ESORICS 2022

Multi-Freq-LDPy: Multiple Frequency Estimation Under Local Differential Privacy in Python

Héber H. Arcolezi‡, Jean-François Couchot†, Sébastien Gambs*, Catuscia Palamidessi‡, Majid Zolfaghari‡
‡ Inria and École Polytechnique † Univ. Bourg. Franche-Comté * Univ. du Québec à Montréal

INTRODUCTION

Privacy model: Local differential privacy (LDP) protects individual's privacy without relying on a trusted third party.

Contribution: Multi-freq-ldpy implements state-of-the-art LDP protocols [1-6] for frequency (or histogram) estimation of single, multidimensional, and/or longitudinal data.

$ pip install multi-freq-ldpy

license MIT

MODULES (TASKS COVERED)

Single Frequency Estimation →

LDP protocols for a single attribute:
- **General Randomized Response (GRR)**: a.k.a. k-RR or direct encoding [1];
- **Unary Encoding (UE)**: Symmetric UE [4] & Optimal UE [1];
- **Local Hashing (LH)**: Binary LH & Optimal LH [1];
- **Subset Selection (SS)** [2].

Multidimensional Frequency Estimation →

Solutions for multiple $d \geq 2$ attributes:
- **Splitting (SPL)**: Sanitizes each attribute with $\epsilon/_d$-LDP;
- **Random Sampling (SMP)**: Sanitizes a single (random) attribute with ϵ-LDP;
- **Random Sampling Plus Fake Data (RS+FD)** [3]: SMP plus random fake data for all $d-1$ non-sampled attributes.

Longitudinal Frequency Estimation →

LDP protocols for multiple collections:
- **Google's RAPPOR** [4];
- **Microsoft's dBitFlipPM** [5];
- **Longitudinal UE (L-UE)** [6] and **Longitudinal GRR (L-GRR)** [6].

Longitudinal & Multidimensional Frequency Estimation →

Combining solutions for multiple attributes and LDP protocols for multiple collections:
- **SPL solution with Longitudinal protocols**;
- **SMP solution with Longitudinal protocols** [6];

USAGE EXAMPLE (LONGITUDINAL FREQUENCY ESTIMATION)

Importing functions and numpy package.

```
# Multi-Freq-LDPy functions for L-SUE (RAPPOR [4]) protocol
from multi_freq_ldpy.long_freq_est.L_SUE import L_SUE_Client, L_SUE_Aggregator

# NumPy library
import numpy as np
```

Setting parameters and **generating** synthetic dataset for simulation.

```
# Parameters for simulation
eps_perm = 2   # longitudinal privacy
eps_1 = 0.5 * eps_perm   # first report privacy
n = int(1e6)   # number of users
k = 5   # attribute's domain size

# Simulation dataset following Uniform distribution
dataset = np.random.randint(k, size=n)
```

One **essentially needs 2 lines of code** to simulate the LDP data collection pipeline.

```
# Simulation of data collection
reports = [L_SUE_Client(user_data, k, eps_perm, eps_1) for user_data in dataset]

# Simulation of server-side aggregation
est_freq = L_SUE_Aggregator(reports, eps_perm, eps_1)
>>> array([0.199, 0.201, 0.200, 0.198, 0.202])
```

RESULTS & PERFORMANCE (LONGITUDINAL FREQUENCY ESTIMATION)

Setup of Experiments

Local machine:
- Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz, 16GB RAM, Windows 11.

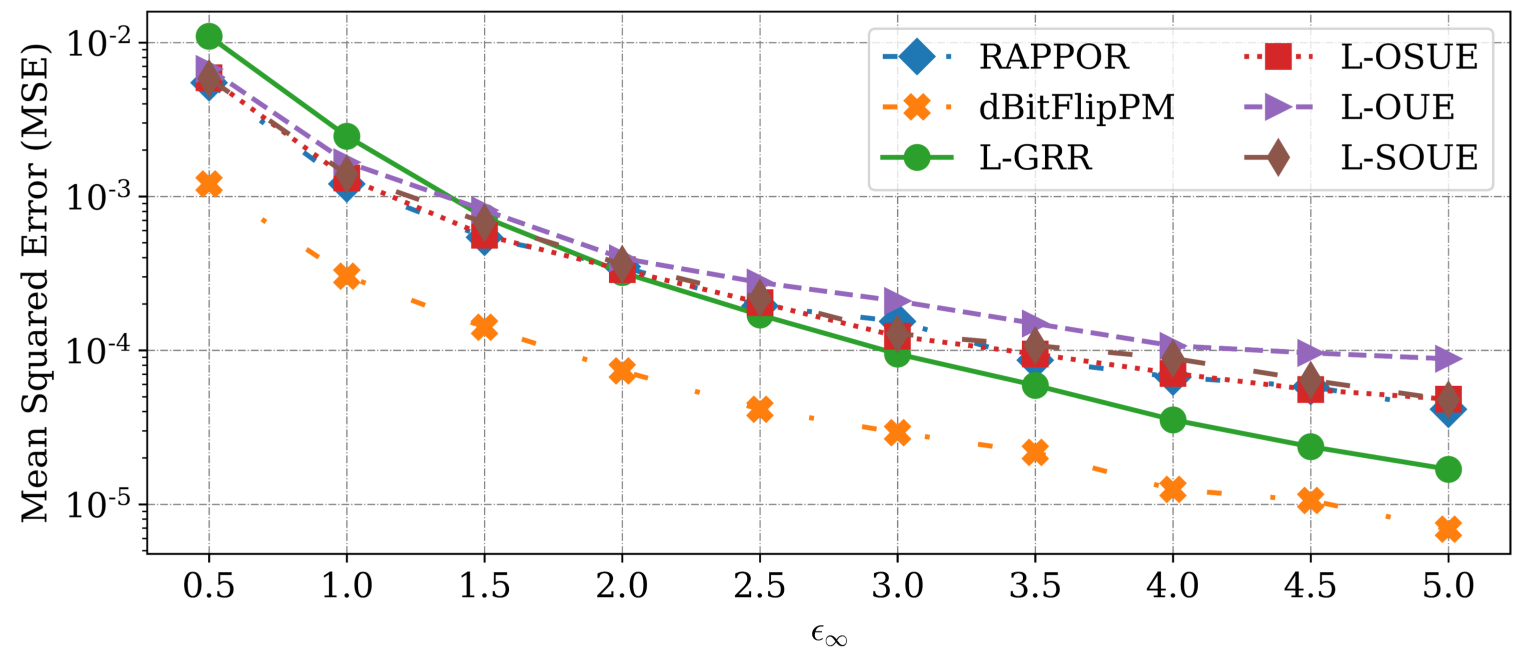
Dataset:
- $n = 10^4$ users, $k = 5$ values, Uniform distribution.

Privacy guarantees:
- $\epsilon_\infty = [0.5, 1, \ldots, 4.5, 5]$ and $\epsilon_1 = 0.5 \cdot \epsilon_\infty$.

 # iterations:
- 50 (as LDP protocols are randomized).

LDP Protocols	Execution Time (s)
RAPPOR [4]	11.15
dBitFlipPM [5]	94.96
L-GRR [6]	82.30
L-OSUE [6]	10.20
L-OUE [6]	10.75
L-SOUE [6]	11.64

Easy-to-use and fast execution toolkit to benchmark state-of-the-art LDP protocols.



CONTACT

heber.hwang-arcolezi@inria.fr
@hharcolezi

ARTIFACT

SCAN ME

REFERENCES

[1] Wang T, Blocki J, Li N, Jha S. Locally differentially private protocols for frequency estimation. In: USENIX Security, 2017.

[2] Wang S, et al. Mutual information optimally local private discrete distribution estimation. In: arXiv:1607.08025, 2016.

[3] Arcolezi HH, et al. Random sampling plus fake data: Multidimensional frequency estimates with local differential privacy. In: CIKM, 2021.

[4] Erlingsson Ú, Pihur V, Korolova A. RAPPOR: Randomized aggregatable privacy-preserving ordinal response. In: CCS, 2014.

[5] Ding B, Kulkarni J, Yekhanin S. Collecting telemetry data privately. In: NIPS, 2017.

[6] Arcolezi HH, et al. Improving the utility of locally differentially private protocols for longitudinal and multidimensional frequency estimates. In: Digital Communications and Networks, 2022.