1 Introduction

Over the past decade the collection and analysis of personal data has increased a lot. This has also raised the question about loss of privacy. Privacy and (useful) data analysis are relatively conflicting terms and thus combining them is a difficult question. In order to discuss the problem, we need to define privacy. Various definitions have been proposed but in the current report we will talk about one of them. Namely differential privacy, which was introduced by Dwork in 2006 [1]. There has been active research of differential privacy after that. In 2014, Dwork and Roth published a comprehensive introduction [2] to the field of differential privacy, which the current report is also based on.

Suppose a trusted curator (e.g. government) gathers sensitive information from a large number of respondents, with the goal of learning (and releasing to the public) statistical facts about the underlying population. The problem is, how to release statistical information without compromising the privacy of the individual respondents. The release phase can also be imagined as a data analyst querying the database.

Roughly speaking, differential privacy ensures that the removal or addition of a single database item does not affect (much) the outcome of any analysis. Thus no privacy is lost by joining the database.

In the current report, we firstly talk briefly about few naive privacy preserving methods and argue that they do not really work. In the second section we set up the model of computation and introduce the formal definition of differential privacy. In the third section, we describe few mechanisms that comply with the differential privacy definition. Finally, we will introduce the notion of composition and show how to analyze the privacy loss over several queries.

1.1 Privacy-preserving data analysis

Before defining differential privacy, we discuss some other possible approaches to privacy in data analysis.

1.1.1 Anonymization

A relatively common belief has been that once we remove direct identification info from data (name, telephone, address), it becomes anonymous and it is not possible to re-identify people. This has been proven incorrect, most famously by linking anonymous Netflix data with Internet Movie Database data [3]. The problem lies in the richness of data, a person can
be often identified by combination of non-obvious data fields (like a combination of zip-code, data of birth and sex; or the names of three movies and the approximate dates they were watched).

1.1.2 Queries over large data sets

One might believe that forcing queries to be summary answers over a large dataset would provide a reasonable privacy guarantee. However, consider answers to two queries (over the same dataset) "How many people in the database have cheated?" and "How many people other than X have cheated?". Both answers are over a large dataset but if the second answer is smaller than first, you know that X has cheated.

1.1.3 Query auditing

It might be tempting to audit the sequence of all queries and responses and limit the answers that would compromise privacy. The main problem with this is that if the query language is sufficiently rich it might not be possible to analyze the queries properly. Secondly, simply not answering the query might leak privacy.

2 Definition

In this section we will describe the model of computation and talk about possible privacy goals. Then, we formulate the definition of differential privacy and talk about the properties following from the definition.

2.1 Model of computation

We assume the existence of a trusted and trustworthy curator who holds the data of individuals in a database $x$. Each row $d_i$ represents data of a single individual and there are $n$ rows

$$x = \{d_i | i = 1...n\}.$$

The privacy goal is to protect each individual row while permitting statistical analysis of the database as a whole.

Secondly we have a data analyst, who can interact with the database through queries.

Privacy mechanism is an algorithm that takes as input a database, a universe $X$ of data types (the set of all possible database rows), random bits and set of queries and produces an output string. In the current report we only consider non-interactive queries where the whole set of queries is known while producing an answer.

2.2 Privacy goals

Based on cryptographic privacy definitions it would be natural to come up with few too strong requirements. First we briefly discuss what cannot be achieved in any setting while maintaining any utility of data analysis.
2.2.1 Participating in the database does not have any effect on results

If the data of any one individual would have no effect, then by induction nobody’s data would have any effect. Thus making the analysis meaningless.

2.2.2 Analyst knows nothing more about individual after analysis

This cannot be achieved if we want the analysis to be useful. Consider a case where analyst before analysis believes that people have 2 left feet, thus he has the same assumption about any specific individual. After analysis he finds out that most people have one left and one right foot. Now, analyst can reasonably assume that any specific individual is also more likely to have one left and one right foot. Thus, his assumption about any specific individual has changed, after getting to know the results of the analysis. If that specific person has one left and one right foot, he has lost some privacy after analyst found out the results of analysis. Notice, that it doesn’t matter if that specific person was in the database or not, if he is similar to database average, his privacy was compromised.

2.3 Formal definition of differential privacy

Differential privacy definition gives a guarantee that based on published results it is very difficult to say whether a specific person participated in the database or not.

Database $X$ is a collection of records from universe $X$. It is often convenient to represent a database as histograms: $x \in \mathbb{N}^{|X|}$, where each entry $x_i$ represents the number of elements in the database of type $i \in X$. Then the natural measure of the distance between databases is the number of rows that are different, which can be described by $\ell_1$ distance

**Definition 1** (Distance Between Databases). The $\ell_1$ norm of a database $x$ is denoted $\|x\|_1$ and is defined to be

$$\|x\|_1 = \sum_{i=1}^{|X|} |x_i|.$$  

In that notation, $\|x\|_1$ shows how many records there are in database $x$ and $\|x - y\|_1$ shows how many records are different between $x$ and $y$.

**Definition 2** (differential privacy). A randomized algorithm $M$ with domain $\mathbb{N}^{|X|}$ is $(\varepsilon, \delta)$-differentially private if for all $S \subseteq \text{Range}(M)$ and for all $x, y \in \mathbb{N}^{|X|}$ such that $\|x - y\|_1 \leq 1$:

$$P[M(x) \in S] \leq \exp(\varepsilon) P[M(y) \in S] + \delta,$$

where the probability space is over the randomness of mechanism $M$. If $\delta = 0$, we say that $M$ is $\varepsilon$-differentially private.

Note that $\varepsilon$ is usually a small constant but not negligible. Mostly we consider cases where $\delta = 0$. Also, the quantity

$$L(r) = \ln \left( \frac{P[M(x) = r]}{P[M(y) = r]} \right)$$

is commonly referred to as privacy loss incurred by observing $r$. It can be positive (when event is more likely under $x$ than under $y$) or negative (event is more likely under $y$ than under $x$).
2.4 Properties of differential privacy

1. Protection against arbitrary risks, moving beyond protection against re-identification.
2. Automatic neutralization of linkage attacks.
3. Quantification of the privacy loss.
4. Composition allows to analyze cumulative privacy loss over multiple computations.
5. Group privacy. If algorithm is $\varepsilon$-private for single records, it is $k\varepsilon$-private for groups of size $k$.
6. Immunity to post-processing. No current or future auxiliary information can increase the privacy loss.

We prove immunity to post-processing for deterministic functions.

Lemma 3. Let $M : \mathbb{N}^{\mathcal{X}} \rightarrow R$ be a randomized algorithm that is $(\varepsilon,\delta)$-differentially private. Let $f : R \rightarrow R'$ be a deterministic function. Then $f \circ M : \mathbb{N}^{\mathcal{X}} \rightarrow R'$ is $(\varepsilon,\delta)$-differentially private.

Proof. For any pair of adjacent databases $x,y$ with $\|x-y\|_1 \leq 1$ and any event $S \subseteq R'$ we have that if $T = \{ r \in R : f(r) \in S \}$
\[
P[f(M(x)) \in S] = P[M(x) \in T] \\
\leq \exp(\varepsilon) P[M(y) \in T] + \delta \\
= \exp(\varepsilon) P[f(M(y)) \in S] + \delta
\]

Lemma 3 holds also for any randomized post-processing functions.

3 Basic Techniques

In this section we describe some of the basic algorithms that satisfy differential privacy definition. These basic building blocks can then be combined to get more sophisticated algorithms. The composition theorem described in the next chapter allows us to easily analyze the cumulative privacy loss.

3.1 Randomized response

When faced with question "Have you done XYZ?", the respondent is instructed to perform the following steps:

1. Flip a coin.
2. If tails, then respond honestly.
3. If heads, flip a second coin and respond "Yes" if heads and "No" if tails.

This allows for "plausible deniability", we cannot say anything sure for any specific answer.
Lemma 4. Randomized response described above is \((\ln 3, 0)\)-differentially private.

Proof. We analyze the worst case, when we only have single respondent answer. Our outcome space is \{Yes, No\}. For both, we want to know what is the probability that observed answer is the same as the actual answer. Clearly, in both cases it is 3/4, leaving 1/4 for the incorrect observation. Thus, the ratio is

\[
\frac{P[\text{Response} = \text{Yes}|\text{Truth} = \text{Yes}]}{P[\text{Response} = \text{Yes}|\text{Truth} = \text{No}]}, \quad \frac{P[\text{Response} = \text{No}|\text{Truth} = \text{No}]}{P[\text{Response} = \text{No}|\text{Truth} = \text{Yes}]} = 3.
\]

Randomized response is a nice technique as it does not require a trusted database curator. However, it has pretty bad privacy parameters, queries based on several fields have low accuracy and it does not work for real valued queries. In the following we look at methods that solve these issues, but do require a trusted database holder who has the 'real' source data.

3.2 Laplace mechanism

Let \(f : \mathbb{N}^{|X|} \rightarrow \mathbb{R}^k\) be a numeric query that maps database to \(k\) different real numbers.

Definition 5 (\(\ell_1\)-sensitivity). The \(\ell_1\)-sensitivity of function \(f\) is

\[
\Delta f = \max_{x,y \in \mathbb{N}^{|X|}} \frac{\|f(x) - f(y)\|_1}{\|x - y\|_1 = 1}
\]

The \(\ell_1\)-sensitivity shows the magnitude how much one individual’s data can change the function result. This is a natural measure of how much noise we must add to the answer to give privacy for any one individual.

Definition 6 (Laplace Distribution). The Laplace Distribution (centered at 0) with scale \(b\) is the distribution with probability density function:

\[
\text{Lap}(x|b) = \frac{1}{2b} \exp \left( \frac{-|x|}{b} \right).
\]

Often we write \(\text{Lap}(b)\) to denote a random variable \(X \sim \text{Lap}(b)\).

Laplace distribution is a symmetric version of the exponential function.

Definition 7 (Laplace Mechanism). Given numeric function \(f : \mathbb{N}^{|X|} \rightarrow \mathbb{R}^k\), the Laplace mechanism is defined as:

\[
M_L(x, f(\cdot), \varepsilon) = f(x) + (Y_1, \ldots, Y_k)
\]

where \(Y_i\) are i.i.d. random variables drawn from \(\text{Lap}(\Delta f/\varepsilon)\).

Theorem 8. The Laplace mechanism provides \((\varepsilon, 0)\)-differential privacy.

Proof. Let \(x, y \in \mathbb{N}^{|X|}\) be such that \(\|x - y\|_1 \leq 1\) and let \(f(\cdot)\) be some function \(f : \mathbb{N}^{|X|} \rightarrow \mathbb{R}^k\). Let \(p_k\) denote the probability density function of \(M_L(y, f, \varepsilon)\) and let \(p_y\) denote the probability
density function of \( M_L(y, f, \varepsilon) \). We compare the two at some arbitrary point \( z \in \mathbb{R}^k \):

\[
\frac{p_x(z)}{p_y(z)} = \prod_{i=1}^{k} \left( \frac{\exp \left( -\frac{\varepsilon |f(x)_i - z|}{\Delta f} \right)}{\exp \left( -\frac{\varepsilon |f(y)_i - z|}{\Delta f} \right)} \right)
\]

\[
= \prod_{i=1}^{k} \exp \left( \frac{\varepsilon (|f(y)_i - z_i| - |f(x)_i - z_i|)}{\Delta f} \right)
\]

\[
\leq \prod_{i=1}^{k} \exp \left( \frac{\varepsilon |f(y)_i - f(x)_i|}{\Delta f} \right)
\]

\[
= \exp \left( \frac{\varepsilon \sum_{i=1}^{k} |f(y)_i - f(x)_i|}{\Delta f} \right)
\]

\[
= \exp \left( \frac{\varepsilon \|f(x) - f(y)\|}{\Delta f} \right) \leq \exp(\varepsilon)
\]

where the first inequality follows from the triangle inequality and the last from definition of sensitivity. The other inequality

\[
\frac{p_y(x)}{p_x(z)} \leq \exp(\varepsilon)
\]

follows by symmetry.

We can also easily analyze the accuracy of the Laplace mechanism.

**Theorem 9.** Let \( f : \mathbb{N}^{|X|} \to \mathbb{R}^k \), and let \( y = M_L(x, f(\cdot), \varepsilon) \). Then \( \forall \delta \in (0, 1) \):

\[
P \left[ \|f(x) - y\|_{\infty} \geq \left( \ln \frac{k}{\delta} \right) \frac{\Delta f}{\varepsilon} \right] \leq \delta.
\]

**Proof.** Since for any \( Y \sim \text{Lap}(b) \) we have that

\[
P[|Y| \geq t \cdot b] = \exp(-t)
\]

and \( Y_i \sim \text{Lap}(\Delta f / \varepsilon) \) we can write

\[
P \left[ \|f(x) - y\|_{\infty} \geq \left( \ln \frac{k}{\delta} \right) \frac{\Delta f}{\varepsilon} \right] = P \left[ \max_{i \leq k} |Y_i| \geq \left( \ln \frac{k}{\delta} \right) \frac{\Delta f}{\varepsilon} \right]
\]

\[
\leq k \cdot P \left[ |Y_i| \geq \left( \ln \frac{k}{\delta} \right) \frac{\Delta f}{\varepsilon} \right]
\]

\[
= k \cdot \exp \left( \ln \frac{k}{\delta} \right) = k \cdot \frac{\delta}{k} = \delta.
\]

\[\square\]

### 3.3 Exponential mechanism

Laplace mechanism gave us a method to answer numerical queries by adding an appropriate amount of noise to the real answer. However, there is a large class of queries where we would like to know the *best* answer from some range and adding noise to that answer would completely destroy its value. Imagine case when answers are not numeric, like *What is most
common favorite color?’. Or the case of selecting an optimal price for auction while preserving privacy of individual bids - adding some noise to the optimal price could severely modify the revenue.

In a very high level view, exponential mechanism samples answer from allowed range where the "better" answer has higher probability.

**Definition 10** (Exponential Mechanism). Given some arbitrary range $R$ and utility function $u : \mathbb{N}^{|X|} \times R \to \mathbb{R}$ (which maps database/output pairs to utility scores), exponential mechanism $M_{E}(x, u, R)$ selects and outputs an element $r \in R$ with probability proportional to

$$\exp \left( \frac{\varepsilon u(x, r)}{2\Delta u} \right),$$

where

$$\Delta u = \max_{r \in R, \|x-y\| \leq 1} |u(x, r) - u(y, r)|$$

**Theorem 11.** The exponential mechanism preserves $(\varepsilon, 0)$-differential privacy.

**Proof.** For adjacent databases $x, y \in \mathbb{N}^{|X|}$

$$\frac{\Pr [M_{E}(x, u, R) = r]}{\Pr [M_{E}(y, u, R) = r]} = \frac{\exp \left( \frac{\varepsilon u(x, r)}{2\Delta u} \right)}{\exp \left( \frac{\varepsilon u(y, r)}{2\Delta u} \right)} \cdot \frac{\sum_{r' \in R} \exp \left( \frac{\varepsilon u(x, r')}{2\Delta u} \right)}{\sum_{r' \in R} \exp \left( \frac{\varepsilon u(y, r')}{2\Delta u} \right)}$$

$$= \exp \left( \frac{\varepsilon (u(x, r) - u(y, r))}{2\Delta u} \right) \cdot \left( \frac{\sum_{r' \in R} \exp \left( \frac{\varepsilon u(x, r')}{2\Delta u} \right)}{\sum_{r' \in R} \exp \left( \frac{\varepsilon u(y, r')}{2\Delta u} \right)} \right)$$

$$\leq \exp \left( \frac{\varepsilon}{2} \right) \cdot \exp \left( \frac{\varepsilon}{2} \right) \cdot \left( \frac{\sum_{r' \in R} \exp \left( \frac{\varepsilon u(x, r')}{2\Delta u} \right)}{\sum_{r' \in R} \exp \left( \frac{\varepsilon u(y, r')}{2\Delta u} \right)} \right)$$

$$= \exp(\varepsilon).$$

Similarly

$$\Pr [M_{E}(y, u, R) = r] \leq \Pr [M_{E}(x, u, R) = r]$$

follows by symmetry.

We give without proof the following result about accuracy of the exponential mechanism.

**Theorem 12.** Fixing a database $x$ we have

$$\Pr \left[ u(M_{E}(x, u, R)) \leq \text{OPT} \frac{u(x, r)}{\varepsilon} (\ln |R| + t) \right] \leq e^{-t},$$

where $\text{OPT} = \max_{r \in R} u(x, r)$ is the maximum utility score over all elements in database $x$.

What theorem 12 says is that it is very unlikely that the element $r$ returned by exponential mechanism has a utility score which is lower than optimal minus additive factor of $O((\Delta/\varepsilon) \ln |R|)$. 7
4 Composition

In the previous chapter we saw some of the building blocks for differential privacy. Now we are interested in how the privacy loss behaves over multiple differentially private analyses. More specifically we are interested in the following scenarios:

1. Repeated use of (different) differentially private algorithms on the same database.

2. Repeated use of differentially private algorithms on different databases that may contain information related to the same individual.

4.1 Basic composition

First we show that independent use of \((\varepsilon_1,0)\)-differentially private algorithm and \((\varepsilon_2,0)\)-differentially private algorithm, when taken together, is \((\varepsilon_1 + \varepsilon_2,0)\)-differentially private.

**Theorem 13** (Basic Composition). Let \(M_1 : \mathbb{N}^{[X]} \rightarrow R_1\) be an \(\varepsilon_1\)-differentially private algorithm and \(M_2 : \mathbb{N}^{[X]} \rightarrow R_2\) be an \(\varepsilon_2\)-differentially private algorithm. Then their combination, defined to be \(M_{1,2} : \mathbb{N}^{[X]} \rightarrow R_1 \times R_2\) by mapping: \(M_{1,2}(x) = (M_1(x), M_2(x))\) is \(\varepsilon_1 + \varepsilon_2\)-differentially private.

**Proof.** Let \(x,y \in \mathbb{N}^{[X]}\) be adjacent databases. Fix any \((r_1, r_2) \in R_1 \times R_2\). Then

\[
\frac{P[M_{1,2}(x) = (r_1, r_2)]}{P[M_{1,2}(y) = (r_1, r_2)]} = \frac{P[M_1(x) = r_1] \cdot P[M_2(x) = r_2]}{P[M_1(y) = r_1] \cdot P[M_2(y) = r_2]}
\]

\[
= \left(\frac{P[M_1(x) = r_1]}{P[M_1(y) = r_1]}\right) \cdot \left(\frac{P[M_2(x) = r_2]}{P[M_2(y) = r_2]}\right)
\]

\[
\leq \exp(\varepsilon_1) \cdot \exp(\varepsilon_2) = \exp(\varepsilon_1 + \varepsilon_2).
\]

By symmetry we also have

\[
\frac{P[M_{1,2}(y) = (r_1, r_2)]}{P[M_{1,2}(x) = (r_1, r_2)]} \leq \exp(\varepsilon_1 + \varepsilon_2)
\]

By repeatedly applying the composition theorem we get the following lemma:

**Lemma 14.** Let \(M_i : \mathbb{N}^{[X]} \rightarrow R_i\) be an \(\varepsilon_i\)-differentially private algorithm for \(i = 1, \ldots, k\). Then if \(M_{[k]} : \mathbb{N}^{[X]} \rightarrow R_1 \times \cdots \times R_k\) is defined to be \(M_{[k]}(x) = (M_1(x), \ldots, M_k(x))\), then \(M_{[k]}\) is \((\sum_{i=1}^{k} \varepsilon_i)\)-differentially private.

Similar result holds for \((\varepsilon, \delta)\)-differentially private algorithms as well.

Now, if we have \(k\) queries and we want the overall algorithm to be \(\varepsilon\)-differentially private, we simply have to require that each individual query would be \(\frac{\varepsilon}{k}\)-differentially private.
4.2 Advanced composition

Basic composition theorem was easy to analyze and gave us a nice result. However, it assumes that privacy loss on each query is positive - each query always decreases the privacy. When we look at the average case, the picture is much much nicer, as we can also gain privacy on some queries (if the output result said that it was less likely that an individual was in the database). Thus, if we allow for some small probability that privacy loss is bigger, we can get a much better bound for composition, i.e. we can trade some small $\delta$ for better $\varepsilon$.

Without proof we present the following theorem:

**Theorem 15** (Advanced Composition). For all $\varepsilon, \delta, \delta' \geq 0$, the class of $(\varepsilon, \delta)$-differentially private mechanisms satisfies $(\varepsilon', k\delta + \delta')$-differential privacy under $k$-fold adaptive composition ($k$ consecutive queries over any adjacent databases) for

$$ \varepsilon' = \varepsilon \cdot \sqrt{2k \ln \left(\frac{1}{\delta'}\right)} + k\varepsilon (e^\varepsilon - 1). $$

When $\varepsilon$ is reasonably small, we have that $\varepsilon' = O(\varepsilon \sqrt{k})$ compared to $k\varepsilon$ for basic composition.

**Example 16.** Suppose that over his lifetime, Bob is a member of $k = 10000$ $(\varepsilon_0, 0)$-differentially private databases. For Bob’s total cumulative privacy loss to be $\varepsilon = 1$ with probability at least $1 - e^{-32}$ and $\delta' = e^{-32}$ he needs to have $\varepsilon_0 \leq 1/801$ for each database.

5 Conclusion

Differential privacy is a mathematically rigorous definition of privacy that allows to quantify privacy loss for an individual person in the database after publishing some statistical results about the database. The definition ensures that no additional privacy loss can occur after the results are published, i.e. no post-processing or additional information affects the privacy. Second important property of differential privacy definition is that it allows to analyze the total privacy loss of several statistical queries and thus set privacy requirements for any individual query. We also described mechanisms that satisfy differential privacy and that can be made arbitrarily private by sacrificing the accuracy of the result.

The use of differentially private algorithms can encourage people to submit their personal data which can be analyzed for the good of society. I believe that the wider application of this concept is important for us to make good decisions.
References

