007/11681878_14
4. Dwork, C., Roth, A.: The algorithmic foundations of differential privacy. Found. Trends Theor. Comput. Sci. **9**(3–4), 211–407 (2014). https://doi.org/10.1561/0400000042
5. Fan, L., Xiong, L.: Real-time aggregate monitoring with differential privacy. In: 21st ACM International Conference on Information and Knowledge Management, CIKM 2012, Maui, HI, USA, 29 October – 02 November 2012, pp. 2169–2173 (2012). https://doi.org/10.1145/2396761.2398595
6. Geng, Q., Ding, W., Guo, R., Kumar, S.: Truncated Laplacian mechanism for approximate differential privacy. arXiv preprint arXiv:1810.00877 (2018)
7. Hardt, M., Talwar, K.: On the geometry of differential privacy. In: Proceedings of the Forty-Second ACM Symposium on Theory of Computing, pp. 705–714. ACM (2010)
8. Hill, K.: How Target Figured Out a Teen Girl was Pregnant Before Her Father Did. Forbes, Inc., Jersey City (2012)
9. Li, C., Hay, M., Rastogi, V., Miklau, G., McGregor, A.: Optimizing linear counting queries under differential privacy. In: Proceedings of the Twenty-Ninth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, PODS 2010, 6–11 June 2010, Indianapolis, Indiana, USA, pp. 123–134 (2010). https://doi.org/10.1145/1807085.1807104

10. Machanavajjhala, A., Gehrke, J., Kifer, D., Venkitasubramaniam, M.: l-diversity: privacy beyond k-anonymity. In: 22nd International Conference on Data Engineering (ICDE 2006), pp. 24–24. IEEE (2006)
11. Narayanan, A., Shmatikov, V.: How to break anonymity of the Netflix prize dataset. CoRR abs/cs/0610105 (2006). http://arxiv.org/abs/cs/0610105
12. Nikolov, A., Talwar, K., Zhang, L.: The geometry of differential privacy: the sparse and approximate cases. In: Proceedings of the Forty-Fifth Annual ACM Symposium on Theory of Computing, pp. 351–360. ACM (2013)
13. Perry, D.: Sex and Uber's "rides of glory": the company tracks your one-night stands–and much more'. Comput. Law Secur. Rev. (2014)
14. Rajendran, K., Jayabalan, M., Rana, M.E.: A study on k-anonymity, l-diversity, and t-closeness techniques. IJCSNS **17**(12), 172 (2017)
15. Smith, D.B., Thilakarathna, K., Kâafar, M.A.: More flexible differential privacy: the application of piecewise mixture distributions in query release. CoRR abs/1707.01189 (2017). http://arxiv.org/abs/1707.01189
16. Sweeney, L.: k-anonymity: a model for protecting privacy. Int. J. Uncertainty Fuzziness Knowl.-Based Syst. **10**(05), 557–570 (2002)
17. Voigt, P., Von dem Bussche, A.: The EU General Data Protection Regulation (GDPR). A Practical Guide, 1st edn. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-57959-7