Differential Privacy
A Survery

20398702 HU, Jiajun
ZHOU, Lei
Department of Computer Science & Engineering
The Hong Kong University of Science and Technology
April 28, 2017

Abstract

Keywords: differential privacy

1 Introduction

2 Differentail Privacy

Over the past ten years, differentail privacy[1, 2] has emerged to become one of the most powerful approaches to ensure data pricacy. Roughly speaking, differential privacy ensures that the removal or insertion of a single record does not significantly affect the outcome of any analysis conducted on the database, thus making it possible to prevent private information from exposing to attackers. It follows a rigorous mathematical deduction to prove it can reduce the risk of privacy breach while remaining the utility of the data. At the beginning of this section, we will illuminate the concept by leveraging a simple example. Then, we will give the mathematical definition of differential privay and introduce two privacy mechanisms to achieve it.

2.1 A Simple Example

Suppose you have access to a database that allows you to compute the total income of all resident in certain area. You know one of your friends, Mr. White is going to move to another area, so simply computing the total

Name	Annual Income
Mr. Richard	0.5 million
Mr. White	1 million
Mr. Brown	2 million
Ms. Lee	0.35 million
Ms. Jean	0.6 million
...	...
Total income = 50 million	

Name	Annual Income
Mr. Richard	0.5 million
Mr. Brown	2 million
Ms. Lee	0.35 million
Ms. Jean	0.6 million
...	...
Total income = 49 million	

Table 1: The table before and after Mr. White's move.

income of all resident before and after Mr. White's move would allow you to guess his real income. As shown in table 1, the total income of all residents before Mr.White's move is 50 million, while the total income of all residents after Mr.White's move is 49 million. One can compute the real income of Mr.White is 1 millon. So from this example we can see even though we are not allowed to retrieve the information of a particular person, we are still able to get the private informtion through certain opertions. So what could one do to stop this? In the next section, we wil see how differential privacy can help resolve this problem.

2.2 Definition of Differential Privacy

Firstly, let us define some notations.

Definition 2.1. D and D' are databases, but they must differs on at most one row.

The reason why D and D' is required to differ on one row is to simulate whether a particular record is in or not in the database.

Definition 2.2. $f(D)$ is a query on D

Refer to the previous example, $f(D)$ is the total income of all residents in the database.

Definition 2.3. $M(D)$ is the privacy mechanism, which is a randomized function that takes the database D as inpiut, and release privatized information with respect to $f(D)$.

Refer to the previous example, $M(D)$ is the privated total income obtained by adding random noise on the total income.

Definition 2.4. ϵ - differential privacy A privacy mechanism M gives ϵ - differential privacy if for all data sets D and D' differing on at most one row, and all $C \in Range(M)$,

$$\frac{Pr[M(D) = C]}{Pr[M(D') = C]} < e^{\epsilon}$$

$\epsilon - differential\ privacy$ is a special case of $(\epsilon, \delta) - differential\ privacy$[3, 4] with $\delta = 0$. Typically, $(\epsilon, \delta) - differential\ privacy$ is simplified to $\epsilon - differential\ privacy$, so we only consider $\epsilon - differential\ privacy$ in this survey. $\epsilon - differential\ privacy$ says that the probability that the privatized result will be C is nearly the same whether or not you are in the database, which means the harm to you is nearly the same regardless of your participation. In the definition, ϵ is the privacy budget, which is a tradeoff that is used to balance the privacy of the result and it's utility. The smaller the ϵ is, the closer $Pr[M(D) = C]$ and $Pr[M(D') = C]$ are, and the stronger protection is.

2.3 Laplace Mechanism

2.4 Exponential Mechanism

Conclusion

References

1. Cynthia Dwork. Differential privacy: A survey of results. In *International Conference on Theory and Applications of Models of Computation*, pages 1–19. Springer, 2008.

2. Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

3. Cynthia Dwork. Differential privacy. In *Encyclopedia of Cryptography and Security*, pages 338–340. Springer, 2011.

4. Cynthia Dwork, Krishnaram Kenthapadi, Frank McSherry, Ilya Mironov, and Moni Naor. Our data, ourselves: Privacy via distributed noise generation. In *Annual International Conference on the Theory and Applications of Cryptographic Techniques*, pages 486–503. Springer, 2006.