Hi, I've created a proof-of-concept contract that implements differentially private COUNT and AVERAGE queries on on-chain data using the Laplace mechanism and pseudorandom generation on chain.

The repo is here: https://github.com/darwinzer0/scrt-diff-privacy-example. It still needs more documentation but basic info for running the tests is in there. When I get time I will probably write up a blog entry about it.

Features

The test script uploads a set of observations to the contract (sample data from the Iris dataset) and calculates fuzzy count and average values for the observations.

The substrate-fixed crate was used to implement noise calculation using fixed-point arithmetic. (The Laplace mechanism requires logarithm on a unit interval random number).

Contract initialization sets the epsilon value (higher epsilon gives closer to true value / less privacy, lower epsilon adds more noise / more privacy), and a privacy budget. Each query will use some of the privacy budget, and once it is exhausted no more queries are allowed on the data.

The queries are implemented as exec

functions in the contract, because they need to update how much of the privacy budget is left.

Next steps

This is a simple example demonstrating the capacity of Secret Network to create contracts that can return answers to differentially private queries on encrypted data. There's much more that could be done to make a general differential privacy toolkit for developers on SN, including:

- Refactoring into library functions, e.g. integration with AppendStore and better data structures to manage privacy budget and other metadata. There are different patterns this could take. Calculating on a list of data in storage could be expensive, e.g. calculating mean, but maybe it is only necessary to store a running count and sum, such as in the example contract.
- Adding more types of queries, e.g. histograms, crosstab.
- Better methods for calculating bounded sensitivity, including on log-scale data.
- Adding other noise mechanisms such as Gaussian and Exponential (for categorical data).
- Documentation of tools and best practices

, (e.g. how to pick an epsilon value, calculate sensitivity, privacy costs) – differential privacy is easy to get wrong!

- Analysis of attack vectors, e.g. possibility of using chain forks to circumvent the privacy budget?, etc.

I note that differential privacy was listed as a topic on the grants page a while ago: Secret Network Grants | Application Ideas for Blockchain Development. I'm not sure what was intended when it was put under the protocol improvement

section, but if a secret toolkit-style helper library is something that would be of interest to the community, please let me know.

A nice primer on differential privacy can be found here: Differential Privacy — Programming Differential Privacy