Revisiting Locally Differentially Private Protocols: Towards Better Trade-offs in Privacy, Utility, and Attack Resistance

Héber H. Arcolezi
Inria Centre at the University Grenoble Alpes
France
heber.hwang-arcolezi@inria.fr

Sébastien Gambs
Université du Québec à Montréal (UQAM)
Canada
gambs.sebastien@uqam.ca

ABSTRACT

Local Differential Privacy (LDP) has become foundational in safeguarding user privacy, especially in settings where the server collecting the data is untrusted. Designing LDP mechanisms that achieve an optimal trade-off between privacy, utility, and robustness to adversarial inference attacks remains a significant challenge. In this work, we introduce a general multi-objective optimization framework for refining LDP protocols, enabling the joint optimization of privacy and utility under various adversarial settings. While our framework is flexible enough to accommodate multiple privacy attacks, security vulnerability, and utility metrics, in this paper, we specifically optimize for Attacker Success Rate (ASR) under distinguishability attack as a measure of privacy and Mean Squared Error (MSE) as a measure of utility. We systematically revisit these trade-offs by analyzing eight state-of-the-art LDP protocols and proposing adaptive and refined counterparts that leverage tailored optimization techniques. Our results demonstrate that our proposed adaptive mechanisms consistently outperform their non-adaptive counterparts, reducing ASR by up to five orders of magnitude while maintaining competitive utility, with MSE increasing by no more than two orders of magnitude. Analytical derivations and empirical validations confirm the effectiveness of our mechanisms, moving them closer to the ASR-MSE Pareto frontier.

PVLDB Reference Format:
Héber H. Arcolezi and Sébastien Gambs. Revisiting Locally Differentially Private Protocols: Towards Better Trade-offs in Privacy, Utility, and Attack Resistance. PVLDB, 14(1): XXX-XXX, 2020.
doi:XX.XX/XXX.XX

PVLDB Artifact Availability:
The source code, data, and/or other artifacts have been made available at https://github.com/hharcolezi/LDP_protocols_refined.

1 INTRODUCTION

Differential Privacy (DP) [20] has become a gold standard for preserving privacy in data analytics and mining, in which the goal is to ensure that an individual's data does not significantly influence the output of any analysis. However, traditional DP relies on a trusted aggregator to apply the privacy mechanism, which is often impractical for decentralized or privacy-sensitive applications. This limitation has led to the rise of Local Differential Privacy (LDP) [27], a model that removes the need for a trusted aggregator by requiring users to perturb their data locally before sharing it with the server.

The local DP model has gained widespread adoption, with major technology companies integrating it into their systems to enhance user privacy. Notable examples include Google Chrome [22] and Windows 10 [18], in which LDP protocols have been used to collect statistics while safeguarding individual data. A fundamental task under LDP guarantees is **frequency estimation**, which forms the basis of many advanced data analysis tasks like heavy hitter estimation [9, 40], frequency monitoring [4, 7, 18, 22], multidimensional queries [3, 15, 24, 38], frequent item-set mining [28, 37], and spatial density estimation [2, 32].

Due to its importance, numerous LDP frequency estimation protocols have been proposed [30], namely, Generalized Randomized Response (GRR) [26], Subset Selection (SS) [33, 39], Symmetric Unary Encoding (SUE) [22], Optimized Unary Encoding (OUE) [34], Summation with Histogram Encoding (SHE) [20], Thresholding with Histogram Encoding (THE) [34], Binary Local Hashing (BLH) [9], and Optimal Local Hashing (OLH) [34]. These protocols mainly focus on improving utility, often quantified in terms of the variance (*i.e.*, Mean Squared Error – MSE), as well as optimizing computational and communication costs [1, 23] to enable efficient data collection in large-scale systems.

While utility and communication costs have been the traditional focus of LDP frequency estimation protocols, ***recent research has shed light on their vulnerabilities in adversarial settings***. Notably, studies have investigated many LDP mechanisms through the lens of distinguishability attacks [5, 6, 21], which allow adversaries to infer an individual's true value based on the perturbed data sent to the server. In addition, re-identification risks [6, 29] have been demonstrated, where adversaries can uniquely identify users within a dataset. Moreover, poisoning attacks [12] have emerged as a significant security concern, enabling adversaries to manipulate aggregated statistics by injecting maliciously crafted data. These threats underscore the pressing need for enhanced privacy protections against a diverse range of attacks targeting LDP mechanisms.

Contributions. In this work, ***we introduce a general multi-objective optimization framework for refining LDP frequency estimation protocols***, enabling the joint optimization of privacy and utility under various adversarial settings. While our framework is flexible enough to incorporate multiple objectives, in this paper, we focus on privacy attacks, quantified by Attacker Success Rate (ASR) under distinguishability attacks [5, 6, 21], and utility, measured via Mean Squared Error (MSE). Distinguishability attacks are particularly relevant as they directly challenge the fundamental goal of LDP–preventing an adversary from inferring a user's value

Proceedings of the VLDB Endowment, Vol. 14, No. 1 ISSN 2150-8097.
doi:XX.XX/XXX.XX

from the obfuscated output. Meanwhile, MSE serves as a widely adopted utility metric in LDP literature [14, 34, 35] due to its analytical tractability and connection to other estimation measures, such as Mean Absolute Error (MAE) and Fisher Information [8, 26]. We further justify the reason for these metrics in Section 5.2.

To demonstrate the practical benefits of our framework, **we refine four state-of-the-art LDP protocols** (SS, OUE, OLH, and THE), proposing adaptive versions, namely, ASS, AUE, ALH, and ATHE, which achieve superior trade-off between privacy and utility compared to their traditional counterparts. Additionally, we derive the analytical ASR for three additional LDP protocols (SHE, THE, and a generic Unary Encoding protocol), extending prior analyses [5, 21]. Our adaptive protocols offer enhanced robustness against privacy attacks while maintaining practical utility levels, demonstrating the effectiveness of our framework in optimizing LDP mechanisms beyond conventional single-objective approaches. In summary, the main contributions of this paper are:

- We introduce a **general multi-objective optimization framework** that enables the joint optimization of privacy and utility in LDP frequency estimation. Our approach is flexible and can incorporate multiple privacy, security, and utility metrics; in this paper, we focus on minimizing ASR and MSE to achieve a comprehensive trade-off between privacy guarantees and data utility.
- We **extend four state-of-the-art LDP frequency estimation protocols**—SS, OUE, OLH, and THE—by proposing adaptive versions named ASS, AUE, ALH, and ATHE. These adaptive protocols offer a significantly better trade-off between robustness to privacy attacks and utility compared to their traditional counterparts (*e.g.*, see Figure 4).
- We derive the **analytical closed-form equation** of the expected ASR for three LDP protocols beyond what was presented in previous works [21]. These derivations are critical for evaluating and optimizing the privacy guarantees of LDP protocols under adversarial inference scenarios.
- We validate our proposed adaptive protocols **through extensive experiments**, demonstrating their effectiveness in achieving a more favorable balance between privacy (ASR) and utility (MSE) compared to existing protocols, under the same ε-LDP guarantees. Our results indicate that the adaptive protocols can substantially reduce adversarial success rates while maintaining competitive estimation accuracy.

Outline. The remainder of this paper is structured as follows. First, Section 2 reviews the most relevant related work, providing context for our contributions. Next, Section 3 defines the problem, introduces the LDP privacy model, and describes the adversarial model used in this study. Section 4 provides an overview of the eight LDP frequency estimation protocols analyzed in this work, along with their attack models. Section 5 then presents our multi-objective framework and the proposed adaptive LDP protocols that refine the original methods to enhance privacy and utility. Afterward, Section 6 details the experiments we conducted and presents a comprehensive analysis of the results. Finally, Section 7 concludes the paper and discusses potential directions for future research.

2 RELATED WORK

Frequency estimation is a primary objective of LDP, serving as a building block for a wide range of advanced applications, such as heavy hitter estimation [9, 40], frequency monitoring [4, 7, 18, 22], multidimensional queries [3, 15, 24, 38], frequent item-set mining [28, 37], and spatial density estimation [2, 32]. Numerous LDP frequency estimation protocols have been proposed in the literature [1, 14, 23, 26, 33, 34, 39], primarily focusing on minimizing estimation error, computational cost, or communication overhead.

However, recent studies started to examine LDP frequency estimation mechanisms from an adversarial perspective such as distinguishability attacks [5, 6, 21], re-identification risks [6, 29], and poisoning attacks [12]. This work focuses on distinguishability attacks, which allow adversaries to predict the user's input based on the observed obfuscated output. However, unlike these prior studies that primarily highlight vulnerabilities, we propose a systematic methodology to refine existing protocols, thus reducing their susceptibility to known and emerging privacy threats.

Specifically, our work differs from the existing distinguishability attack literature [5, 6, 21] in the following aspects. First, we formally analyze three LDP protocols' expected ASR beyond the ones in [21]. Second, we extensively analyzed the privacy, utility, and robustness against privacy attacks of eight state-of-the-art LDP protocols. Third, we formulate a multi-objective optimization problem for LDP frequency estimation protocols instead of the single-objective one (*i.e.*, utility-driven). This allowed us to propose four new adaptive and refined LDP protocols, which provide better trade-offs in terms of privacy, utility, and robustness against adversarial attacks.

3 BACKGROUND

Notation. We use italic uppercase letters (*e.g.*, U) to denote sets, and write $[n] = \{1, \ldots, n\}$ to represent a set of n positive integers. Vectors are denoted by bold lowercase letters (*e.g.*, $\mathbf{x}$), where $\mathbf{x}_i$ represents the value of the i-th coordinate of $\mathbf{x}$. Finally, randomized mechanisms are denoted by $\mathcal{M}$, the input domain is denoted by $\mathcal{X}$, and the output domain by $\mathcal{Y}$. Both $\mathcal{X}$ and $\mathcal{Y}$ are discrete, where $|\mathcal{X}| = k$ and $|\mathcal{Y}|$ depends on the randomized mechanism $\mathcal{M}$.

3.1 Problem Statement

We consider an untrusted server collecting data from a distributed group of users while preserving their privacy. Formally, there are n users, and each user holds a value x from a discrete domain $\mathcal{X} = \{1, 2, \ldots, k\}$. The task is **frequency estimation**, *i.e.*, the server aims to learn the frequencies of each value across all users, denoted as $\mathbf{f} = \{f_i\}_{i \in [k]}$. More precisely:

- **Users' goal.** Each user wants to protect their privacy. To achieve this, users apply an obfuscation mechanism $\mathcal{M}$ that perturbs their value x before sending it to the server.
- **Server's goal.** The server aims to estimate the frequency distribution $\mathbf{f}$ of the values held by all users while minimizing the estimation error. After receiving the obfuscated values from all n users, the server estimates a k-bins histogram $\hat{\mathbf{f}} = \{\hat{f}_i\}_{i \in [k]}$, representing the estimated frequencies.
- **Adversary's goal.** The adversary aims to accurately infer each user's true value $x \in \mathcal{X}$ based on the obfuscated output $y \in \mathcal{Y}$ sent to the server (*i.e.*, "value inference attack").

3.2 Local Differential Privacy

Local Differential Privacy (LDP) [27] ensures that the output of a randomized mechanism does not significantly reveal information about the input. Formally:

Definition 1 (ε-Local Differential Privacy). *An algorithm $\mathcal{M}$ satisfies ε-local differential privacy (ε-LDP), where $\varepsilon \geq 0$, if and only if for any two distinct inputs $x, x' \in \mathcal{X}$, we have:*

$$\forall y \in Range(\mathcal{M}) : \Pr[\mathcal{M}(x) = y] \leq e^{\varepsilon} \cdot \Pr[\mathcal{M}(x') = y], \quad (1)$$

where $Range(\mathcal{M})$ denotes the set of all possible outputs of $\mathcal{M}$.

Smaller values of ε indicate stronger privacy guarantees, as it limits how much more likely the output is for one input compared to another given the same observed value. In other words, ε controls the level of indistinguishability between inputs x and x', providing a formal measure of privacy.

3.3 Pure LDP Framework

We consider the pure LDP framework proposed by Wang *et al.* [34] to analyze LDP frequency estimation protocols. Formally, an LDP protocol is called pure if it satisfies the following definition:

Definition 2 (Pure LDP Protocols [34]). *A protocol is considered pure if and only if there exist two probability values, $p^* > q^*$, in its perturbation mechanism $\mathcal{M}$, such that for all inputs $x \in \mathcal{X}$:*

$$\Pr[\mathcal{M}(x) \in \{y \mid x \in Support(y)\}] = p^*$$
$$\forall_{x' \neq x} \Pr[\mathcal{M}(x') \in \{y \mid x \in Support(y)\}] = q^*$$

where the set $\{y \mid x \in Support(y)\}$ includes all possible outputs y that "support" the input value x.

Thus, p^* represents the probability that the input x is mapped to an output supporting it, whereas q^* represents the probability that any other input $x' \neq x$ is mapped to an output that supports x. Given n users, let y^j denote the obfuscated value of user $j \in [n]$, the frequency estimate $\hat{f}_i$ of the input value $i \in [k]$ is calculated as:

$$\hat{f}_i = \frac{\sum_{j=1}^{n} \mathbb{1}_{\text{Support}(y^j)}(i) - nq^*}{n(p^* - q^*)}. \quad (2)$$

As shown in [34], the estimator in Equation (2) is unbiased (*i.e.*, $\mathbb{E}[\hat{f}_i] = f_i$) and the variance of the estimation $\hat{f}_i$ is:

$$\text{Var}[\hat{f}_i] = \frac{q^*(1 - q^*)}{n(p^* - q^*)^2} + \frac{f_i(1 - p^* - q^*)}{n(p^* - q^*)}. \quad (3)$$

With a sufficiently large domain size and no dominant frequency f_i, the second term of Equation (3) can be ignored. Thus, as commonly used in the LDP literature [4, 34, 38], we will consider the approximate variance, which is given by:

$$\text{Var}[\hat{f}_i] = \frac{q^*(1 - q^*)}{n(p^* - q^*)^2}. \quad (4)$$

Furthermore, as the estimation is unbiased, we will interchangeably refer to the variance as the Mean Squared Error (MSE):

$$\text{MSE} = \frac{1}{k}\sum_{i=0}^{k} \mathbb{E}\left[\left(\hat{\mathbf{f}}_i - \mathbf{f}_i\right)^2\right] = \frac{1}{k}\sum_{i=0}^{k} \text{Var}[\hat{\mathbf{f}}_i].$$

3.4 Adversarial Model for LDP

The fundamental premise of ε-LDP, as stated in Equation (1), is that the input to $\mathcal{M}$ cannot be confidently determined from its output, with the level of confidence determined by e^{ε}. Therefore, the user's privacy is considered compromised if an adversary $\mathcal{A}$ can successfully infer the user's original input $x \in \mathcal{X}$ from the obfuscated output $y \in \mathcal{Y}$. While various adversarial models have been proposed to exploit vulnerabilities in LDP mechanisms, including re-identification [6, 29] and poisoning [12] attacks, in this paper, and without loss of generality, we focus on distinguishability attacks [5, 6, 21]. These attacks provide a fundamental measure of privacy leakage in LDP, as they directly **quantify how well an adversary can infer the true input from the obfuscated response**. Formally, the adversary's prediction $\hat{x}$ can be defined as:

$$\hat{x} = \arg\max_{x \in \mathcal{X}} \Pr[x \mid y].$$

By applying Bayes' theorem, we can rewrite the expression as:

$$\hat{x} = \arg\max_{x \in \mathcal{X}} \frac{\Pr[y \mid x] \cdot \Pr[x]}{\Pr[y]}.$$

Assuming a uniform prior distribution over x (*i.e.*, $\Pr[x] = \frac{1}{k}$, where $k = |\mathcal{X}|$), and noting that $\Pr[y]$ is constant for a given observation, the expression simplifies to:

$$\hat{x} = \arg\max_{x \in \mathcal{X}} \Pr[y \mid x].$$

To evaluate the accuracy of the attack, we will use the **Attacker Success Rate (ASR)** metric, which represents the probability that the adversary's prediction $\hat{x}$ matches the input value x:

$$\text{ASR} = \Pr[\hat{x} = x] = \frac{1}{n} \cdot \sum_{j=1}^{n} \mathbb{1}\left(\hat{x}^j = x^j\right).$$

Mathematically, it is also possible to derive the **expected ASR** for each protocol through formal expected value analysis [21]:

$$\mathbb{E}[\text{ASR}] = \mathbb{E}[\Pr[\hat{x} = x]].$$

With ASR, we can assess LDP protocols' effectiveness in protecting privacy, uncovering vulnerabilities not evident from ε alone.

4 LDP FREQUENCY ESTIMATION PROTOCOLS

In this section, we provide a concise overview of eight state-of-the-art LDP frequency estimation protocols. We systematically describe each mechanism based on three key functions: **Encoding**, **Perturbation**, and **Aggregation** which collectively define their operation. Additionally, we introduce an **Attacking** function for each protocol, highlighting potential vulnerabilities to privacy attacks. For protocols where the expected ASR is not available in the literature [21], we derive it and present the detailed analyses in Appendix A.

4.1 Generalized Randomized Response (GRR)

GRR [26] extends the randomized response method proposed by Warner [36] to a domain size of $k \geq 2$, while ensuring ε-LDP.

Encoding. In GRR, $\texttt{Encode}(x) = x$ and $x \in [k]$.

Perturbation. The perturbation function of GRR is given by:

$$\Pr[\text{GRR}(x) = y] = \begin{cases} p = \frac{e^{\varepsilon}}{e^{\varepsilon}+k-1} & \text{if } y = x, \\ q = \frac{1}{e^{\varepsilon}+k-1} & \text{if } y \neq x, \end{cases} \quad (5)$$

where $y \in [k]$ is the perturbed value sent to the server.

Aggregation. In GRR, each output value i supports the corresponding input i, resulting in the support set $\mathbb{1}_{\mathrm{GRR}} = y$. GRR is a pure protocol with $p^* = p$ and $q^* = q$. The server estimates the frequency using Equation (2), with the following analytical MSE:

$$\mathrm{MSE}_{\mathrm{GRR}} = \frac{e^\varepsilon + k - 2}{(e^\varepsilon - 1)^2}. \tag{6}$$

Attacking. From Equation (5), it follows that $\Pr[y = x] > \Pr[y = x']$ for all $x' \in \mathcal{X} \setminus \{x\}$. Thus, the optimal attack strategy for GRR is to predict $\hat{x} = \mathcal{A}_{\mathrm{GRR}}(y) = y$. The expected ASR for GRR is given by:

$$\mathbb{E}[\mathrm{ASR}]_{\mathrm{GRR}} = \frac{e^\varepsilon}{e^\varepsilon + k - 1}. \tag{7}$$

4.2 Subset Selection (SS)

SS [33, 39] outputs a randomly selected subset $\mathbf{y}$ of size ω from the original domain $\mathcal{X}$. SS can be seen as a generalization and optimization of GRR, where SS is equivalent to GRR when $\omega = 1$.

Encoding and Perturbation. Starting with an empty subset $\mathrm{y} = \emptyset$, the true value x is added to y with probability: $p = \frac{\omega e^\varepsilon}{\omega e^\varepsilon + k - \omega}$. Finally, values are added to y as follows:

- If $x \in \mathbf{y}$, then $\omega - 1$ values are sampled from $\mathcal{X} \setminus \{x\}$ uniformly at random (without replacement) and are added to y;
- If $x \notin \mathbf{y}$, then ω values are sampled from $\mathcal{X} \setminus \{x\}$ uniformly at random (without replacement) and are added to y,

The user then sends the subset y to the server.

Aggregation. In SS, each value i in the output subset y supports the corresponding input value i. Thus, the support set for SS is $\mathbb{1}_{\mathrm{SS}} = \{x \mid x \in \mathrm{y}\}$. This protocol is pure, with: $p^* = p = \frac{\omega e^\varepsilon}{\omega e^\varepsilon + k - \omega}$ and $q^* = \frac{\omega e^\varepsilon(\omega - 1) + (k - \omega)\omega}{(k-1)(\omega e^\varepsilon + k - \omega)}$. The server estimates the frequency using Equation (2), with the following analytical MSE:

$$\mathrm{MSE}_{\mathrm{SS}} = \frac{\begin{array}{c}(k - \omega + (\omega - 1)e^\varepsilon)\,(-\omega(k - \omega) - \omega(\omega - 1)e^\varepsilon \\ +(k-1)(k + 2\omega e^\varepsilon - \omega))\end{array}}{n\omega(-k + \omega + (k-1)e^\varepsilon - (\omega - 1)e^\varepsilon)^2}. \tag{8}$$

The optimal subset size that minimizes the MSE of SS in Equation (8) is $\omega = \max\left(1, \left\lfloor \frac{k}{e^\varepsilon + 1} \right\rceil\right)$ [33, 39].

Attacking. With the support set of each user's report $\mathbb{1}_{\mathrm{SS}}$, the optimal attack strategy $\mathcal{A}_{\mathrm{SS}}$ is to predict $\hat{x} = \mathrm{Uniform}\,(\mathbb{1}_{\mathrm{SS}})$ [6, 21]. The expected ASR for SS is given by [21]:

$$\mathbb{E}[\mathrm{ASR}]_{\mathrm{SS}} = \frac{e^\varepsilon}{\omega e^\varepsilon + k - \omega}. \tag{9}$$

4.3 Unary Encoding (UE)

UE protocols [22, 34] encode the user's input $x \in \mathcal{X}$ as a k-dimensional one-hot vector $\mathbf{x}$, where each bit is subsequently obfuscated independently.

Encoding. $\mathsf{Encode}(x) = [0, \ldots, 0, 1, 0, \ldots, 0]$ is a binary vector with a single 1 at position x and all other positions set to 0.

Perturbation. The obfuscation function of UE mechanisms randomizes the bits from $\mathbf{x}$ independently to generate $\mathbf{y}$ as follows:

$$\forall i \in [k]: \quad \Pr[\mathbf{y}_i = 1] = \begin{cases} p, & \text{if } \mathbf{x}_i = 1, \\ q, & \text{if } \mathbf{x}_i = 0, \end{cases} \tag{10}$$

where $\mathbf{y}$ is sent to the server. There are two variations of UE mechanisms: (i) Symmetric UE (SUE) [22] that selects $p = \frac{e^{\varepsilon/2}}{e^{\varepsilon/2}+1}$ and $q = \frac{1}{e^{\varepsilon/2}+1}$; and (ii) Optimized UE (OUE) [34] that selects $p = \frac{1}{2}$ and $q = \frac{1}{e^\varepsilon + 1}$ to minimize the MSE in Equation (11) below.

Aggregation. A reported bit vector $\mathbf{y}$ is considered to support an input i if $\mathbf{y}_i = 1$. Therefore, the support set for UE protocols is defined as $\mathbb{1}_{\mathrm{UE}} = \{i | \mathbf{y}_i = 1\}$. UE protocols are pure with $p^* = p$ and $q^* = q$. The server estimates the frequency using Equation (2), with the corresponding MSE given by:

$$\mathrm{MSE}_{\mathrm{UE}} = \frac{((e^\varepsilon - 1)\,q + 1)^2}{n\,(e^\varepsilon - 1)^2\,(1 - q)q}. \tag{11}$$

Attacking. Given the support set of each user's report, $\mathbb{1}_{\mathrm{UE}}$, the adversary can adopt two possible attack strategies $\mathcal{A}_{\mathrm{UE}}$ [6, 21]:

- $\mathcal{A}^0_{\mathrm{UE}}$ is a random choice $\hat{x} = \mathrm{Uniform}\,([k])$, if $\mathbb{1}_{\mathrm{UE}} = \emptyset$;
- $\mathcal{A}^1_{\mathrm{UE}}$ is a random choice $\hat{x} = \mathrm{Uniform}\,(\mathbb{1}_{\mathrm{UE}})$, otherwise.

In this paper, we generalized the expected ASR of SUE and OUE given in [21] for any UE protocol as:

$$\mathbb{E}[\mathrm{ASR}]_{\mathrm{UE}} = (1 - p) \cdot (1 - q)^{k-1} \cdot \frac{1}{k} + \sum_{m=1}^{k} p \cdot \frac{1}{m} \cdot \binom{k-1}{m-1} q^{m-1} (1 - q)^{(k-1)-(m-1)}. \tag{12}$$

We defer the derivation of Equation 12 to Appendix A.1.

4.4 Local Hashing (LH)

LH protocols [9, 34] use hash functions to map the input data $x \in \mathcal{X}$ to a new domain $[g]$, and then obfuscates the hash value with GRR. Let $\mathscr{H}$ be a universal hash function family such that each hash function $\mathrm{H} \in \mathscr{H}$ hashes a value $x \in \mathcal{X}$ into $[g]$ (*i.e.*, $\mathrm{H} : [k] \to [g]$).

Encoding. $\mathsf{Encode}(x) = \langle \mathrm{H}, h \rangle$, where $\mathrm{H} \in \mathscr{H}$ is chosen uniformly at random, and $h = \mathrm{H}(x)$. There are two variations of LH mechanisms: (i) Binary LH (BLH) [9] that just sets $g = 2$, and (ii) Optimized LH (OLH) [34] that selects $g = \lfloor e^\varepsilon + 1 \rceil$ to minimize the MSE in Equation (14) below.

Perturbation. LH protocols perturb $\langle \mathrm{H}, h \rangle$ into $\langle \mathrm{H}, y \rangle$, just like GRR, as follows:

$$\forall i \in [g],\ \Pr[y = i] = \begin{cases} p = \frac{e^\varepsilon}{e^\varepsilon + g - 1}, & \text{if } h = i, \\ q = \frac{1}{e^\varepsilon + g - 1}, & \text{if } h \neq i. \end{cases} \tag{13}$$

Aggregation. For each reported tuple $\langle \mathrm{H}, y \rangle$, the support set for LH protocols consists of all values $x \in \mathcal{X}$ that hash to y, denoted as $\mathbb{1}_{\mathrm{LH}} = \{x \mid \mathrm{H}(x) = y\}$. LH protocols are pure with $p^* = p$ and $q^* = \frac{1}{g}$. The server estimates the frequency using Equation (2) with the following analytical MSE:

$$\mathrm{MSE}_{\mathrm{LH}} = \frac{(e^\varepsilon - 1 + g)^2}{n(e^\varepsilon - 1)^2(g - 1)}. \tag{14}$$

Attacking. Based on the support set of each user's report, $\mathbb{1}_{\mathrm{LH}}$, the adversary can employ one of two possible attack strategies, denoted by $\mathcal{A}_{\mathrm{LH}}$ [6, 21]:

- $\mathcal{A}^0_{\mathrm{LH}}$ is a random choice $\hat{x} = \mathrm{Uniform}\,([k])$, if $\mathbb{1}_{\mathrm{LH}} = \emptyset$;
- $\mathcal{A}^1_{\mathrm{LH}}$ is a random choice $\hat{x} = \mathrm{Uniform}\,(\mathbb{1}_{\mathrm{LH}})$, otherwise.

The expected ASR of LH protocols is given by [21]:

$$\mathbb{E}[\text{ASR}]_{\text{LH}} = \frac{e^{\varepsilon}}{(e^{\varepsilon} + g - 1) \cdot \max\left\{\frac{k}{g}, 1\right\}}. \tag{15}$$

4.5 Histogram Encoding (HE)

HE protocols [34] encode the user's input data $x \in \mathcal{X}$, as a one-hot k-dimensional histogram before obfuscating each bit independently.

Encoding. $\mathtt{Encode}(x) = [0.0, 0.0, \ldots, 1.0, 0.0, \ldots, 0.0]$ in which only the x-th component is 1.0. Two different input values $x, x' \in \mathcal{X}$ will result in two vectors with L1 distance of $\Delta_1 = 2$.

Perturbation. The perturbation function $\mathtt{Perturb}(\mathbf{x})$ generates the output vector $\mathbf{y}$, where each component is given by $\mathbf{y}_i = \mathbf{x}_i + \text{Lap}\left(\frac{\Delta_1}{\varepsilon}\right)$, with $\text{Lap}(\cdot)$ representing the Laplace mechanism [20].

The following subsections describe two HE-based mechanisms: Summation with HE (SHE) and Thresholding with HE (THE). These mechanisms differ in their aggregation and attack strategies.

4.5.1 *Summation with HE (SHE).*

With SHE, there is no post-processing of $\mathbf{y}$ at the server-side.

Aggregation. Since Laplace noise with mean 0 is added to each vector independently, the server estimates the frequency using the sum of the noisy reports: $\hat{\mathbf{f}} = \left\{\sum_{j=1}^{n} \mathbf{y}_i^j\right\}_{i \in [k]}$. This aggregation method for SHE does not provide a support set and is not pure. The analytical MSE of this estimation is:

$$\text{MSE}_{\text{SHE}} = \frac{8}{n\varepsilon^2}. \tag{16}$$

Attacking. The optimal attack strategy for SHE is to predict the user's value by selecting the index corresponding to the maximum component of the obfuscated vector: $\hat{x} = \underset{i \in [k]}{\text{argmax}}\ y_i$ [5]. Following this attack strategy, we deduce the expected ASR for the SHE protocol as the probability that the noisy value y_x at the true index exceeds all other noisy values y_i for $i \neq x$:

$$\mathbb{E}[\text{ASR}]_{\text{SHE}} = \Pr\left[y_x > \max_{i \neq x} y_i\right]. \tag{17}$$

We defer the derivation of Equation (17) to Appendix A.2.

4.5.2 *Thresholding with HE (THE).*

In THE, each perturbed component of the vector $\mathbf{y}$ is compared to a threshold value θ to generate the final output vector. More precisely:

$$\forall i \in [k]: \quad \mathbf{y}_i = \begin{cases} 1, & \text{if } \mathbf{y}_i > \theta \\ 0, & \text{if } \mathbf{y}_i \leq \theta \end{cases}$$

Thus, the resulting output vector $\mathbf{y}$ is a binary vector in $\{0, 1\}^k$, where we have the following probabilities:

$$p = \Pr[\mathbf{y}_i = 1 \mid \mathbf{x}_i = 1] = 1 - \frac{1}{2} e^{\frac{\varepsilon}{2}(\theta - 1)}. \tag{18}$$

$$q = \Pr[\mathbf{y}_i = 1 \mid \mathbf{x}_i = 0] = \frac{1}{2} e^{-\frac{\varepsilon}{2}\theta}. \tag{19}$$

Aggregation. A reported bit vector $\mathbf{y}$ is viewed as supporting an input i if $\mathbf{y}_i > \theta$. Therefore, the support set for THE is $\mathbb{1}_{\text{THE}} = \{i \mid \mathbf{y}_i > \theta\}$. The THE mechanism is pure with $p^* = p$ and $q^* = q$. The server estimates the frequency using Equation (2) with the following analytical MSE:

$$\text{MSE}_{\text{THE}} = \frac{2e^{\varepsilon\theta/2} - 1}{(1 + e^{\varepsilon(\theta - 1/2)} - 2e^{\varepsilon\theta/2})^2}. \tag{20}$$

The optimal threshold value that minimizes the protocol's MSE in Equation (20) is within $\theta \in (0.5, 1)$ [34].

Attacking. Based on the support set $\mathbb{1}_{\text{THE}}$, the adversary can use one of two attack strategies denoted by $\mathcal{A}_{\text{THE}}$ [5]:

- $\mathcal{A}^0_{\text{THE}}$ is a random choice $\hat{x} = \text{Uniform}([k])$, if $\mathbb{1}_{\text{THE}} = \emptyset$;
- $\mathcal{A}^1_{\text{THE}}$ is a random choice $\hat{x} = \text{Uniform}(\mathbb{1}_{\text{THE}})$, otherwise.

In this paper, we obtained the expected ASR for THE as:

$$\mathbb{E}[\text{ASR}]_{\text{THE}} = (1-p)(1-q)^{k-1} \cdot \frac{1}{k} + \sum_{m=1}^{k} \binom{k-1}{m-1} p(q)^{m-1}(1-q)^{k-m} \cdot \frac{1}{m}. \tag{21}$$

We defer the derivation of Equation (21) to Appendix A.3.

5 REFINING LDP PROTOCOLS

The existing literature has traditionally proposed mechanisms like SS, OUE, OLH, and THE, focusing solely on minimizing MSE for a given privacy budget ε. However, these protocols exhibit varying vulnerabilities to privacy [5, 21] and security attacks [12, 13] under the same ε value, leaving room for improvement in both privacy guarantees and estimation accuracy. To address this, we introduce a **general multi-objective optimization framework** for refining LDP frequency estimation protocols, enabling adaptive parameter tuning based on multiple privacy and utility considerations.

5.1 Multi-Objective Optimization Framework

We define our general optimization objective for refining LDP protocols as:

$$\begin{aligned} \min_{\Theta} \quad & J(\Theta) = w_{\text{priv-atk}} \cdot \mathcal{P} + w_{\text{utility}} \cdot \mathcal{U} + w_{\text{security}} \cdot \mathcal{S}, \\ s.t. \quad & w_{\text{priv-atk}} + w_{\text{utility}} + w_{\text{security}} = 1. \end{aligned} \tag{22}$$

where $w_{\text{priv-atk}}, w_{\text{utility}}, w_{\text{security}}$, and w_{comm} are user-defined weights reflecting the relative importance of privacy attack, utility, and security, respectively. The terms $\mathcal{P}, \mathcal{U}, \mathcal{S}$, and C represent **general privacy attack, utility, and security vulnerability metrics**, respectively. For instance, re-identification risk [6, 29] or inference attacks [7] can serve as a privacy attack metric ($\mathcal{P}$). Fisher information [8] or mean squared/absolute error [26] can be used measure for utility ($\mathcal{U}$). For security vulnerability ($\mathcal{S}$), we can consider resilience to poisoning [12] or manipulation [13] attacks, where an adversary injects fake users to manipulate aggregated statistics. Finally, communication cost, such as the number of transmitted bits per user, can also be incorporated in Equation (22) as a measure of bandwidth overhead in LDP mechanisms.

5.2 Specialization to ASR and MSE

While our proposed framework in Section 5.1 is flexible, in this work, we focus specifically on **ASR under distinguishability attack for privacy and MSE for utility**. These metrics are widely used in the LDP literature [6, 21, 34] and have closed-form expressions (see Section 4), making them analytically tractable for optimization.

As described in Section 3.4, ASR captures an adversary's ability to infer the true input from the obfuscated response, which can **directly impact other privacy threats** such as re-identification risks [6, 29] and inference attacks in iterative data collections [7, 25]. Meanwhile, as described in Section 3.3, variance (*i.e.*, MSE) measures the expected squared deviation from the true value, making it a fundamental utility metric in frequency estimation [14, 34, 35]. MSE serves as a central utility metric because it not only quantifies estimation accuracy but also **directly influences other key statistical measures**, such as:

- **Mean Absolute Error (MAE).** Under the Central Limit Theorem (large n), estimation errors follow an approximately Gaussian distribution, leading to:

  $$\mathrm{MAE} \approx \sqrt{\tfrac{2}{\pi}} \cdot \sqrt{\mathrm{Var}(\hat{f}_i)}.$$

  Thus, minimizing variance/MSE inherently reduces MAE.
- **Fisher Information.** Fisher information quantifies the amount of information an observable variable carries about an unknown parameter. A key connection between Fisher information and variance is given by the Cramér–Rao bound [16, 17, 31]: for any unbiased estimator,

  $$I \ \geq \ \frac{1}{\mathrm{Var}(\hat{f}_i)},$$

  where I is the Fisher information. Since variance is equivalent to MSE for an unbiased estimator, maximizing Fisher information corresponds to minimizing MSE, thereby making it a fundamental objective in optimizing statistical accuracy.

ASR-MSE Optimization. To explicitly define this two-objective formulation, we rewrite Equation (22) as:

$$\begin{aligned} \min_{\Theta} \quad & J(\Theta) = w_{\mathrm{ASR}} \cdot \mathbb{E}[\mathrm{ASR}] + w_{\mathrm{MSE}} \cdot \mathrm{MSE}, \\ \textit{s.t.} \quad & w_{\mathrm{ASR}} + w_{\mathrm{MSE}} = 1. \end{aligned} \tag{23}$$

The weight parameters w_{ASR} and w_{MSE} in Equation (23) allow for adaptive trade-offs between privacy and utility, depending on application-specific requirements. For instance, in scenarios where protecting individuals' input values is critical, assigning a higher weight to w_{ASR} prioritizes privacy over estimation accuracy. Conversely, when precise analytics is the primary goal, increasing w_{MSE} ensures minimal degradation in utility. Notably, setting $w_{\mathrm{MSE}} = 1$ and $w_{\mathrm{ASR}} = 0$ recovers the original protocol parameters, demonstrating that ***our adaptive versions serve as extensions rather than replacements of existing mechanisms***. This flexibility enables LDP deployments to be tailored to different risk-utility trade-offs, enhancing applicability across diverse real-world scenarios.

5.3 Adaptive Parameter Optimization for ASR-MSE Trade-offs in LDP Protocols

Using our two-objective framework defined in Equation (23), we extend four state-of-the-art LDP protocols to introduce adaptive counterparts. Each adaptive protocol selects an optimal parameter Θ to achieve a better trade-off between ASR and MSE, rather than focusing solely on utility. The optimization process varies across protocols, adapting their internal parameters to balance privacy protection and estimation accuracy. We now describe each adaptive mechanism in detail.

5.3.1 *Adaptive Subset Selection (ASS)*.

The ASS mechanism extends the SS protocol. Unlike SS, which aims to minimize the estimation error alone when selecting ω, ASS jointly optimizes it for both MSE and ASR. Formally, the optimization problem for ASS is defined as:

$$\begin{aligned} \min_{\omega \in \mathbb{Z}} \quad & J(\omega) = w_{\mathrm{ASR}} \cdot \left(\frac{\omega e^{\varepsilon}}{\omega e^{\varepsilon} + k - \omega} \right) \\ & \quad + w_{\mathrm{MSE}} \cdot \left(\frac{\begin{gathered}(k - \omega + (\omega - 1)e^{\varepsilon}) \left(-\omega(k - \omega) - \omega(\omega - 1)e^{\varepsilon}\right. \\ \left. +(k-1)(k + 2\omega e^{\varepsilon} - \omega)\right)\end{gathered}}{n\omega(-k + \omega + (k-1)e^{\varepsilon} - (\omega - 1)e^{\varepsilon})^2} \right), \\ \textit{s.t.} \quad & 1 \leq \omega < k, \ w_{\mathrm{ASR}} + w_{\mathrm{MSE}} = 1. \end{aligned} \tag{24}$$

The range $[1 \leq \omega < k]$ ensures that at least one value is selected while keeping the subset smaller than the total domain size. This prevents trivial cases where $\omega = k$ would result in full randomness in the response.

5.3.2 *Adaptive Unary Encoding (AUE)*.

The AUE mechanism is a generalization of UE protocols. Unlike the optimized UE protocol (*i.e.*, OUE [34]), which only aims to minimize the estimation error setting a fixed $p = 1/2$ and $q = \frac{1}{e^{\varepsilon}+1}$, AUE jointly optimizes both MSE and ASR by adapting the probabilities p and q. Formally, the optimization problem for AUE is defined as:

$$\begin{aligned} \min_{p \in \mathbb{R}} \quad & J(p) = w_{\mathrm{MSE}} \cdot \left(\frac{\left(\left(e^{\varepsilon} - 1\right) q + 1\right)^2}{n \left(e^{\varepsilon} - 1\right)^2 (1 - q) q} \right) \\ & \quad + w_{\mathrm{ASR}} \cdot \left((1-p)(1-q)^{k-1} \cdot \frac{1}{k} \right. \\ & \quad \left. + \sum_{m=1}^{k} p \cdot \frac{1}{m} \cdot \binom{k-1}{m-1} q^{m-1} (1-q)^{(k-1)-(m-1)} \right), \\ \textit{s.t.} \quad & 0.5 \leq p < 1, \ q = \frac{p}{e^{\varepsilon}(1-p) + p}, \ w_{\mathrm{ASR}} + w_{\mathrm{MSE}} = 1. \end{aligned} \tag{25}$$

The equality constraint for q is due to the ε-LDP requirement for UE protocols: $\varepsilon = \ln\left(\frac{p(1-q)}{(1-p)q}\right)$ [22, 34].

5.3.3 *Adaptive LH (ALH)*.

The ALH mechanism extends the LH protocol. Unlike the optimized LH protocol (*i.e.*, OLH [34]), which aims solely to minimize estimation error by selecting $g = \lfloor e^{\varepsilon} + 1 \rceil$, ALH jointly optimizes both MSE and ASR by adapting the hash domain size parameter g. Formally the optimization problem for ALH is defined as:

$$\begin{aligned} \min_{g \in \mathbb{Z}} \quad & J(g) = w_{\mathrm{MSE}} \cdot \left(\frac{(e^{\varepsilon} - 1 + g)^2}{n(e^{\varepsilon} - 1)^2 (g - 1)} \right) \\ & \quad + w_{\mathrm{ASR}} \cdot \left(\frac{e^{\varepsilon}}{(e^{\varepsilon} + g - 1) \cdot \max\left\{\frac{k}{g}, 1\right\}} \right), \\ \textit{s.t.} \quad & 2 \leq g \leq \max\left(k, \lfloor e^{\varepsilon} + 1 \rceil\right), \ w_{\mathrm{ASR}} + w_{\mathrm{MSE}} = 1, \end{aligned} \tag{26}$$

The upper bound for g is set to $\max\left(k, \lfloor e^{\varepsilon} + 1 \rceil\right)$, ensuring that g is not unnecessarily smaller than the original domain size k. This avoids under-hashing, allowing for better differentiation and randomness. When k is large, g should ideally match or exceed k, ensuring that hashing effectively introduces randomness to maintain privacy guarantees.

5.3.4 ***Adaptive THE (ATHE)***. The ATHE mechanism extends the THE protocol. Unlike THE [34], which only aims to minimize the MSE in Equation (20), ATHE jointly optimizes both MSE and ASR by adapting the threshold parameter θ. More formally, the optimization problem for ATHE is defined as:

$$\begin{aligned}\min_{\theta\in\mathbb{R}} \quad & J(\theta) = w_{\text{MSE}} \cdot \left(\frac{2e^{\varepsilon\theta/2} - 1}{(1 + e^{\varepsilon(\theta - 1/2)} - 2e^{\varepsilon\theta/2})^2} \right) \\ & \quad + w_{\text{ASR}} \cdot \Bigg((1-p)(1-q)^{k-1} \cdot \frac{1}{k} \\ & \quad + \sum_{m=1}^{k} \binom{k-1}{m-1} p(q)^{m-1}(1-q)^{k-m} \cdot \frac{1}{m} \Bigg), \\ \text{s.t.} \quad & 0.5 \le \theta \le 1, \; w_{\text{ASR}} + w_{\text{MSE}} = 1, \end{aligned} \tag{27}$$

where p and q are given in Equation (18). The constraint $0.5 \le \theta \le 1$ follows the settings used in prior work [34].

5.4 Optimization Strategy

Our adaptive protocols are optimized to jointly minimize ASR and MSE while considering the interdependence of parameters. Due to the non-linearity of the cost functions (see Equations (24), (25), (26), and (27)), deriving closed-form solutions is infeasible in most cases. Instead, we employ numerical optimization to efficiently explore the parameter space and avoid convergence to suboptimal solutions.

General Optimization Techniques. We leverage the following numerical methods based on the characteristics of the search space:

- **Grid search** [10] (*a.k.a.* brute force methods) for exhaustive parameter evaluation, ensuring an optimal solution but at the cost of increased computation time.
- **Constrained optimization methods** [11], specifically the `Bounded Brent method` in `scipy.optimize.minimize_scalar`, to refine parameter selection when the search space is large and an exhaustive search is impractical.

6 EXPERIMENTS AND ANALYSIS

The objective of our experiments is to thoroughly evaluate the performance of the proposed adaptive LDP protocols in terms of privacy, utility, and resilience against privacy attacks. Specifically, in Section 6.2, we conduct an ***ASR analysis*** for each LDP protocol to quantify their vulnerability to distinguishability attacks, thereby assessing the privacy guarantees offered by existing and newly proposed mechanisms. Subsequently, in Section 6.3, we perform an ***MSE analysis*** to evaluate the utility guarantees provided by existing and our newly proposed mechanisms. Next, in Section 6.4, we explore the ***trade-off between ASR and MSE*** (*i.e.*, Pareto frontier), highlighting how our adaptive protocols compare to traditional ones in balancing privacy and utility under different scenarios. We then assess in Section 6.5 the ***impact of different weights*** w_{ASR} and w_{MSE} in the objective function to provide insights into the influence of prioritizing privacy versus utility. Afterward, we briefly analyze in Section 6.6 the ***parameter optimization*** approach for each adaptive protocol.

We highlight that to systematize these aforementioned experiments, we rely primarily on the *analytical closed-form equations* for both MSE and ASR, as prior research has demonstrated a strong agreement between analytical and empirical results [3, 21, 34]. This choice allows us to comprehensively evaluate the performance of our protocols across a wide range of scenarios while avoiding computational overhead. Nonetheless, we also include a concise ***empirical validation*** in Section 6.7 and Appendix C to ensure consistency and reinforce the validity of our analytical findings.

6.1 Setup of Analytical Experiments

For all experiments, we have used the following settings:

- **Environment.** All algorithms are implemented in Python 3 and run on a desktop machine with 3.2GHz Intel Core i9 and 64GB RAM.
- **LDP protocols.** We experiment with the eight LDP protocols described in Section 4 and our four adaptive LDP protocols described in Section 5.3.
- **Number of users.** For the analytical variance/MSE derived from Equation (4) (*e.g.*, Equation (6) for GRR, Equation (20) for THE, ...), we report variance per user, corresponding to the analytical derivation of $\mathrm{Var}[]/n$. This allows us to examine the fundamental properties of each protocol independently of the dataset size.
- **Privacy parameter.** The LDP frequency estimation protocols were evaluated under two privacy regimes:
  a) **High privacy regime** with $\varepsilon \in \{0.5, 0.6, \ldots, 1.9, 2.0\}$.
  b) **Medium to low privacy regime** with $\varepsilon \in \{2.0, 2.5, \ldots, 9.5, 10.0\}$.
- **Domain size.** We vary the domain size in three ranges:
  a) **Small domain:** $k \in \{25, 50, 75, 100\}$.
  b) **Medium domain:** $k \in \{250, 500, 750, 1000\}$.
  c) **Large domain:** $k \in \{2500, 5000, 7500, 10000\}$.
- **Weights for optimization.** Unless otherwise mentioned, we fix the weights for the two-objective optimization framework to $w_{\text{ASR}} = 0.5$ and $w_{\text{MSE}} = 0.5$, aiming for a balanced trade-off between privacy and accuracy. Experiments in Section 6.5 will focus on varying these weights.
- **Optimization method.** Parameters for our adaptive protocols (*e.g.*, subset size ω for ASS) were optimized using a grid search approach [10] as described in Section 5.4.

6.2 ASR Analysis for LDP Protocols

In Figure 1, we evaluate the ASR for various LDP frequency estimation protocols as a function of the privacy budget ε from high to low privacy regimes across small to big domain sizes k. The goal of this analysis is to assess the robustness of each protocol against adversarial inference attacks under varying privacy levels and domain sizes. Specifically, we compare state-of-the-art protocols (GRR, SUE, BLH, OUE, OLH, SS, SHE, and THE) against our proposed adaptive protocols (ASS, AUE, ALH, and ATHE) to understand how effective these methods are at balancing privacy and utility.

General trend across protocols: For all protocols, we observe that the ASR generally increases with increasing privacy budget ε in Figure 1. This behavior is expected, as higher values of ε correspond to weaker privacy guarantees, allowing the adversary to infer user data more effectively. For smaller domain sizes (*i.e.*, $k \le 100$, the

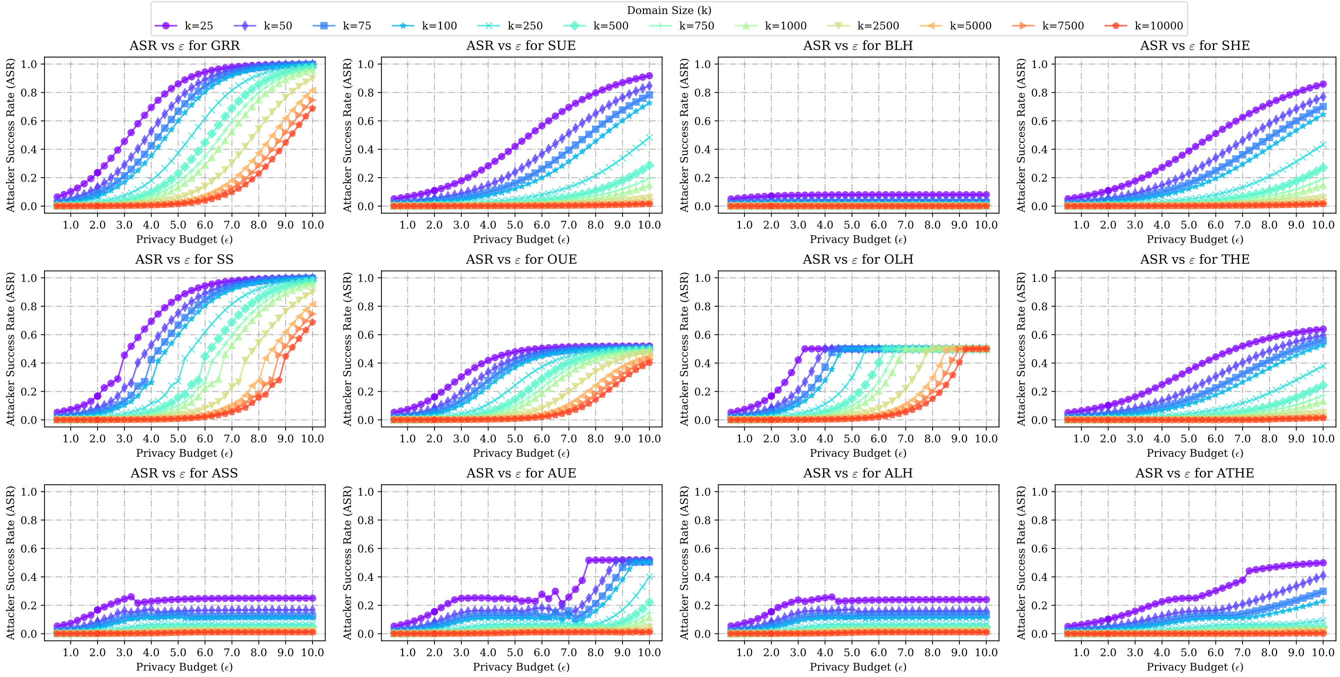


Figure 1: Attacker Success Rate (ASR) *vs.* privacy budget (ε) for different LDP frequency estimation protocols across varying domain sizes (k). The plots compare state-of-the-art LDP protocols, including GRR, SUE, BLH, OUE, OLH, SS, SHE, and THE, against our newly proposed adaptive protocols (ASS, AUE, ALH, ATHE). Each curve represents a different domain size, with k ranging from 25 to 10000. The figure highlights the trade-offs between privacy and adversarial resilience for each protocol, showing how ASR evolves as the privacy budget and domain size change.

ASR rises sharply, suggesting that the adversary's ability to correctly guess the user's input improves significantly as ε increases. In contrast, larger domain sizes (*i.e.*, $k \geq 1000$) show a more gradual increase in ASR, indicating that a larger domain inherently offers greater privacy protection. Nevertheless, ***our adaptive methods effectively counteract privacy threats regardless of*** k ***and*** ε, underscoring the flexibility and robustness offered by our double-objective optimization framework.

Comparing traditional protocols: Among traditional protocols, GRR and SS demonstrate the highest ASR across all privacy budgets and domain sizes, making them the most vulnerable to privacy attacks. In contrast, SUE displays a more gradual increase in ASR, suggesting better resilience than GRR and SS. Protocols such as OUE and OLH exhibit a maximum ASR of approximately 0.5, even as ε increases, showing a clear boundary in their ASR. Notably, for SHE, ASR grows gradually as ε increases, while THE's thresholding strategy yields moderate ASR increments. Overall, BLH consistently achieves an exceptionally low ASR across all privacy budgets and domain sizes, highlighting its robustness against adversarial inference attacks.

Our refined and adaptive protocols: Our adaptive protocols (*i.e.*, ASS, AUE, ALH, and ATHE) demonstrate significantly lower ASR compared to their traditional counterparts across all privacy budgets. For example, ASS effectively mitigates the ASR vulnerability of the state-of-the-art SS by capping the ASR below 0.25, in contrast to the traditional SS where the ASR gets to 1 (*i.e.*, the adversary can fully infer the user's value). Similarly, AUE consistently achieves a lower or comparable ASR relative to the state-of-the-art OUE protocol, highlighting the benefits of its adaptive parameter optimization. For ALH, the adaptive mechanism achieves a balanced compromise between the low ASR of BLH and the moderate ASR of OLH by dynamically optimizing the hash domain size, resulting in a substantial reduction in ASR compared to traditional OLH. Finally, ATHE demonstrates more resilience than THE across all domain sizes, thereby showcasing the effectiveness of our framework in enhancing privacy protection.

> **ASR increases with ε, but our adaptive protocols resist:** As the privacy budget ε increases, ASR generally rises for all protocols due to weaker privacy guarantees. However, our adaptive protocols (ASS, AUE, ALH, ATHE) exhibit significantly lower (*i.e.*, $\leq$ 5 orders of magnitude) ASRs across a broad range of ε values, underscoring their enhanced resilience against privacy attacks and their ability to maintain robust privacy protection.

6.3 MSE Analysis for LDP Protocols

In Figure 2, we examine the MSE behavior of the SS protocol and its adaptive counterpart ASS across varying privacy budgets ε. In Figure 3, we extend this analysis to compare MSE trends for UE-, LH-, and HE-based LDP protocols, including our adaptive versions (AUE, ALH, ATHE). These analyses aim to determine whether the adaptive protocols' improved robustness to privacy attacks results in substantial estimation accuracy loss or if they maintain competitive MSE values. Notice that the MSE curves of SUE, OUE, BLH, OLH, SHE, and THE remain independent of the domain size k due to their fixed parameterization, as established in previous literature [34].

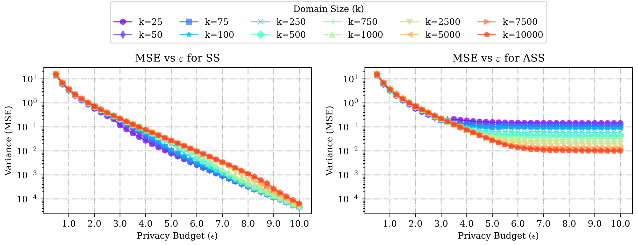


Figure 2: Variance (MSE) *vs.* privacy budget (ε) for the state-of-the-art SS protocol and our adaptive version ASS across various domain sizes k. Each curve represents a distinct domain size, illustrating how each protocol balances estimation accuracy with privacy as ε changes.

For the SS protocol, we observe in Figure 2 that ASS exhibits an increase in MSE by up to two orders of magnitude compared to SS, particularly in higher privacy regimes ($\varepsilon \geq 4$). This increase is expected, as SS's minimal MSE comes at the cost of extreme vulnerability, with ASR approaching 1 (*e.g.*, see Figure 1), meaning an adversary can fully infer the user's value. ASS, in contrast, balances this trade-off by introducing adaptive parameterization that mitigates privacy attacks while moderately increasing the MSE.

In Figure 3, we observe similar trends for the other adaptive protocols (AUE, ALH, and ATHE). For UE protocols, our adaptive AUE version achieves slightly higher MSE compared to OUE across all domain sizes k, with the gap becoming more pronounced as ε increases. This is expected, as AUE optimizes its parameters to enhance robustness to privacy attacks, leading to a slight trade-off in utility. Notably, AUE remains competitive with SUE in high privacy regimes ($\varepsilon \leq 2$), offering similar MSE levels while achieving significantly improved ASR performance as shown earlier. In low privacy regimes ($\varepsilon > 2$), AUE incurs a modest increase in MSE, which stays within an order of magnitude compared to OUE.

Moreover, for LH- and HE-based protocols, we observe in Figure 3 that our adaptive protocols (ALH and ATHE) demonstrate a variance (MSE) behavior that is consistently "sandwiched" between the two corresponding state-of-the-art protocols in their respective groups. Specifically, ALH achieves MSE values between those of BLH (which minimizes ASR at the cost of higher MSE) and OLH (which minimizes MSE but is more vulnerable to privacy attacks). Similarly, ATHE's variance lies between SHE and THE, showing a trade-off where ATHE retains competitive MSE while prioritizing adversarial resilience. This positioning highlights how adaptivity allows our protocols to achieve a better privacy-utility trade-off.

> **Adaptive protocols maintain competitive MSE:** Our adaptive protocols (ASS, AUE, ALH, ATHE) achieve higher MSE compared to their non-adaptive counterparts. However, these increases remain within acceptable bounds ($\leq$ 2 orders of magnitude), highlighting that the improved adversarial resilience (ASR $\leq$ 5 orders of magnitude) comes at a reasonable cost to the utility. This suggests that enhanced privacy protection against adversaries can be achieved without prohibitive increases in variance.

6.4 Pareto Frontier for ASR and MSE

In addition to evaluating how the ASR and the MSE vary with the privacy budget ε separately (see Figures 1, 2 and 3), it is insightful to examine how ASR changes as a function of utility loss, measured by the variance (MSE) of the frequency estimation. Figure 4 presents the Pareto frontier between ASR and MSE for the considered LDP protocols, plotting ASR against MSE across small domain sizes and medium to low privacy regimes. This analysis highlights the performance of state-of-the-art protocols (GRR, SUE, BLH, SHE, SS, OUE, OLH, and THE) compared to our adaptive variants (ASS, AUE, ALH, and ATHE), providing insights into how effectively each method navigates the trade-off between user privacy and data utility. To address space limitations, additional results exploring the ASR-MSE trade-off under varying privacy regimes (high, medium, low) and domain sizes (small, medium, large) are provided in Appendix B. The findings and discussions in this section are supported by a comprehensive analysis that includes these extended results.

General trend across protocols: For all protocols, there is a clear inverse relationship between ASR and MSE – lower variance estimates correspond to higher ASR, and vice versa. As the privacy budget ε increases, protocols tend to yield estimates with lower MSE but simultaneously face higher ASRs, reflecting a direct cost to privacy when pursuing higher accuracy. Conversely, under stricter privacy regimes (lower ε), while ASR remains lower, the resulting estimates incur greater MSE. This fundamental tension highlights that simply tuning ε does not guarantee a balanced privacy-utility outcome. Thus, understanding the ASR-MSE relationship is crucial for informed protocol selection.

Comparing traditional protocols: Among the traditional protocols, GRR and SS tend to cluster in regions with relatively lower MSE but higher ASR, especially at moderate-to-high ε values. In contrast, SUE, SHE, and THE provide a more gradual trade-off curve, achieving lower ASR at slightly higher MSE levels, suggesting that these methods preserve more privacy when aiming for moderate accuracy. OUE and OLH occupy intermediate positions: they can reach points of low MSE but at the cost of increasing ε. BLH generally exhibits a low ASR while not bounding the MSE, indicating a low privacy-utility trade-off. Overall, none of the traditional protocols dominate the entire ASR-MSE space, and their effectiveness varies with the chosen privacy budget and domain size.

Our refined and adaptive protocols: Our adaptive protocols achieve more favorable ASR-MSE trade-offs, especially providing low-ASR even in low-privacy regimes (*i.e.*, high ε values). For instance, ASS, derived from SS, effectively prevents the sharp ASR increases observed in SS's low-MSE regions by fine-tuning protocol parameters, resulting in points that align closer to a Pareto frontier between ASR and MSE. AUE, adapting from OUE, displays a particularly strong trade-off, often settling into low-ASR regimes

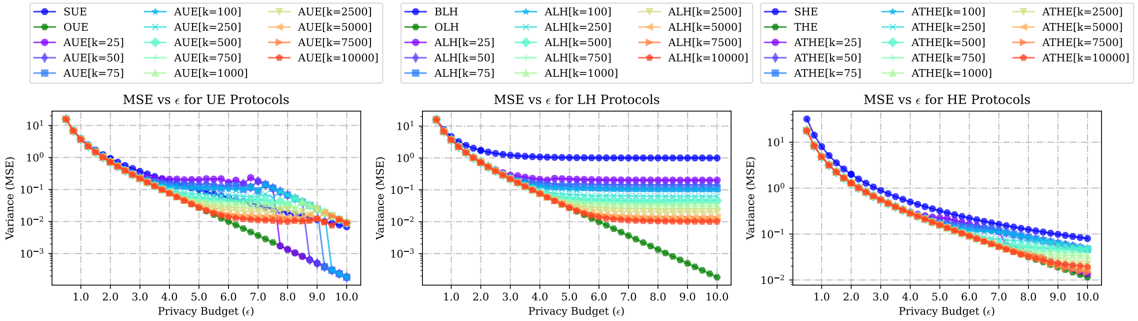


Figure 3: Variance (MSE) *vs.* privacy budget (ε) for the state-of-the-art LDP protocols (UE-, LH-, and HE-based) and our adaptive versions (AUE, ALH, and ATHE) across various domain sizes k. For our adaptive protocols, each curve represents a distinct domain size, illustrating how each protocol balances estimation accuracy with privacy as ε changes.

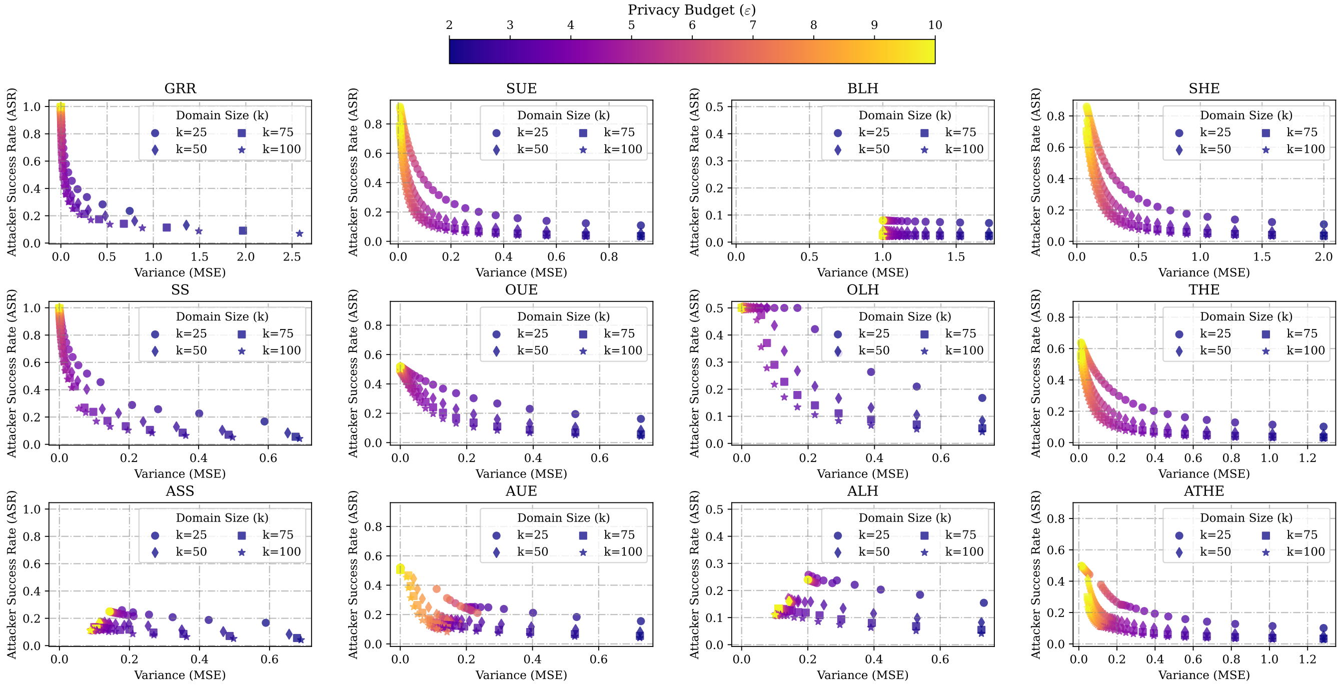


Figure 4: Attacker Success Rate (ASR) *vs.* Variance (MSE) for numerous LDP frequency estimation protocols. Each plot shows how each protocol performs under varying privacy budgets ε and domain sizes (k), illustrating the trade-off between adversarial success rate (ASR) and utility (MSE). State-of-the-art LDP protocols (*i.e.*, GRR, SUE, BLH, SHE, SS, OUE, OLH, and THE) are compared against our adaptive counterparts (*i.e.*, ASS, AUE, ALH, and ATHE). Each point represents a different configuration of ε (in medium to low privacy regimes) and k (small domain), with colors indicating the privacy budget level.

without disproportionately large MSE. ALH, building on hashing-based approaches, maintains ASR reductions comparable to BLH or OLH but achieves them with more controlled MSE levels, ensuring that the benefits of hashing-based schemes are not compromised. ATHE consistently achieves configurations that lower ASR without excessively increasing MSE, indicating a more harmonious balance.

> **ASR-MSE trade-offs are protocol-dependent:** Our findings show that each protocol exhibits a distinct ASR-MSE profile. By jointly optimizing for both privacy (ASR) and utility (MSE), our adaptive protocols consistently push these curves toward more favorable regimes in the ASR-MSE Pareto frontier.

6.5 Impact of Weights in the Objective Function

Thus far, we have focused on analyzing the performance of adaptive protocols under fixed objective configurations for the weights ($w_{\text{ASR}} = w_{\text{MSE}} = 0.5$) in Equation (23). However, our proposed multi-objective framework introduces a new degree of freedom: practitioners can adjust the relative importance of ASR *vs.* variance (MSE) when optimizing LDP protocols. Figure 5 illustrates how varying these weight combinations ($w_{\text{ASR}}, w_{\text{MSE}}$) influences the ASR, MSE, and optimal parameter choices for AUE (p), ALH (g), ASS (ω), and ATHE (θ) under a fixed setting ($k = 100$, $\varepsilon = 4$).

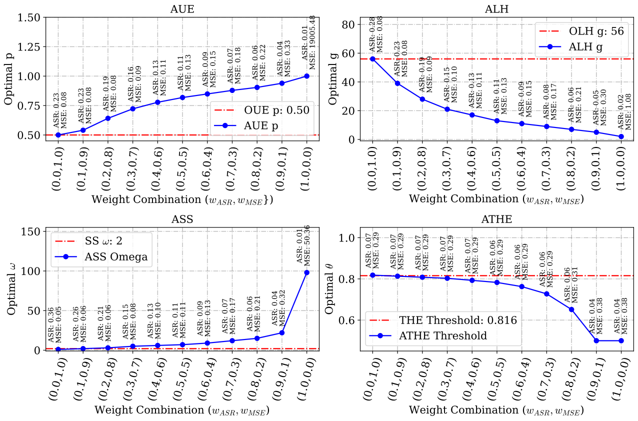


Figure 5: Optimal parameter choices for each adaptive protocol as a function of the weight combination ($w_{\text{ASR}}, w_{\text{MSE}}$), evaluated at $k = 100$ and $\varepsilon = 4$. Each sub-figure compares the adaptive protocol's chosen parameter (blue color curve) against the corresponding parameter choice in the original, non-adaptive protocol (red color dashed line).

From Figure 5, one can notice that as the weight on ASR w_{ASR} increases, each adaptive protocol tends to choose parameter values that more aggressively reduce the ASR at the cost of increasing the MSE. Conversely, placing greater emphasis on MSE drives parameters toward configurations closer to or equal to those of the original protocols (*i.e.*, SS, OUE, OLH, and THE), aiming to preserve utility even if it elevates the ASR. These results confirm the benefits of our two-objective optimization framework: rather than a static parameter choice, practitioners can tune the protocol parameters in response to changing priorities, achieving a more flexible trade-off between privacy (ASR) and utility (MSE).

6.6 Adaptive and Optimized Parameters

In this section, we analyze the optimization of parameters in our adaptive protocols (AUE, ALH, ASS, and ATHE) by examining the behavior of their objective functions (Equations (24)–(27)), which balance the ASR-MSE trade-off. Figure 6 illustrates the objective function as a function of key parameters: AUE (p), ALH (g), ASS (ω), and ATHE (θ), with a fixed $k = 100$ and $\varepsilon = 4$. The selected parameters for our adaptive protocols are compared against the fixed state-of-the-art parameters of OUE, OLH, SS, and THE.

For **AUE**, we observe in the top-left plot that the objective function reaches its minimum at $p = 0.818$, which is notably higher than the fixed $p = 0.5$ used by OUE. This also means that parameter q will increase to satisfy ε-LDP, *i.e.*, increasing the probability of reporting random bits. For **ALH**, as shown in the top-right plot, the optimal hash domain size is reduced to $g = 13$ in our adaptive protocol compared to $g = 56$ in OLH. This reduction in g lowers the ASR at the expense of slightly increased variance, aligning with our adaptive objective to achieve a better balance between privacy and utility. For **ASS**, as depicted in the bottom-left plot, the optimal subset size $\omega = 7$ contrasts with the fixed $\omega = 2$ used in SS. The larger ω effectively spreads the probability mass across a larger subset, reducing ASR while incurring a higher variance. For **ATHE**, the bottom-right plot shows that the adaptive threshold $\theta = 0.783$ is slightly lower than the fixed threshold $\theta = 0.816$ in THE. This subtle adjustment enables ATHE to reduce privacy attacks while maintaining competitive variance levels, showcasing the precision of our adaptive optimization.

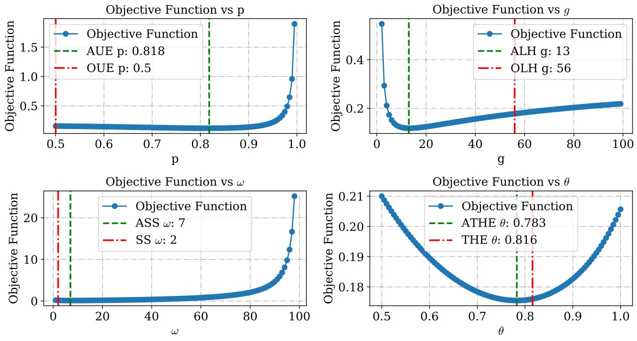


Figure 6: Objective function value as a function of key parameters for our adaptive protocols (AUE, ALH, ASS, and ATHE) compared with their state-of-the-art counterparts (OUE, OLH, SS, and THE). Vertical dashed lines (green color) indicate the optimal parameter values selected by our adaptive protocols, while vertical dash-dotted lines (red color) represent the fixed parameter values of the state-of-the-art protocols. Without loss of generality, we set $k = 100$ and $\varepsilon = 4$.

> **Adaptive protocols optimize key parameters for better trade-offs:** Our findings demonstrate that the optimized parameters selected by adaptive protocols significantly differ from the fixed parameters of state-of-the-art protocols, resulting in improved robustness to privacy attacks with controlled increases in variance. This optimization highlights the flexibility and effectiveness of our adaptive mechanisms in better navigating the privacy-utility trade-off space.

6.7 Empirical Pareto Frontier for ASR and MSE

To assess the accuracy of our closed-form ASR and MSE equations within the ASR-MSE two-objective framework (Section 5.2), we conduct empirical experiments using the **Adult** dataset from the UCI machine learning repository [19]. We select the Age attribute with domain size $k = 100$ (*i.e.*, Age $\in [0, 99]$) and $n = 48842$ users, reporting results averaged over 100 independent runs.

Figure 7 presents the comparison between **empirical** (○ makers) and **analytical** (red dashed lines) Pareto frontiers for ASR versus MSE across various LDP protocols, including both state-of-the-art and our adaptive variants. Each subplot represents a specific protocol (*i.e.*, GRR, SUE, BLH, SHE, SS, OUE, OLH, THE, ASS, AUE,

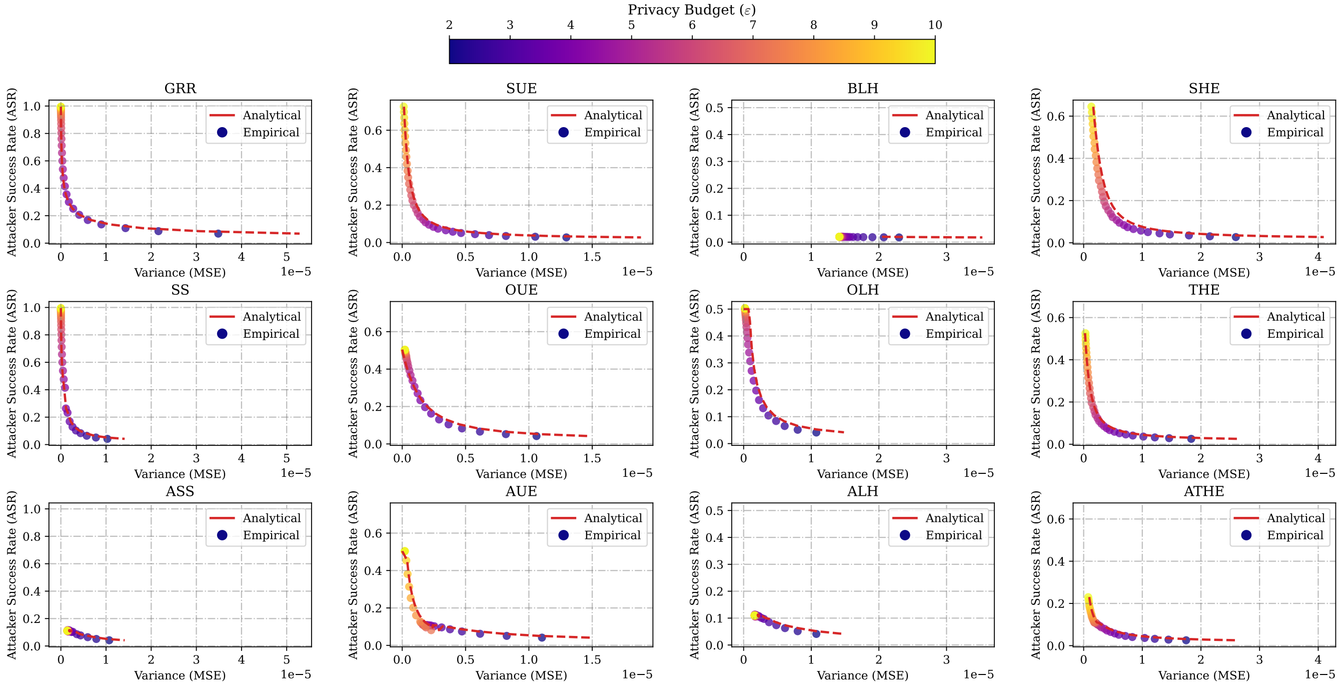


Figure 7: Comparison of empirical and analytical Pareto frontiers for ASR *vs.* Variance (MSE) across various LDP protocols (state-of-the-art and adaptive). Each subplot considers a range of privacy budgets $\varepsilon \in (2, 10)$ and a fixed domain size $k = 100$. Empirical results (○ makers) are averaged over 100 independent runs with the Age attribute of the Adult dataset [19], while analytical results (red dashed lines) are computed using closed-form equations.

ALH, and ATHE) with the privacy budget varying as $\varepsilon \in (2, 10)$ with the domain size fixed at $k = 100$. One can notice that across all protocols, the empirical results closely align with the analytical calculation, showcasing the robustness of the closed-form equations for ASR and MSE. This alignment highlights the reliability of our theoretical framework and previous analytical experiments across different privacy budgets and data distributions. To further support this claim, Appendix C provides additional experiments using a synthetic dataset generated from a Dirichlet distribution (**1**), exploring variations in the number of users ($n \in \{5000, 50000, 500000\}$).

> **Analytical *vs.* empirical validation:** Our findings demonstrate that the analytical results closely align with empirical ones, reaffirming the reliability of the closed-form equations across various privacy budgets and scenarios.

7 CONCLUSION AND PERSPECTIVES

In this work, we introduced a **general multi-objective optimization framework** for refining LDP frequency estimation protocols, enabling adaptive parameter tuning based on multiple privacy and utility considerations. While the classical LDP paradigm primarily focuses on minimizing the estimation error (*i.e.*, utility-driven), our framework provides a flexible optimization approach that balances multiple objectives in adversarial settings. As an instantiation of this framework, we focused on a two-objective formulation that jointly minimizes the Attacker Success Rate (ASR) under distinguishability attacks [5, 6, 21] and Mean Squared Error (MSE), demonstrating that existing protocols (*i.e.*, SS, OUE, OLH, and THE) can be beneficially re-optimized. Our adaptive protocols, namely, ASS, AUE, ALH, and ATHE, significantly reduce adversarial success rates while maintaining competitive estimation accuracy (*e.g.*, see Figure 4).

Beyond this specific instantiation, our findings pave the way for broader applications of our framework. First, our approach naturally extends to **alternative privacy, utility, and security objectives**, such as robustness to poisoning attacks [12, 13], resilience against inference [7, 25] and re-identification [6, 29] risks, or the integration of entropy-based utility metrics [8]. Additionally, incorporating constraints on communication cost and efficiency would enhance its applicability to real-world, resource-constrained environments. Finally, we envision the development of an LDP optimization suite that accommodates various objectives, such as constrained utility, constrained ε, double-objective problems, and so on. This suite could serve as a practical tool for deploying LDP mechanisms tailored to specific real-world requirements.

ACKNOWLEDGMENTS

This work has been partially supported by the French National Research Agency (ANR), under contract “ANR-24-CE23-6239” JCJC project AI-PULSE and by the “ANR 22-PECY-0002” IPOP (Interdisciplinary Project on Privacy) project of the Cybersecurity PEPR. Sébastien Gambs is supported by the Canada Research Chair program as well as a Discovery Grant from NSERC.

REFERENCES

[1] Jayadev Acharya, Ziteng Sun, and Huanyu Zhang. 2019. Hadamard Response: Estimating Distributions Privately, Efficiently, and with Little Communication. In *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics (Proceedings of Machine Learning Research)*, Kamalika Chaudhuri and Masashi Sugiyama (Eds.), Vol. 89. PMLR, 1120–1129.

[2] Ece Alptekin and M Emre Gursoy. 2023. Building quadtrees for spatial data under local differential privacy. In *IFIP Annual Conference on Data and Applications Security and Privacy*. Springer, 22–39. https://doi.org/10.1007/978-3-031-37586-6_2

[3] Héber H. Arcolezi, Jean-François Couchot, Bechara Al Bouna, and Xiaokui Xiao. 2021. Random Sampling Plus Fake Data: Multidimensional Frequency Estimates With Local Differential Privacy. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*. ACM, 47–57. https://doi.org/10.1145/3459637.3482467

[4] Héber H. Arcolezi, Jean-François Couchot, Bechara Al Bouna, and Xiaokui Xiao. 2024. Improving the utility of locally differentially private protocols for longitudinal and multidimensional frequency estimates. *Digital Communications and Networks* 10, 2 (2024), 369–379. https://doi.org/10.1016/j.dcan.2022.07.003

[5] Héber Hwang Arcolezi and Sébastien Gambs. 2024. Revealing the True Cost of Locally Differentially Private Protocols: An Auditing Perspective. *Proceedings on Privacy Enhancing Technologies* 2024, 4 (2024), 123–141. https://doi.org/10.56553/popets-2024-0110

[6] Héber H. Arcolezi, Sébastien Gambs, Jean-François Couchot, and Catuscia Palamidessi. 2023. On the Risks of Collecting Multidimensional Data Under Local Differential Privacy. *Proc. VLDB Endow.* 16, 5 (jan 2023), 1126–1139. https://doi.org/10.14778/3579075.3579086

[7] Héber H. Arcolezi, Carlos A Pinzón, Catuscia Palamidessi, and Sébastien Gambs. 2023. Frequency Estimation of Evolving Data Under Local Differential Privacy. In *Proceedings of the 26th International Conference on Extending Database Technology, EDBT 2023, Ioannina, Greece, March 28 - March 31, 2023*. OpenProceedings.org, 512–525. https://doi.org/10.48786/EDBT.2023.44

[8] Leighton Pate Barnes, Wei-Ning Chen, and Ayfer Özgür. 2020. Fisher Information Under Local Differential Privacy. *IEEE Journal on Selected Areas in Information Theory* 1, 3 (2020), 645–659. https://doi.org/10.1109/JSAIT.2020.3039461

[9] Raef Bassily and Adam Smith. 2015. Local, Private, Efficient Protocols for Succinct Histograms. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing* (Portland, Oregon, USA) *(STOC '15)*. Association for Computing Machinery, New York, NY, USA, 127–135. https://doi.org/10.1145/2746539.2746632

[10] James Bergstra and Yoshua Bengio. 2012. Random search for hyper-parameter optimization. *Journal of machine learning research* 13, 2 (2012).

[11] Richard P Brent. 2013. *Algorithms for minimization without derivatives*. Courier Corporation.

[12] Xiaoyu Cao, Jinyuan Jia, and Neil Zhenqiang Gong. 2021. Data Poisoning Attacks to Local Differential Privacy Protocols. In *30th USENIX Security Symposium (USENIX Security 21)*. USENIX Association, 947–964.

[13] Albert Cheu, Adam Smith, and Jonathan Ullman. 2021. Manipulation Attacks in Local Differential Privacy. In *2021 IEEE Symposium on Security and Privacy (SP)*. IEEE. https://doi.org/10.1109/sp40001.2021.00001

[14] 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. https://doi.org/10.14778/3476249.3476261

[15] José Serafim Costa Filho and Javam C Machado. 2023. FELIP: A local Differentially Private approach to frequency estimation on multidimensional datasets. In *Proceedings of the 26th International Conference on Extending Database Technology, EDBT 2023, Ioannina, Greece, March 28 - March 31, 2023*. OpenProceedings.org, 671–683. https://doi.org/10.48786/EDBT.2023.56

[16] Thomas M. Cover and Joy A. Thomas. 2006. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, USA.

[17] Harald Cramér. 1999. *Mathematical Methods of Statistics (PMS-9)*. Princeton University Press.

[18] Bolin Ding, Janardhan Kulkarni, and Sergey Yekhanin. 2017. Collecting Telemetry Data Privately. In *Advances in Neural Information Processing Systems 30*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett (Eds.). Curran Associates, Inc., 3571–3580.

[19] Dheeru Dua and Casey Graff. 2017. UCI Machine Learning Repository. Available online: http://archive.ics.uci.edu/ml.

[20] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating Noise to Sensitivity in Private Data Analysis. In *Theory of Cryptography*. Springer Berlin Heidelberg, 265–284. https://doi.org/10.1007/11681878_14

[21] 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 Transactions on Information Forensics and Security* 17 (2022), 1785–1799. https://doi.org/10.1109/TIFS.2022.3170242

[22] Úlfar Erlingsson, Vasyl Pihur, and Aleksandra Korolova. 2014. RAPPOR: Randomized Aggregatable Privacy-Preserving Ordinal Response. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (Scottsdale, Arizona, USA). ACM, New York, NY, USA, 1054–1067. https://doi.org/10.1145/2660267.2660348

[23] Vitaly Feldman, Jelani Nelson, Huy Nguyen, and Kunal Talwar. 2022. Private frequency estimation via projective geometry. In *Proceedings of the 39th International Conference on Machine Learning (Proceedings of Machine Learning Research)*, Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (Eds.), Vol. 162. PMLR, 6418–6433.

[24] Xiaolan Gu, Ming Li, Yang Cao, and Li Xiong. 2019. Supporting Both Range Queries and Frequency Estimation with Local Differential Privacy. In *2019 IEEE Conference on Communications and Network Security (CNS)*. 124–132. https://doi.org/10.1109/CNS.2019.8802778

[25] Mehmet Emre GÜRSOY. 2024. Longitudinal attacks against iterative data collection with local differential privacy. *Turkish Journal of Electrical Engineering and Computer Sciences* 32, 1 (Feb. 2024), 198–218. https://doi.org/10.55730/1300-0632.4063

[26] Peter Kairouz, Keith Bonawitz, and Daniel Ramage. 2016. Discrete distribution estimation under local privacy. In *International Conference on Machine Learning*. PMLR, 2436–2444.

[27] 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. https://doi.org/10.1137/090756090

[28] Junhui Li, Wensheng Gan, Yijie Gui, Yongdong Wu, and Philip S. Yu. 2022. Frequent Itemset Mining with Local Differential Privacy. In *Proceedings of the 31st ACM International Conference on Information & Knowledge Management* (Atlanta, GA, USA) *(CIKM '22)*. Association for Computing Machinery, New York, NY, USA, 1146–1155. https://doi.org/10.1145/3511808.3557327

[29] Takao Murakami and Kenta Takahashi. 2021. Toward Evaluating Re-identification Risks in the Local Privacy Model. *Transactions on Data Privacy* 14, 3 (2021), 79–116.

[30] René Raab, Pascal Berrang, Paul Gerhart, and Dominique Schröder. 2025. SoK: Descriptive Statistics Under Local Differential Privacy. *Proceedings on Privacy Enhancing Technologies* 2025, 1 (Jan. 2025), 118–149. https://doi.org/10.56553/popets-2025-0008

[31] C Radhakrishna Rao. 1992. Information and the accuracy attainable in the estimation of statistical parameters. In *Breakthroughs in Statistics: Foundations and basic theory*. Springer, 235–247.

[32] Ekin Tire and M. Emre Gursoy. 2024. Answering Spatial Density Queries Under Local Differential Privacy. *IEEE Internet of Things Journal* 11, 10 (2024), 17419–17436. https://doi.org/10.1109/JIOT.2024.3357570

[33] Shaowei Wang, Liusheng Huang, Pengzhan Wang, Yiwen Nie, Hongli Xu, Wei Yang, Xiang-Yang Li, and Chunming Qiao. 2016. Mutual information optimally local private discrete distribution estimation. *arXiv preprint arXiv:1607.08025* (2016).

[34] Tianhao Wang, Jeremiah Blocki, Ninghui Li, and Somesh Jha. 2017. Locally Differentially Private Protocols for Frequency Estimation. In *26th USENIX Security Symposium (USENIX Security 17)*. USENIX Association, Vancouver, BC, 729–745.

[35] Yue Wang, Xintao Wu, and Donghui Hu. 2016. Using randomized response for differential privacy preserving data collection. In *EDBT/ICDT Workshops*, Vol. 1558. 0090–6778.

[36] Stanley L. Warner. 1965. Randomized Response: A Survey Technique for Eliminating Evasive Answer Bias. *J. Amer. Statist. Assoc.* 60, 309 (March 1965), 63–69. https://doi.org/10.1080/01621459.1965.10480775

[37] Haonan Wu, Ruisheng Ran, Shunshun Peng, Mengmeng Yang, and Taolin Guo. 2023. Mining frequent items from high-dimensional set-valued data under local differential privacy protection. *Expert Systems with Applications* 234 (Dec. 2023), 121105. https://doi.org/10.1016/j.eswa.2023.121105

[38] Jianyu Yang, Tianhao Wang, Ninghui Li, Xiang Cheng, and Sen Su. 2020. Answering multi-dimensional range queries under local differential privacy. *arXiv preprint arXiv:2009.06538* (2020).

[39] Min Ye and Alexander Barg. 2018. Optimal Schemes for Discrete Distribution Estimation Under Locally Differential Privacy. *IEEE Transactions on Information Theory* 64, 8 (2018), 5662–5676. https://doi.org/10.1109/TIT.2018.2809790

[40] Yuemin Zhang, Qingqing Ye, and Haibo Hu. 2025. Federated Heavy Hitter Analytics with Local Differential Privacy. *Proc. ACM Manag. Data* 3, 1, Article 42 (Feb. 2025), 27 pages. https://doi.org/10.1145/3709739

A EXPECTED ASR ANALYSES

A.1 UE Protocols

Following the attack strategy for UE in Section 4.3, we consider two events:

- **Event 0:** The bit corresponding to the user's value x is flipped from 1 to 0, and all other bits remain 0.
  - The attacker's guess is uniformly distributed over all k possible values.
  - Success rate: $\frac{1}{k}$.
  - Probability: $\Pr(\text{Event 0}) = (1-p)(1-q)^{k-1}$.
- **Event 1:** The bit corresponding to the user's value x remains 1, and $m-1$ of the remaining $k-1$ bits are flipped from 0 to 1.
  - The attacker's guess is uniformly distributed over the bits set to 1.
  - If m bits are set to 1, the success rate is $\frac{1}{m}$.
  - Probability: $\Pr(\text{Event 1 with } m \text{ bits set to 1}) = \binom{k-1}{m-1} p(q)^{m-1}(1-q)^{k-m}$.

Thus, combining these two events, we can derive the expected ASR as:

$$\mathbb{E}[\text{ASR}]_{\text{UE}} = \Pr(\text{Event 0}) \cdot \frac{1}{k} + \sum_{m=1}^{k} \Pr(\text{Event 1 with } m \text{ bits set to 1}) \cdot \frac{1}{m}.$$

More formally, the probability calculations of each event are:

(1) **Probability of Event 0:**

$$\Pr(\text{Event 0}) = (1-p)(1-q)^{k-1}.$$

(2) **Probability of Event 1 with m bits set to 1:**

$$\Pr(\text{Event 1 with } m \text{ bits set to 1}) = \binom{k-1}{m-1} p(q)^{m-1}(1-q)^{k-m}.$$

Combining these probabilities, the expected ASR for UE is:

$$\mathbb{E}[\text{ASR}]_{\text{UE}} = (1-p) \cdot (1-q)^{k-1} \cdot \frac{1}{k} + \sum_{m=1}^{k} p \cdot \frac{1}{m} \cdot \binom{k-1}{m-1} q^{m-1}(1-q)^{(k-1)-(m-1)}.$$

A.2 SHE Protocol

The expected ASR of SHE is defined as the probability that $\hat{x} = x$:

$$\mathbb{E}[\text{ASR}]_{\text{SHE}} = \Pr[\hat{x} = x] = \Pr\left[y_x > \max_{i \neq x} y_i\right].$$

Define the random variables:

$$y_x = 1 + Z_x, \quad \text{where } Z_x \sim \text{Laplace}(0, b),$$
$$y_i = 0 + Z_i = Z_i, \quad \text{where } Z_i \sim \text{Laplace}(0, b), \quad \forall i \neq x$$

All Z_i and Z_x are independent random variables. Let $M = \max_{i \neq x} y_i = \max_{i \neq x} Z_i$. Then, the expected ASR becomes:

$$\mathbb{E}[\text{ASR}]_{\text{SHE}} = \Pr[y_x > M] = \Pr[1 + Z_x > M] = \Pr[Z_x > M - 1].$$

The cumulative distribution function (CDF) of the Laplace distribution $Z_x \sim \text{Laplace}(0, b)$ is:

$$F_Z(z) = \begin{cases} \frac{1}{2} \exp\left[\frac{z}{b}\right], & \text{if } z \leq 0, \\ 1 - \frac{1}{2} \exp\left[-\frac{z}{b}\right], & \text{if } z > 0. \end{cases}$$

The probability density function (PDF) of Z_x is:

$$f_Z(z) = \frac{1}{2b} \exp\left[-\frac{|z|}{b}\right].$$

For $M = \max_{i \neq x} Z_i$, since Z_i are independent and identically distributed (i.i.d.), the CDF of M is:

$$F_M(m) = [F_Z(m)]^{k-1}. \tag{28}$$

The PDF of M is then:

$$f_M(m) = (k-1)\,[F_Z(m)]^{k-2}\, f_Z(m). \tag{29}$$

The ASR can be expressed as:

$$\mathbb{E}[\text{ASR}]_{\text{SHE}} = \int_{-\infty}^{\infty} \Pr[Z_x > m - 1]\, f_M(m)\, dm. \tag{30}$$

Since Z_x and M are independent, $\Pr[Z_x > m-1] = 1 - F_Z(m-1)$. Therefore:

$$\mathbb{E}[\text{ASR}]_{\text{SHE}} = \int_{-\infty}^{\infty} [1 - F_Z(m-1)]\, f_M(m)\, dm. \tag{31}$$

Empirical Estimation via Simulation. In this work, we estimate the expected ASR in Equation (31) empirically using Monte Carlo simulations, following:

(1) **Generate Samples**:
- Sample Z_x from $\text{Laplace}(0, b)$.
- Sample Z_i for $i \neq x$ and compute $M = \max_{i \neq x} Z_i$.

(2) **Compute Success Indicator**:
- For each sample, check if $1 + Z_x > M$.

(3) **Estimate ASR**:
- The ASR is estimated as the proportion of times $1 + Z_x > M$ holds over all samples.

A.3 THE Protocol

Following the attack strategy for THE in Section 4.5.2, we have:

- **Event 0:** The bit corresponding to the user's value x is less than θ (*i.e.*, remains 0) and all other bits also remain 0.
  - The attacker's guess is uniformly distributed over all k possible values.
  - Success rate: $\frac{1}{k}$.
  - Probability: $\Pr(\text{Event 0}) = (1-p)(1-q)^{k-1}$.
- **Event 1:** The bit corresponding to the user's value x is greater than θ (*i.e.*, flips to 1) and $m-1$ other bits also flip to 1.
  - The attacker's guess is uniformly distributed over the bits set to 1.
  - If m bits are set to 1, the success rate is $\frac{1}{m}$
  - Probability: $\Pr(\text{Event 1 with } m \text{ bits set to 1}) = \binom{k-1}{m-1} p(q)^{m-1}(1-q)^{k-m}$.

Thus, by combining these two events, we can derive the expected ASR as:

$$\mathbb{E}[\text{ASR}]_{\text{THE}} = \Pr(\text{Event 0}) \cdot \frac{1}{k} + \sum_{m=1}^{k} \Pr(\text{Event 1 with } m \text{ bits set to 1}) \cdot \frac{1}{m}.$$

More formally, the probability calculations of each event are:

(1) **Probability of Event 0:**

$$\Pr(\text{Event 0}) = (1-p)(1-q)^{k-1}.$$

(2) **Probability of Event 1 with m bits set to 1:**

$$\Pr(\text{Event 1 with } m \text{ bits set to 1}) = \binom{k-1}{m-1} p(q)^{m-1}(1-q)^{k-m}.$$

Combining these probabilities, the expected ASR for THE is:

$$\mathbb{E}[\text{ASR}]_{\text{THE}} = (1-p)(1-q)^{k-1} \cdot \frac{1}{k} + \sum_{m=1}^{k} \binom{k-1}{m-1} p(q)^{m-1}(1-q)^{k-m} \cdot \frac{1}{m}.$$

B ADDITIONAL ANALYTICAL RESULTS FOR THE ASR *VS.* MSE TRADE-OFF

To complement the results of Figure 4 (medium to low privacy regimes and small domain size) in Section 6.4, Figures 8 to 12 illustrates the ASR-MSE trade-off considering:

- Figure 8: high privacy regime and small domain size.
- Figure 9: high privacy regime and medium domain size.
- Figure 10: high privacy regime and large domain size.
- Figure 11: medium to low privacy regimes and medium domain size.
- Figure 12: medium to low privacy regimes and large domain size.

C ADDITIONAL EMPIRICAL RESULTS FOR THE ASR *VS.* MSE TRADE-OFF

To complement the results presented in Figure 7 (**Adult** dataset [19] with $n = 48842$ and $k = 100$ for the Age attribute), we now conduct experiments using a synthetic dataset generated from a Dirichlet distribution with parameter **1**. Specifically, we compare empirical results obtained with varying numbers of users against analytical predictions computed using our closed-form equations from Section 5.3. Figure 13 presents the **analytical** Pareto frontier for ASR vs. MSE for SS, OUE, OLH, and THE against our adaptive counterparts, *i.e.*, ASS, AUE, ALH, and ATHE, providing a baseline for evaluating the empirical results. Meanwhile, Figure 14 presents the empirical Pareto frontiers for the same protocols across different user counts ($n \in \{5000, 50000, 500000\}$). Each subplot of Figures 13 and 14 evaluates the trade-off between ASR and MSE under varying privacy budgets $\varepsilon \in (2, 10)$ and a fixed domain size $k = 100$.

Consistency across user counts: Notably, Figure 14 shows that while the absolute variance (MSE) scales inversely with the number of users, the overall shape of the ASR-MSE Pareto frontier remains consistent across different values of n. This behavior aligns with theoretical expectations, as increasing the number of users reduces the variance but does not alter the fundamental trade-off between privacy and utility.

Empirical and analytical alignment: Similar to the results in Figure 7, the empirical ASR-MSE trends in Figure 14 closely follow the analytical ones in Figure 13, reinforcing the validity of our closed-form equations across various dataset distributions and user settings. These findings confirm that our analytical framework provides a reliable approximation for real-world deployments of LDP protocols, allowing practitioners to anticipate privacy-utility trade-offs without requiring extensive empirical evaluations.

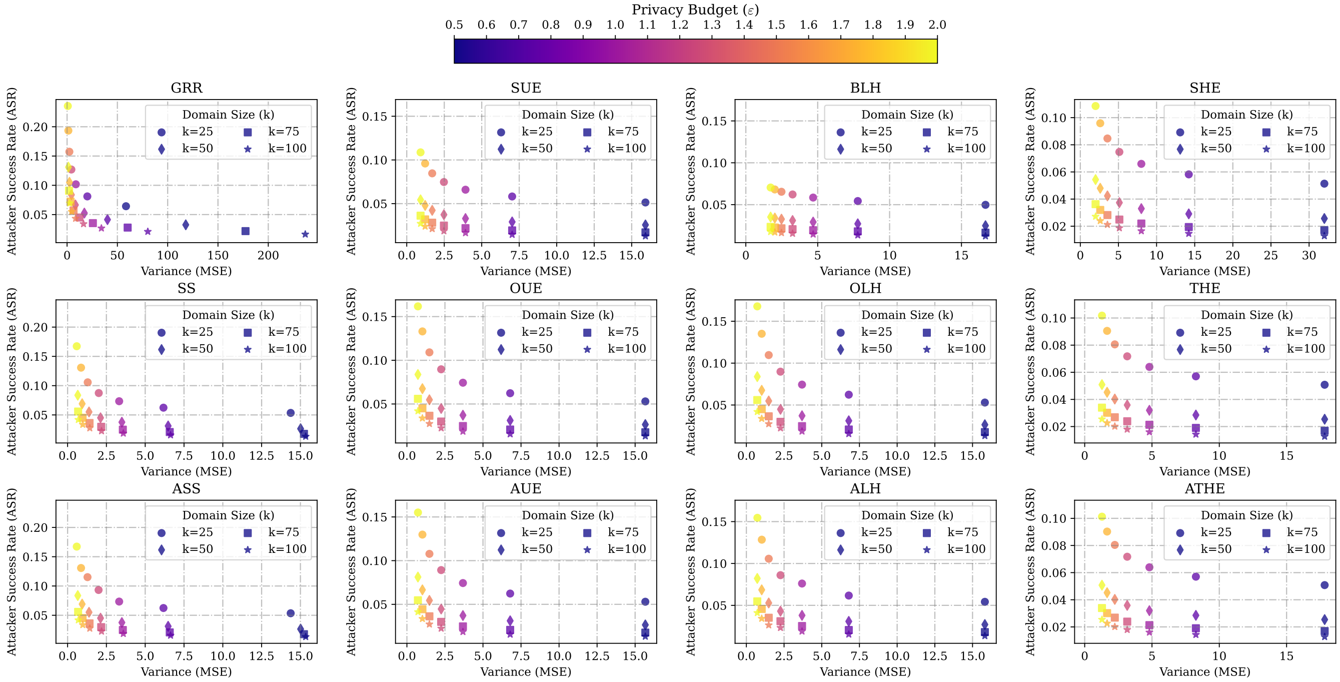


Figure 8: Attacker Success Rate (ASR) *vs.* Variance (MSE) for numerous LDP frequency estimation protocols. Each plot shows how each protocol performs under varying privacy budgets ε and domain sizes (k), illustrating the trade-off between adversarial success rate (ASR) and utility (MSE). State-of-the-art LDP protocols (*i.e.*, GRR, SUE, BLH, SHE, SS, OUE, OLH, and THE) are compared against our adaptive counterparts (*i.e.*, ASS, AUE, ALH, and ATHE). Each point represents a different configuration of ε (in high privacy regimes) and k (small domain), with colors indicating the privacy budget level.

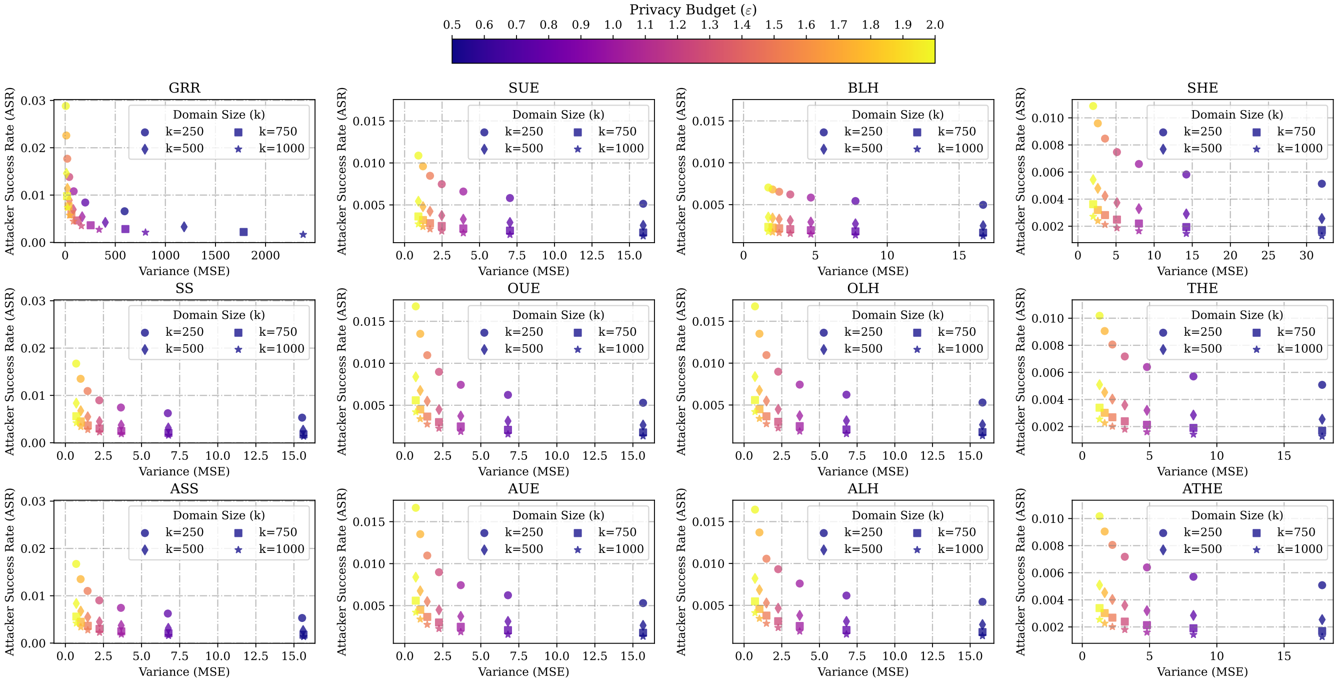


Figure 9: Attacker Success Rate (ASR) *vs.* Variance (MSE) for numerous LDP frequency estimation protocols. Each plot shows how each protocol performs under varying privacy budgets ε and domain sizes (k), illustrating the trade-off between adversarial success rate (ASR) and utility (MSE). State-of-the-art LDP protocols (*i.e.*, GRR, SUE, BLH, SHE, SS, OUE, OLH, and THE) are compared against our adaptive counterparts (*i.e.*, ASS, AUE, ALH, and ATHE). Each point represents a different configuration of ε (in high privacy regimes) and k (medium domain), with colors indicating the privacy budget level.

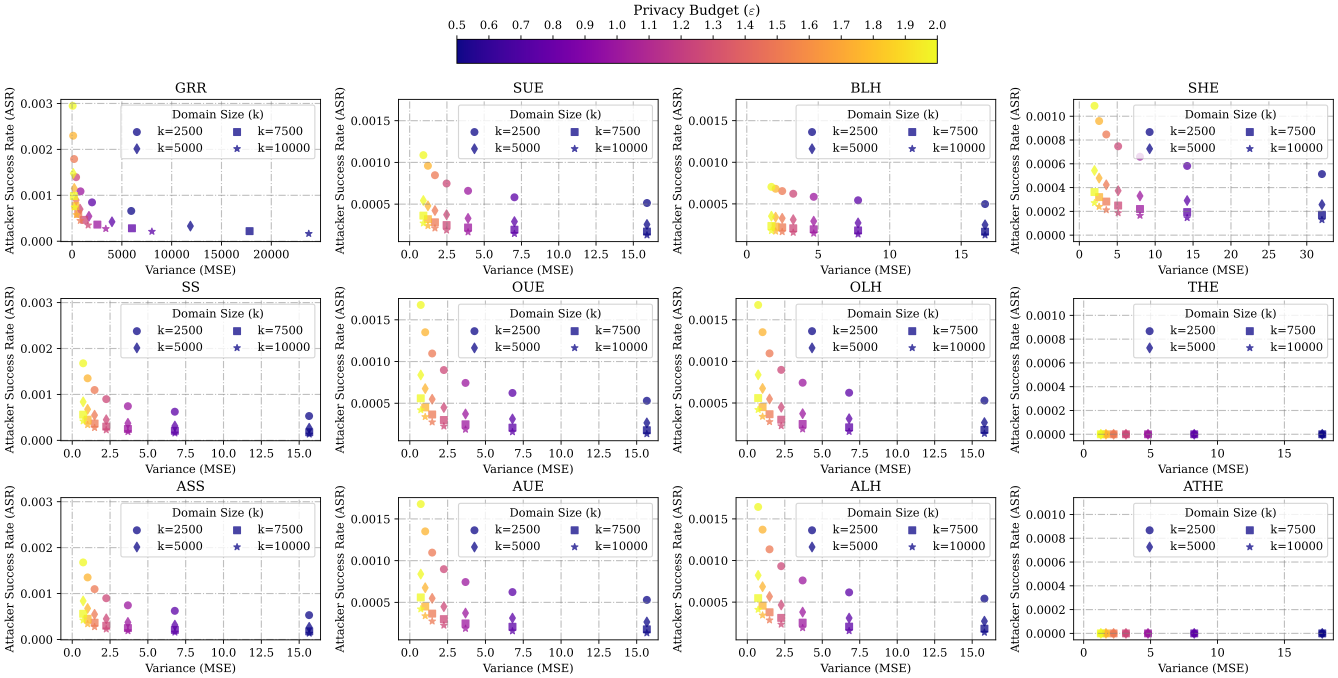


Figure 10: Attacker Success Rate (ASR) *vs.* Variance (MSE) for numerous LDP frequency estimation protocols. Each plot shows how each protocol performs under varying privacy budgets ε and domain sizes (k), illustrating the trade-off between adversarial success rate (ASR) and utility (MSE). State-of-the-art LDP protocols (*i.e.*, GRR, SUE, BLH, SHE, SS, OUE, OLH, and THE) are compared against our adaptive counterparts (*i.e.*, ASS, AUE, ALH, and ATHE). Each point represents a different configuration of ε (in high privacy regimes) and k (large domain), with colors indicating the privacy budget level.

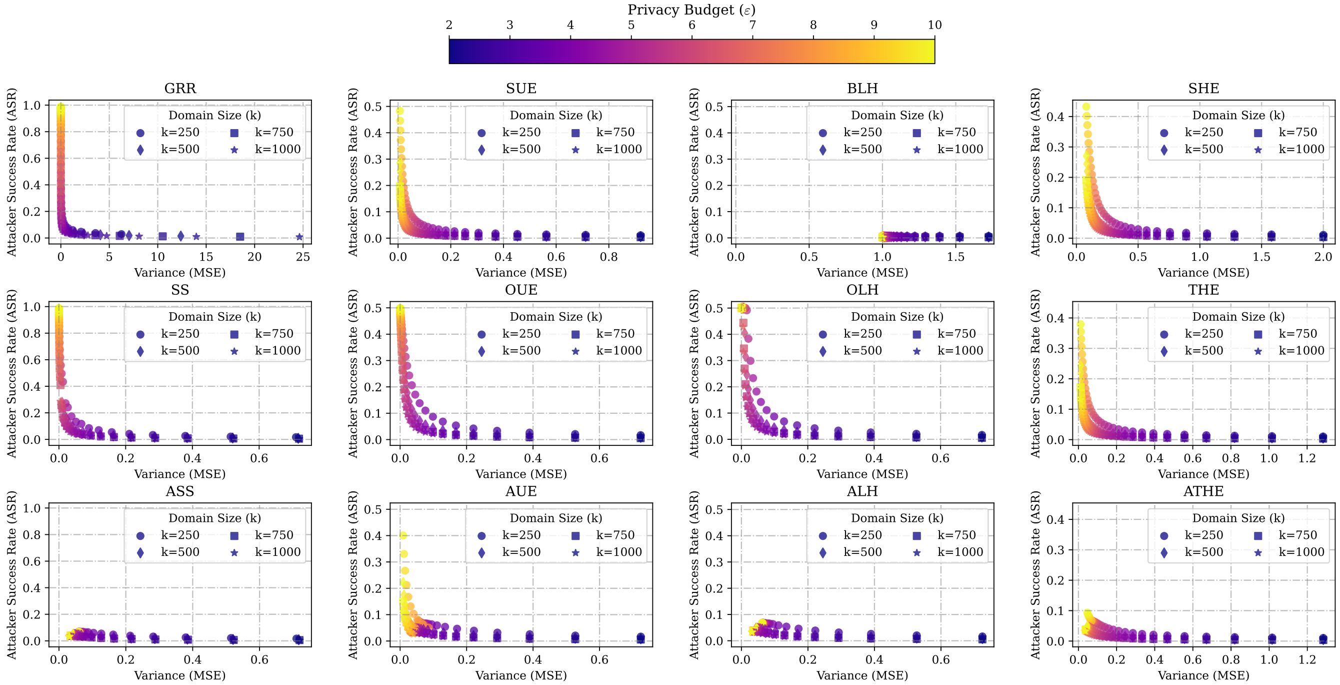


Figure 11: Attacker Success Rate (ASR) *vs.* Variance (MSE) for numerous LDP frequency estimation protocols. Each plot shows how each protocol performs under varying privacy budgets ε and domain sizes (k), illustrating the trade-off between adversarial success rate (ASR) and utility (MSE). State-of-the-art LDP protocols (*i.e.*, GRR, SUE, BLH, SHE, SS, OUE, OLH, and THE) are compared against our adaptive counterparts (*i.e.*, ASS, AUE, ALH, and ATHE). Each point represents a different configuration of ε (in medium to low privacy regimes) and k (medium domain), with colors indicating the privacy budget level.

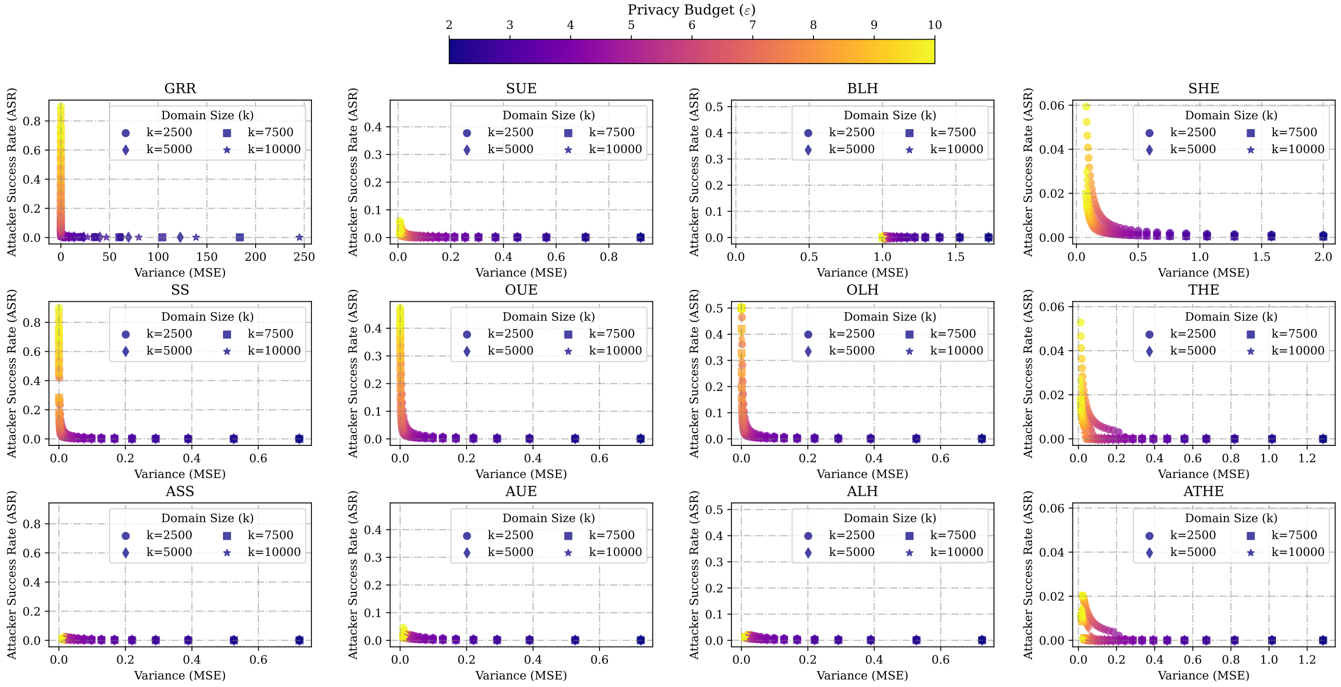


Figure 12: Attacker Success Rate (ASR) *vs.* Variance (MSE) for numerous LDP frequency estimation protocols. Each plot shows how each protocol performs under varying privacy budgets ε and domain sizes (k), illustrating the trade-off between adversarial success rate (ASR) and utility (MSE). State-of-the-art LDP protocols (*i.e.*, GRR, SUE, BLH, SHE, SS, OUE, OLH, and THE) are compared against our adaptive counterparts (*i.e.*, ASS, AUE, ALH, and ATHE). Each point represents a different configuration of ε (in medium to low privacy regimes) and k (large domain), with colors indicating the privacy budget level.

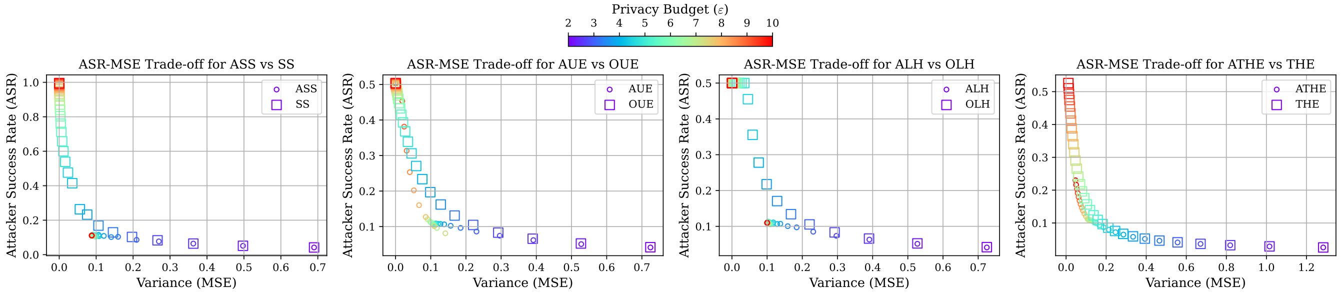
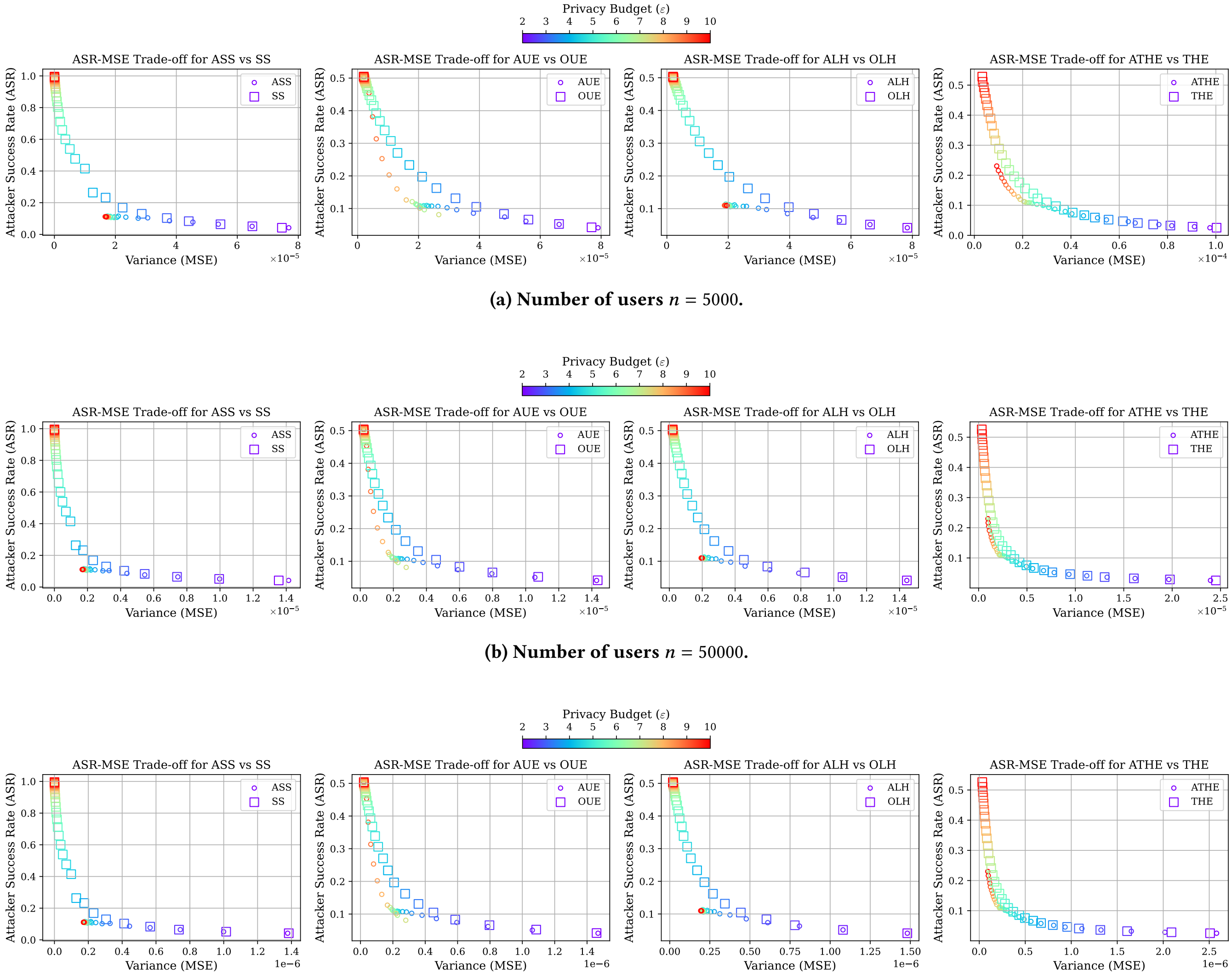


Figure 13: Analytical comparison of ASR *vs.* MSE Pareto frontier for four state-of-the-art LDP protocols (*i.e.*, SS, OUE, OLH, and THE) and our proposed adaptive versions (*i.e.*, ASS, AUE, ALH, ATHE). Each subplot considers a range of privacy budgets $\varepsilon \in (2, 10)$ and a fixed domain size $k = 100$.



(a) Number of users $n = 5000$.

(b) Number of users $n = 50000$.

(c) Number of users $n = 500000$.

Figure 14: Empirical comparison of ASR *vs.* MSE Pareto frontier for four state-of-the-art LDP protocols (*i.e.*, SS, OUE, OLH, and THE) and our proposed adaptive versions (*i.e.*, ASS, AUE, ALH, ATHE). Each subplot considers a range of privacy budgets $\varepsilon \in (2, 10)$, a fixed domain size $k = 100$, and varying numbers of users $n \in \{5000, 50000, 500000\}$. The dataset follows a Dirichlet distribution with parameter 1. Results are averaged over 100 independent runs.