Applications and Implementation of Differential Privacy on Statistical Databases

Jonathan Sumner Evans, Victoria Girkins, and Sam Sartor

***Abstract*—Statistical databases have long been know to be vulnerable to inference attacks. One method of mitigating these attacks is Differential Privacy. Our final project will implement a working Differential Privacy statistical database and then perform experiments using the database to analyze how well our implementation achieves the objectives of Differential Privacy databases.**

I. Introduction

Statistical databases are designed for statistical analysis of the datasets contained within the database. One of the desired properties of statistical databases is that nothing about an individual should be learnable from the database that cannot be learned without access to the database [2]. In other words, gathering information about the records underlying the statistics using allowed queries on the database should be impossible.

Without any protections, well crafted queries to a statistical database can reveal information about a single row. An attack of this form is called an *inference attack*. Take, for example, a class of 30 people that stores the students' homestate and age. If it is known from some other source that only one of them is from a certain state, determining the age of that person is easy and merely requires a calculation of the `sum` of the ages of people with that homestate. In this case, since there is only one person, the sum will be equal to that person's age.

One way to alleviate this problem and reduce the risk of a successful inference attack is to use a technique known as *Differential Privacy* — a term first coined by Cynthia Dwork at Microsoft Research in 2006 [3]. The goal of Differential Privacy is to maximize the statistical accuracy of data while also maximizing the privacy of the individual records in the database. It does this by introducing randomness into the dataset while still maintaining the statistical accuracy of the queries on that dataset. This randomness reduces the likelihood of obtaining information about the actual individual records in the database.

II. Approach

Differential Privacy has received adoption in many applications including the United States Census Bureau to show commuting patterns [4]. Apple also advertises its use of Differential Privacy on their Privacy page [1]. Because of this, our team has decided to take the previous research on the topic and attempt to implement a Differential Privacy statistical database.

Our team will research the best approach for implementing this, but currently, we are considering two approaches:

1) Taking an existing statistical database and adding a Differential Privacy component; or
2) Creating a simple statistical database with a Differential Privacy component built in.

Initially, we will limit the number of statistical functions that we implement to the minimum required to determine whether or not our implementation functions as intended. If time allows, we will add more advanced statistical functions to our implementation.

III. Expected Results

After implementing this database, our team will run experiments to determine whether or not our implementation achieves the following objectives:

1) The individual records are reasonably difficult to determine.
2) The statistical data returned by the queries from the database are sufficiently accurate.

If these two objectives are met, then the project will be considered a success.

IV. Conclusion

Statistical databases are useful tools for allowing users to analyze the datasets contained within the database while maintaining the privacy of individual records. There are challenges with this including the possibility of inference attacks but these can be mitigated by using techniques such as Differential Privacy. Our team's objective is to create a Differential Privacy statistical database and perform experiments to determine if our implementation achieves the objectives described in Section III.

References

[1] Apple. *Privacy - Approach to Privacy*. 2017. URL: https : / / www . apple . com / privacy / approach-to-privacy/.

[2] Cynthia Dwork. "Differential Privacy". In: *Microsoft Research* (2006). URL: https : / / pdfs . semanticscholar . org / 4c99 / 097af05e8de39370dd287c74653b715c8f6a . pdf.

[3] Michael Hilton. *Differential Privacy: A Historical Survey*. URL: http://www.cs.uky.edu/~jzhang/CS689/PPDM-differential.pdf.

[4] Ashwin Machanavajjhala et al. "Privacy: Theory meets Practice on the Map". In: (2008). URL: http://www.cse.psu.edu/~duk17/papers/PrivacyOnTheMap.pdf.