Towards General-Purpose Infrastructure for Protecting Scientific Data Under Study

Andrew Trask, Kritika Prakash OpenMined **PRIML - PPML NeurIPS 2020**

An end-to-end system for the automatic protection of data under study, such that data can be studied without being shared and without data owners actively participating in experiments.

Privacy Enhancing Technologies
provide guarantees to data owner & data scientist:
1. Protect data from being copied by data scientist
2. Protect queries from being copied by data owner
3. Prevent the algorithm from memorizing the data

Tools for end-to-end Private Data Analysis

1. RPC Federated Learning
2. Pre-publish & Post-publish composition
3. Object & user level permissions
4. Adaptive DP filter
5. Approximate DP Odometer
6. Privacy budget simulations
7. Individual Differential Privacy

Syft

Private Sensitivity Scalar
n entities contribute to a single scalar value ***y*** with metadata {g, x, f, c}
g: polynomial function over the set of entities
x: input vector to polynomial g, such that g(x) = **y**
f: lower bound of each component of x
c: upper bound of each component of x

This polynomial based representation of data allows us to keep track of the Lipschitz factor automatically when performing non-additive operations, allowing us to achieve individual Renyi DP guarantees.

https://github.com/OpenMined/PySyft/