CS 208 Project Proposal:
A Differentially Private Query System

Andrew Shackelford and Peter Chang

March 13, 2019

We plan to work on creating a differentially private query system for use by other data analysts. However, rather than start from scratch, we plan to work with current systems like Ektelo or PinQ, and extend them to new use cases that would be relevant for the Harvard community.

Differential privacy is a difficult concept to understand, let alone implement. As a result, we aim to create a system that abstracts away the difficult and hard-to-comprehend privacy mechanisms so that data analysts can focus their energy on the analysis itself. More importantly, using a robust differentially private query system will ensure that privacy is never compromised, as our system will not allow analysts to accidentally perform queries that might violate privacy.

We will then test our data privacy system on different databases, both Harvard-affiliated and non-Harvard affiliated. We plan to use both the US Census sample we used in the first problem set, as well as a dataset from Harvard's edX program. We may also consider other datasets that pose interesting problems we aim to tackle.

We plan to study the following works as the starting point for our project.

- Zhang et al: *Ektelo: A framework for defining differentially-private computations* (2018)
- Kotsogiannis et al: *Architecting a differentially private SQL engine* (2019)
- Gaboardi et al: *Psi (Ψ): a private data sharing interface* (2016)

Although we anticipate learning more about this in Friday's practicum, as well as researching it ourselves, we seek advice from the instructors on what current differentially private query systems lack. Only by understanding what data analysts want from their querying systems will we be able to architect our improvements to current frameworks.