A Unified View of Frequency Estimation and their Attacks on Local Differential Privacy

Al Mehdi Saadat Chowdhury* Dhaval Pankaj Tanna† Deepak Vellanki‡
Chirag Manjeshwar§

Abstract

Protecting individual's privacy while providing statistical summary of a population is a central goal in privacy-preserving data analysis. Local Differential Privacy (LDP) achieves this by perturbing user's encoded data before aggregation. This report examines how perturbed, encoded user responses can be combined for frequency estimation and how these estimators can be intentionally corrupted by injecting fake users to the system with the goal of increasing frequency of some target items (known as the frequency estimation problem). We focus on three LDP frequency-estimation protocols—kRR, OUE, and OLH—and study how they behave under a data-poisoning strategy known as the maximum-gain attack. For each protocol, we derive the corresponding maximum-gain expression and study how the attack's effectiveness varies with key parameters.

1 Introduction

Generating meaningful statistical summaries about a population without revealing information about any individual is the central goal of privacy-preserving data analysis. Since its introduction, Differential Privacy (DP) has become the gold-standard framework for analyzing sensitive data while providing rigorous privacy guarantees. For any randomized algorithm M, DP is defined [?] as the following:

Definition 1.1 (Differential Privacy). Consider any database x as a collection of records taken from a universe $\mathcal{X}$, and is represented by their histograms: $x \in \mathbb{N}^{|\mathcal{X}|}$ in which each entry x_i represents the number of elements in x of type $i \in \mathcal{X}$. A randomized algorithm $\mathcal{M}$ with domain $\mathbb{N}^{|\mathcal{X}|}$ is (ϵ, δ)-differentially private if for all $\mathcal{S} \subseteq Range(\mathcal{M})$ and for all $x, y \in \mathbb{N}^{|\mathcal{X}|}$ such that $\|x - y\|_1 \leq 1$:

$$Pr[\mathcal{M}(x) \in \mathcal{S}] \leq e^{\epsilon} Pr[\mathcal{M}(y) \in \mathcal{S}] + \delta$$

Stronger privacy guarantee is achieved by using smaller privacy loss bound parameter ϵ; the parameter δ represents the probability that the guarantee fails to hold.

DP requires a central curator who collects the dataset and perturbs it to preserve privacy. This not only creates legal, ethical, and technical burden on the curator, but also the privacy itself becomes vulnerable if the curator is compromised. Local differential privacy (LDP) can solve this issue by asking each user to encrypt their data before sending to the curator. The only job of the curator remains is to aggregate the data from all users.

Definition 1.2 (Local Differential Privacy [?]). An algorithm $\mathcal{M}$ satisfies ϵ-local differential privacay (ϵ-LDP), where $\epsilon \geq 0$, iff for any input v_1 and v_2, we have:

$$\forall y \in Range(\mathcal{M}) : Pr[\mathcal{M}(v_1) = y] \leq e^{\epsilon} Pr[\mathcal{M}(v_2) = y]$$

LDP can be ensured by following three protocols. The `Encode` protocol takes an input value v and outputs an encoded value x. The `Perturb` protocol returns a noisy version of the encoded x as $y = Perturb(Encode(v))$. The `Aggregate` protocol takes perturbed values from all users and returns any required aggregate information. The first two protocols, `Encode` and `Perturb`, are executed by the user (we will combine them into one as `PE(v)`), and the curator executes `Aggregate`.

[*] CSE PhD 3rd year

[†] CSE Master's 2nd year

[‡] CSE Master's 2nd year

[§] CSE Master's 2nd year