Demo: Exploring Utility and Attackability Trade-offs in Local Differential Privacy

Haoying Zhang, Abhishek K Mishra, Héber H. Arcolezi

Demo: Exploring Utility and Attackability Trade-offs in Local Differential Privacy

Haoying Zhang*
Inria Saclay, INSA CVL
Palaiseau, France
haoying.zhang@inria.fr

Abhishek K. Mishra*
Inria Lyon
Lyon, France
abhishek.mishra@inria.fr

Héber H. Arcolezi
Inria Grenoble
Grenoble, France
heber.hwang-arcolezi@inria.fr

Abstract

Local Differential Privacy (LDP) provides strong, formal privacy guarantees without requiring a trusted curator, making it a promising approach for privacy-preserving data collection and analysis. However, despite extensive research, practitioners may struggle to understand how to tune LDP parameters and anticipate the impact on data utility and attack risks for their specific scenarios. To address this gap, we demonstrate LDP-Toolbox, the first interactive, web-based toolbox (implemented in Python) that enables practical, analytical visualization of trade-offs between privacy loss (ε), utility loss, and vulnerability to attacks. The toolbox supports exploration of these trade-offs using real-world datasets from different domains; in this demonstration, we focus on discrete personal attributes and location-based scenarios. By providing intuitive, visual insights, LDP-Toolbox lowers the barrier to deploying LDP in real applications and helps bridge the gap between theoretical guarantees and practical adoption. The toolbox is open-source on PyPI (https://pypi.org/project/ldp-toolbox) and a video is available on our GitHub repository (https://github.com/hharcolezi/ldp-toolbox).

CCS Concepts

• **Security and privacy → Software and application security**; **Privacy-preserving protocols**.

Keywords

Local Differential Privacy, Privacy-Utility Trade-Off, Open Source.

ACM Reference Format:
Haoying Zhang, Abhishek K. Mishra, and Héber H. Arcolezi. 2025. Demo: Exploring Utility and Attackability Trade-offs in Local Differential Privacy. In *Proceedings of the 2025 ACM SIGSAC Conference on Computer and Communications Security (CCS '25), October 13–17, 2025, Taipei, Taiwan.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3719027.3760706

1 Introduction

With growing public concern over data privacy and strict regulations such as the GDPR, Big tech companies have increasingly deployed some of their systems under Local Differential Privacy (LDP) [7] to ensure user-level protection. LDP provides strong, formal guarantees by perturbing each user's data before collection, enabling population statistics and machine learning without requiring a trusted curator. Formally, a mechanism $\mathcal{M}: \mathcal{X} \to \mathcal{Y}$ is ε-LDP if for every measurable $S \subseteq \mathcal{Y}$ and every $x, x' \in \mathcal{X}$: $\Pr[\mathcal{M}(x) \in S] \leq e^{\varepsilon} \Pr[\mathcal{M}(x') \in S]$.

In practice, LDP has been deployed at scale by Big tech companies; for example, Google used LDP to estimate the frequency of unsafe browser settings, and Apple applies it to measure emoji usage and app crash rates on iOS devices. Both deployments are *frequency estimation tasks*, which remain one of the most widely studied and deployed use cases for LDP. In frequency estimation, each user holds a private value x_i drawn from a finite domain $[k] = \{0, \ldots, k-1\}$. The goal is for an untrusted server to recover an accurate estimate of the population histogram $\mathbf{f} \in \mathbb{R}^k$, where $f_v = \frac{1}{n}\#\{i : x_i = v\}$. After collecting the randomized reports $\{Y_i\}_{i=1}^n$, the server computes an estimate $\hat{\mathbf{f}}$ that minimizes its distance from $\mathbf{f}$ under some utility loss metric.

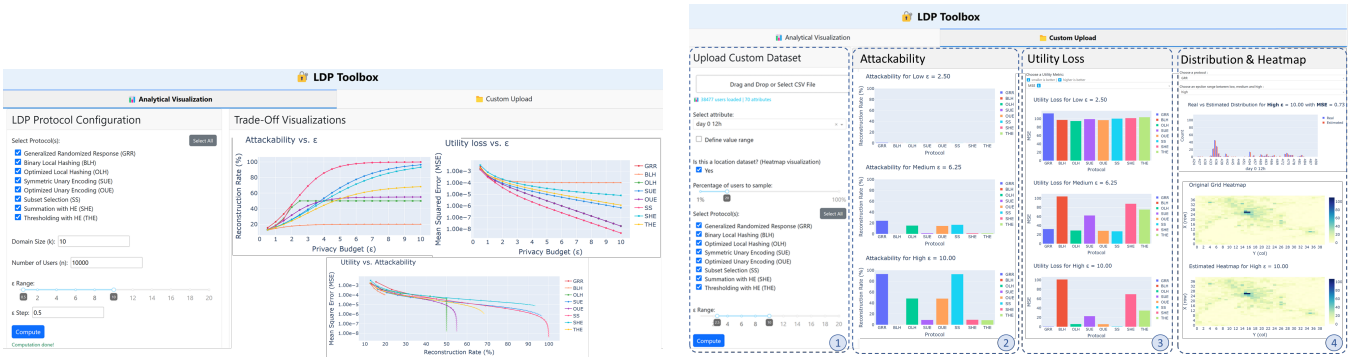
While much work has focused on maximizing utility, another fundamental concern in LDP is *Data Reconstruction Attacks* (DRA), *i.e.*, the ability of an adversary to infer a user's true input x from a single privatized report Y [2, 6]. In central differential privacy, *membership inference attacks* are a natural concern, as they determine whether a specific individual is part of a trusted central dataset D. Such an attack is inherently meaningless under LDP, since data remain under the control of each individual. Here, the primary privacy risk shifts to input-level inference: *can an attacker guess a user's original value from their privatized report?* DRA is therefore intrinsic to the LDP threat model and arguably the most direct and relevant form of privacy leakage in this setting. Moreover, successful data reconstruction can serve as a gateway for a variety of downstream privacy attacks, such as *attribute inference* (predicting unknown sensitive features), *linkage attacks* (matching users across datasets), or *broader profiling threats* that apply existing attacks on the reconstructed data. For all these reasons, DRA is a critical evaluation metric for assessing the robustness of LDP mechanisms.

However, selecting a protocol that balances utility and robustness against data reconstruction attacks is non-trivial in practice, especially when utility requirements are subjective, loosely defined, and highly case-dependent. While Big tech companies can afford in-house experts to tune protocols and privacy budgets carefully, **smaller organizations may lack practical guidance** on how to select an appropriate protocol and privacy parameter ε for their specific use case. Poorly chosen parameters or protocols can result in high utility loss or increased vulnerability to attacks, making the deployment of LDP challenging for practitioners. Indeed, selecting appropriate parameters for LDP protocols has no one-size-fits-all solution. It depends not only on the choice of ε, but also on the

*Equal contribution.

CCS '25, Taipei, Taiwan

ACM ISBN 979-8-4007-1525-9/2025/10
https://doi.org/10.1145/3719027.3760706



(a) Protocol-level trade-off visualization.

(b) Custom dataset analysis dashboard.

Figure 1: Overview of the LDP-Toolbox web interface. (a) The *Analytical Visualization* page enables users to compare privacy-utility-attackability trade-offs across multiple LDP protocols and data configurations. (b) The *Custom Upload* page allows users to upload their own datasets (*e.g.*, tabular, synthetic, location data, etc), experiment with different parameter settings, and visualize attackability, utility loss, and estimated distributions in an interactive dashboard.

specific LDP protocol and the nature of the underlying data. These factors make protocol selection and ε-tuning largely *ad-hoc* and empirically driven, with little theoretical or practical guidance that generalizes across tasks, datasets, or needs.

Related Work. A rich literature exists on LDP protocols for frequency estimation and other tasks, and several libraries implement these primitives in Python (*e.g.*, `pure-LDP` [3], `multi-freq-ldpy` [1]). However, these tools focus on protocol implementations only, without providing interactive interfaces for real-world data exploration, metric visualization, or parameter tuning. In the broader differential privacy space, recent efforts have explored visualization and interactive support for practitioners [9], but does not target LDP protocols or attackability metrics. To the best of our knowledge, no prior work combines interactive visualization, protocol selection, and attackability evaluation specifically for LDP. A practical toolbox is therefore needed to automatically benchmark utility and attackability trade-offs across protocols, parameters, and datasets, a gap which our demonstration paper fills.

Contributions. In this paper, we demonstrate LDP-Toolbox, the first benchmarking web-based system for LDP protocols that evaluates both utility and attackability in an integrated framework. The toolbox offers comprehensive benchmarking across multiple protocols and data configurations, providing practitioners with clear, visual insights into the trade-offs between privacy guarantees and practical performance. Designed for usability, it includes a ready-to-use, customizable data loader, flexible parameter tuning, and intuitive visualizations to support informed decision-making. The toolbox is easily installed via `pip install ldp-toolbox`, lowering the barrier to adoption. We demonstrate the functionality of LDP-Toolbox on two representative frequency estimation tasks using real-world discrete personal attribute data and location-based time-series data, showcasing its versatility across applications.

2 LDP-Toolbox

We develop LDP-Toolbox as an interactive interface via Python Dash, composed of two main components. The first is a benchmark visualization interface that provides a global overview of utility (MSE), attackability (vulnerability to reconstruction attacks), and their trade-offs across eight standard LDP protocols [2]: `GRR`, `OUE`, `SUE`, `OLH`, `BLH`, `SS`, `SHE`, and `THE`. The second interface allows users to upload their own dataset and perform protocol-specific analysis based on their selected utility metrics and privacy tolerance range.

2.1 Analytical Visualization

The objective of this interface is to provide a global overview of the behavior of various LDP protocols from a predefined set, with respect to utility, attackability, and the trade-offs between them. The visualization interface yields the plots shown in Figure 1a. Both utility loss, quantified by the expected mean squared error (MSE), and attackability, measured by the expected success rate of data reconstruction attacks (DRA) using optimal strategies, can be expressed analytically for each protocol [2]. These expressions are functions of three key parameters: domain size (k), number of users (n), and privacy budget (ε). We allow users to choose k, n, and a range of ε values, then plot how each protocol behaves as ε increases. This enables users to compare how quickly a protocol becomes more vulnerable (higher DRA) or loses utility (higher MSE) as privacy decreases (higher ε). The third graph visualizes the Pareto frontier between utility loss and attackability. Intuitively, a protocol demonstrates a better privacy-utility trade-off if it approaches the origin point (0,0), *i.e.*, showing minimal variation along both axes.

2.2 Custom Dataset Analysis

In the second interface, users can upload their own dataset and perform more flexible and interactive analyses in four steps. The layout is shown in Figure 1b, with each step numbered at the bottom.

Step ① Upload and Set Parameters. Users can upload their dataset in CSV format, with each row representing a data record and each column corresponding to an attribute. The attributes are automatically detected and populated in a dropdown menu, allowing the user to select which attribute they wish to perturb and estimate the histogram. The domain size k is also automatically

computed based on the data, but it can be customized by checking a box and manually entering a minimum and maximum value, overriding the default range if needed. An additional checkbox allows users to specify whether the dataset contains location data, in which case the interface will generate heatmaps in the output visualization. Users can also adjust the percentage of individuals to include if they wish to focus on a random subset of the data.

Next, users select the LDP protocols they wish to evaluate and define a privacy budget range (ε). Since it is often difficult for non-experts to choose an exact value for ε, the interface allows users to input a range instead. The system then evaluates three representative points within this range: the minimum, mean, and maximum. Once configured, clicking the "Compute" button launches the backend process to calculate the metrics. Unlike the *Analytical Visualization* page, these calculations are performed directly on the user-uploaded, customized dataset.

Step ② and ③ Visualize Benchmark Results. Once computation is complete, the **attackability** results for each selected protocol are displayed for the three privacy budget values within the user-defined ε range. These attackability plots show the **empirical** reconstruction rate, helping users assess which protocols are more vulnerable to data reconstruction under their specific dataset and parameter settings. For utility loss, users can choose from multiple histogram distance metrics depending on their needs: **MSE**, **RMSE**, **KL-divergence**, and **Kendall rank correlation**. Each utility metric highlights different aspects of distribution similarity, enabling practitioners to align the evaluation with the requirements of their target application. The results are presented in a format similar to the attackability plots, allowing side-by-side comparison of protocols and privacy budgets.

Step ④ Select and Compare Results. Based on the two diagrams, users can decide which protocol and which ε level best suit their privacy-utility trade-off requirements. After selecting a protocol and privacy level, the interface displays a comparison between the original and the estimated distributions of the uploaded data. This offers a concrete visualization of how the frequency estimation is affected by perturbation. For spatial datasets, a heatmap comparison is also shown to capture space effects better. Finally, users may iterate by refining protocol choices or parameter settings until the desired balance is achieved.

3 Demonstration

The audience will experience the demonstration through: (1) A **poster** outlining the motivation, the architecture of LDP-Toolbox, its usability-focused design, and the evaluation workflow for comparing privacy-utility trade-offs. (2) Attendees will interact with a **live system** to explore and compare different LDP mechanisms in practice. They will be able to visualize estimated distributions, utility loss, and attack risks using real-world examples. Specifically, we demonstrate the capabilities of LDP-Toolbox through the following representative frequency estimation use cases.

Tabular Data with Discrete Attributes. Tabular datasets containing discrete attributes, such as age, gender, or preferences, are widely collected and analyzed across many real-world applications, including web services, recommender systems, and social networks. A common objective is to release histograms or joint distributions of these attributes to support population-level insights, personalization, or policy decisions, while still preserving privacy. However, these attributes are often sensitive, and reconstruction attacks can recover individual records from noisy values [5]. Small noise often fails to prevent leakage in multi-attribute data, making it difficult for non-experts to balance privacy and utility via ε-selection. We illustrate this use case using a subset of the U.S. Census dataset, which includes 125,789 individuals and 14 attributes. With our toolbox, users can generate locally differentially private reports using protocols and parameter settings that are resistant to reconstruction attacks. They can then validate custom utility constraints by comparing the aggregated, estimated results to the original data, *e.g.*, ensuring that utility loss remains acceptable under metrics such as MSE, RMSE, KL-divergence, or Kendall's rank correlation.

Time-Series Location Data. Spatiotemporal location data are valuable for applications such as urban planning, transportation, and epidemiology, but pose significant privacy risks due to their fine-grained and highly correlated nature. Even when protected with differential privacy, location datasets can remain vulnerable to reconstruction attacks, which amplify threats such as trajectory recovery [8], re-identification [4], or semantic inference (*e.g.*, home/workplace inference). We illustrate this scenario using `YJMob100K` (https://zenodo.org/records/10142719), a real-world dataset capturing mobility traces in a 200×200 grid at 30-minute intervals for 100,000 individuals in a Japanese city. LDP-Toolbox allows users to upload such spatial data, apply LDP protocols, and visualize the impact of perturbation. As shown in Figure 1b, users can select noise mechanisms and privacy budgets that balance utility (*e.g.*, identifying top-visited locations or temporal peaks) with protection against reconstruction attacks. The system provides interpretable spatiotemporal *heatmaps* to compare original and estimated distributions for any given time slot, helping practitioners to assess how well patterns are preserved while maintaining privacy.

Acknowledgement. This work has been supported by the French National Research Agency (ANR): "ANR-24-CE23-6239" and "ANR-22-PECY-0002".

References

[1] Héber H Arcolezi, Jean-François Couchot, Sébastien Gambs, Catuscia Palamidessi, and Majid Zolfaghari. 2022. Multi-Freq-LDPy: multiple frequency estimation under local differential privacy in python. In *European Symposium on Research in Computer Security*. Springer, 770–775. doi:10.1007/978-3-031-17143-7_40

[2] Héber H Arcolezi and Sébastien Gambs. 2025. Revisiting Locally Differentially Private Protocols: Towards Better Trade-offs in Privacy, Utility, and Attack Resistance. *arXiv preprint arXiv:2503.01482* (2025).

[3] Graham Cormode, Samuel Maddock, and Carsten Maple. 2021. Frequency estimation under local differential privacy. *Proceedings of the VLDB Endowment* 14, 11 (July 2021), 2046–2058. doi:10.14778/3476249.3476261

[4] Yves-Alexandre De Montjoye, César A Hidalgo, Michel Verleysen, and Vincent D Blondel. 2013. Unique in the crowd: The privacy bounds of human mobility. *Scientific reports* 3, 1 (2013), 1376.

[5] Irit Dinur and Kobbi Nissim. 2003. Revealing information while preserving privacy. In *Proceedings of the twenty-second ACM PODS*. 202–210.

[6] M. Emre Gursoy, Ling Liu, Ka-Ho Chow, Stacey Truex, and Wenqi Wei. 2022. An Adversarial Approach to Protocol Analysis and Selection in Local Differential Privacy. *IEEE TIFS* 17 (2022), 1785–1799. doi:10.1109/TIFS.2022.3170242

[7] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. 2011. What Can We Learn Privately? *SIAM J. Comput.* 40, 3 (2011), 793–826. doi:10.1137/090756090

[8] Abhishek Kumar Mishra, Mathieu Cunche, and Heber H Arcolezi. 2025. Breaking Anonymity at Scale: Re-identifying the Trajectories of 100K Real Users in Japan. *arXiv preprint arXiv:2506.05611* (2025).

[9] Liudas Panavas, Saeyoung Rho, Hari Bhimaraju, Wynee Pintado, Rebecca N. Wright, and Rachel Cummings. 2024. A Visualization Tool to Help Technical Practitioners of Differential Privacy. TPDP 2024.