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. For our final project we implemented a simple Differential Privacy statistical database and then performed experiments using the database to analyze how well our implementation achieves the objectives of Differential Privacy databases.

I. Introduction

A. Previous Attempts at Anonymizing Statistical Data

Many attempts have been made to privatize data in the past and many are still in use today. However everything from simply removing columns containing personally identifiable information to advanced techniques such as as k-anonymity and l-diversity have been shown to be vulnerable to attack [2]. K-anonymity, for example, does not include any randomization, attackers can still make inferences about data sets [1].

B. Differential Privacy

II. Approach

III. Experiment

A. Dataset

B. Crafted Queries

C. Database Implementation

IV. Results

V. Limitations and Future Work

VI. Conclusion

References

[1] Charu Aggarwal. “On k-Anonymity and the Curse of Dimensionality.” In: VLDB 2005 - Proceedings of 31st International Conference on Very Large Data Bases. Vol. 2. Jan. 2005, pp. 901–909.

[2] Atockar. “Differential Privacy: The Basics”. In: (2014). url: https://research.neustar.biz/2014/09/08/differential-privacy-the-basics/.