Hiding Numerical Vectors in Local Private and Shuffled Messages

Shaowei Wang1, Jin Li1,*, Yuqiu Qian2, Jiachun Du2, Wenqing Lin2, Wei Yang3

1Institute of Artificial Intelligence and Blockchain, Guangzhou University
2Interactive Entertainment Group, Tencent Inc.
3Department of Computer Science and Technology, University of Science and Technology of China
{wangsw,lijin}@gzh.edu.cn, {yuqiuqian,kevinjcd,edwlin}@tencent.com, qubit@ustc.edu.cn

Abstract

Numerical vector aggregation has numerous applications in privacy-sensitive scenarios, such as distributed gradient estimation in federated learning, and statistical analysis on key-value data. Within the framework of local differential privacy, this work gives tight minimax error bounds of $O\left(\frac{d^2}{n^2}\right)$, where $d$ is the dimension of the numerical vector and $s$ is the number of non-zero entries. An attainable mechanism is then designed to improve from existing approaches suffering error rate of $O\left(\frac{d^2}{n^2}\right)$ or $O\left(\frac{d^2}{n^2}\right)$. To break the error barrier in the local privacy, this work further consider privacy amplification in the shuffle model with anonymous channels, and shows the mechanism satisfies centralized $(\sqrt{14\ln(2/\delta)}\frac{se^{c+2s-1}}{n-1},\delta)$-differential privacy, which is domain independent and thus scales to federated learning of large models. We experimentally validate and compare it with existing approaches, and demonstrate its significant error reduction.

1 Introduction

With the enacting of increasingly rigid regulations on data privacy (e.g., the General Data Protection Regulation [Voigt and Von dem Bussche, 2017] in the Europe Union, the Consumer Privacy Act, and the Civil Code of the People’s Republic of China), local differential privacy (LDP) has become the de facto notion for data privacy preservation over the Internet. It originates from the classical notion of differential privacy in the database community [Dwork, 2008] without the trust of the data aggregator or other third parties. LDP allows every user/agent to sanitize their personal data locally (e.g., on mobile devices, IoT sensors or edge servers) and provides information-theoretically rigorous privacy protection. Currently, many giant internet service providers (such as Apple [Greenberg, 2016], Google [Erlingsson et al., 2014] and Microsoft [Ding et al., 2017]) are deploying LDP for regulation compliance when collecting and analyzing user data. As a remedy to the unacceptable error barrier due to stringent LDP constraints, researchers recently introduce shuffle model [Erlingsson et al., 2019] where messages from users are permuted (by a shuffler, e.g., anonymous channels) before sent to the aggregator. The linkage between users and their messages are cut off and messages could hide in the crowd. [Erlingsson et al., 2019] show privacy is amplified with shuffling, thus a lower privacy level can be adopted locally to satisfy a relatively higher privacy level as in the analogised central model.

Plenty of user data are in the form of numerical vectors. Let $x^i$ denote the numerical vector of user $i$. For simplicity but without loss of generality, $x^i$ can be assumed as a $d$-dimensional $s$-sparse ternary vector [Wen et al., 2017; Ye et al., 2019; Sun et al., 2019; Gu et al., 2020], that is:

$$X^i = \{x \mid x \in \{-1, 0, 1\}^d \text{ and } \|x\|_0 = s\}.$$

This work studies the problem of numerical vector aggregation within the local and shuffled differential privacy framework. Many real-world data aggregation tasks could be formulated as this problem, such as gradient estimation in federated learning and sensitive key-value data aggregation for user profile/usage analyses in online services.

1.1 Federated Gradient Estimation

Federated learning [Konečný et al., 2016] studies machine learning systems in the distributed setting so that each party keeps its own data locally for privacy preserving. At each gradient descent iteration for training/updating a machine learning model, locally computed gradients $x^i$ from participating parties (e.g., from $n$ mobile users) need to be averaged by the federation server (e.g., a parameter server):

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x^i. \quad (1)$$

For communication efficiency, local gradients are usually discretized and sparsified [Wen et al., 2017; Wangni et al., 2018].

The original work [Konečný et al., 2016] deems sharing gradient to be more privacy-resistant than sharing raw data, but a recent work demonstrates that the gradient $x^i$ is also privacy risky [Zhu et al., 2019] and local raw data might be derived with confidence from several transmitted gradients. This calls for rigid privacy protection on local gradients.
1.2 Key-value Data Aggregation
We refer to key-value data as any paired (key, value) mappings, where the key \( j \in [1, d] \) is an index and the value \( x_j \) is numerical. Note that, the value is deemed as 0 when and only when the corresponding key is missing from or not defined in key-value data. For any existing or defined keys, their corresponding values are binary as \( \{-1, 1\} \). For example, a user might represent preferences on watched movies as key-value data, in which movies the user likes are assigned with value 1 and movies the user dislikes are assigned with value -1.

Common analysis on key-value data includes estimating both unconditional mean statistics and conditional mean statistics. The unconditional mean estimation about the key \( j \) is
\[
x_j = \frac{1}{n} \sum_{i=1}^{n} x_{ij},
\]
and the conditional mean estimation about the key \( j \) is
\[
x_j^c = \frac{\sum_{i=1}^{n} x_{ij}^c}{\# \{ x_{ij}^c \mid x_{ij}^c \text{ for } i \in [1, n] \text{ and } x_{ij}^c \neq 0 \}}.
\]

1.3 Existing Results
Within the framework of \( \epsilon \)-LDP, theoretical minimax lower bounds for many statistical estimation problems have been established, such as multinomial distribution estimation [Duchi et al., 2013], logistic regression/generalized linear model estimation [Duchi et al., 2018] and sparse covariance matrix estimation [Wang and Xu, 2019]. Specifically, the work of [Duchi et al., 2018] derives minimax lower bounds for multi-dimensional mean estimation for numerical vectors with bounded \( \ell_1 \)-norm or \( \ell_2 \)-norm. However, an \( s \)-sparse numerical vector with bounded \( \ell_1 \)-norm or \( \ell_2 \)-norm is a special case of [Duchi et al., 2018] with identical absolute non-zero entries. Whether it holds the same bounds as the general case or has tighter bounds is still an open question. Recently, for a broad family of \( \epsilon \)-LDP estimation problems that can be cast as a mean estimation one, the work of [Blasiok et al., 2019] studies sample complexity lower bounds under certain error tolerance \( \alpha \), but their sample complexity result for \( s \)-sparse numerical vectors has at least a \( 1/\alpha \) gap to ours result of minimax optimal sample complexity.

Practically, plenty of \( \epsilon \)-LDP mechanisms have been proposed for statistical data estimation, such as multinomial distribution estimation on categorical data [Duchi et al., 2013; Erlingsson et al., 2014; Kairouz et al., 2016; Wang et al., 2020b], and one-dimensional mean estimation on numerical values [Wang et al., 2019; Sun et al., 2020]. For \( \epsilon \)-LDP numerical vector or key-value data aggregation, existing approaches deal with both dense numerical vectors (e.g., in [Nguyễn et al., 2016; Duchi et al., 2018]) and sparse numerical vectors (e.g., in [Ye et al., 2019; Sun et al., 2019; Gu et al., 2020]). Specifically, the works of [Ye et al., 2019; Sun et al., 2019] uniformly randomly select one dimension from \([1, d]\) and transform the multi-dimensional estimation problem to a one-dimensional numerical/categorical problem. The work of [Gu et al., 2020] follows a similar paradigm, but randomly selects one non-empty dimension from \( s \) dimensions. However, as we will show in Section 4, these mechanisms are sub-optimal.

To mitigate the high noise needed for LDP, [Erlingsson et al., 2019] introduces a trusted shuffler to hide private views in the crowd. Recent works [Balle et al., 2020; Ghazi et al., 2020] propose sending multiple unary-binary messages with distributed noise to the shuffler for achieving centralized differential privacy (CDP), while other works [Cheu et al., 2019; Balle et al., 2019] study privacy amplification effects for achieving a lower level of LDP and a higher level of CDP simultaneously via shuffling. Specifically, [Balle et al., 2019; Wang et al., 2020a; Liu et al., 2021] utilize the technique of privacy blanket for privacy amplification on binomial/multinomial distribution estimation. This work further show domain-independent privacy amplification is achievable for sparse numerical vector.

1.4 Our Contributions
Minimax Lower Bounds. The MSE lower bound of \( \epsilon \)-LDP \( s \)-sparse numerical vector mean estimation is \( O(\frac{d_n}{\epsilon^2}) \). Our proof considers \( s \)-sparse numerical vectors that are decomposable, hence reduces the bounding procedure to cases of multiple multinomial distribution estimations.

An Optimal Mechanism. Since existing approaches are sub-optimal, we design a mechanism that matches the minimax lower bound. The mechanism has computational complexity of \( O(s) \) and communication complexity of \( O(\log s) \).

Domain-independent Privacy Amplification. For the shuffle model of \( s \)-sparse numerical vector aggregation, the proposed optimal mechanism satisfies centralized \((\sqrt{14\ln(2/\delta)} \frac{s + 2s - 1}{n - 1}, \delta)\)-differential privacy. The privacy loss \( \epsilon_c \) in CDP is independent of domain size \( d \), thus fits federated learning of large models. In turn, when privacy budget \( \epsilon_c \) is given, we derive local parameter for optimal utility and show the proposed mechanism is asymptotic near-optimal in terms of user population.

The remainder of the paper is organized as follows. Section 2 provides preliminary knowledge on local differential privacy and minimax risk framework of statistical estimation. The minimax lower bound on the \( \epsilon \)-LDP numerical vector aggregation problem is then given in Section 3. Next, Section 4 reviews the design of existing mechanisms and shows their sub-optimality, and then propose a new mechanism and prove its optimality. Section 5 shows the proposed mechanism enjoys domain-independent privacy amplification in the shuffle model, and prove its asymptotic optimality. Later, Section 6 demonstrates the superior performance of the proposed mechanism against existing mechanisms. Finally, Section 7 concludes the paper.

2 Preliminaries
2.1 Differential Privacy
For datasets \( D, D' \) that are of the same size and differ only in one element, they are called neighboring datasets. The centralized differential privacy with budget/level \((\epsilon, \delta)\) is as follows.

**Definition 1** \((\epsilon, \delta)\)-CDP [Dwork, 2008]. Let \( D_K \) denote the output domain, a randomized mechanism \( K \) satisfies \( \epsilon \)-differential privacy iff for any neighboring datasets \( D, D' \), and any outputs \( z \subseteq D_K \),
\[
P[K(D) \in z] \leq \exp(\epsilon) \cdot P[K(D') \in z] + \delta.
\]
Let $K$ denote a randomized mechanism for sanitizing a single user data, the LDP with privacy budget $\epsilon$ is as follows.

**Definition 2** ($\epsilon$-LDP [Duchi et al., 2013]). Let $D_K$ denote the output domain, a randomized mechanism $K$ satisfies local $\epsilon$-differential privacy iff for any data pair $x, x' \in X^*$, and any output $z \in D_K$,

$$\mathbb{P}[K(x) = z] \leq \exp(\epsilon) \cdot \mathbb{P}[K(x') = z].$$

### 2.2 Local Private Minimax Risks

Assuming samples $\{x^1, x^2, \ldots, x^n\}$ that are $n$ i.i.d. drawn from a distribution $P \in \mathcal{P}$. Let $K_n$ denote the set of all possible mechanisms $K = \{K^1, K^2, \ldots, K^K\}$ that satisfy $\epsilon$-LDP for every sample in $\{x^1, x^2, \ldots, x^n\}$. Taking as input the samples, some mechanism $K_i \in K_n$ produces a list of sanitized views $\{z^1, z^2, \ldots, z^n\}$. If the parameter estimator:

$$\hat{\theta} = \hat{\theta}(\{z^1, z^2, \ldots, z^n\})$$

is derived from these private views while having no access to input samples $\{x^j\}_{j=1}^n$, the minimax MSE risk (under privacy budget $\epsilon$) is then:

$$\mathcal{M}_n(\theta(P), \| \cdot \|_2^2, \epsilon) := \inf_{K \in K_n} \inf_{P \in \mathcal{P}} \sup \mathbb{E}_{P,K}[\|\hat{\theta}(z^1, z^2, \ldots, z^n) - \theta(P)\|_2^2].$$

### 3 Minimax Lower Bounds

The Assouad’s method [Yu, 1997] is a common tool for lower bounding via multiple hypothesis testing. It defines a hypercube $V = \{-1, 1\}^d$ ($d \in \mathbb{N}$), then defines a family of distributions $\{P_x\}_{x \in V}$ indexed by the hypercube, where each $P_x$ is defined on a common space. It’s said that the distribution family induces a $2\tau$-Hamming separation for the loss $\|\cdot\|_2^2$ if there exists a vertex mapping function $\kappa : \theta(P) \mapsto \{-1, 1\}^d$ satisfying:

$$\|\theta - \theta(P)\|_2^2 \geq 2\tau \sum_{j=1}^d 1\{[\kappa(\theta)]_j \neq \nu_j\}.$$  

Assume that the nature first uniform-randomly chooses a vector $V \in \{-1, 1\}^d$, and the samples $\{x^1, \ldots, x^n\}$ are drawn from the distribution $P_V$ with $V = \nu$. These samples are then taken as input into $\epsilon$-LDP mechanisms $K$. The literature [Duchi et al., 2018] gives an $\epsilon$-LDP version of Assouad’s method as follows.

**Lemma 1** (Private Assouad bound [Duchi et al., 2018]). Let $P_{\nu + j} = \frac{1}{2^d} \sum_{\nu' : \nu' = \nu \oplus \text{vector}} P_\nu$ and $P_{\nu - j} = \frac{1}{2^d} \sum_{\nu' : \nu' = \nu \ominus \text{vector}} P_\nu$, we have

$$\mathcal{M}_n(\theta(P), \| \cdot \|_2^2) \geq d \cdot \tau[1 - \left(\frac{n^{(e-1)^2}}{2d}\right) F_{\mathbb{B}_\infty(X^*)}(P)]^{\frac{1}{2}},$$

where $\mathbb{B}_\infty(X^*)$ denote the collection of function $\gamma$ with supremum norm bounded by 1 as:

$$\mathbb{B}_\infty(X^*) := \{ \gamma : X^* \mapsto \mathbb{R} \mid \|\gamma\|_\infty \leq 1 \},$$

and maximum possible discrepancy $F_{\mathbb{B}_\infty(X^*)}$ is defined as:

$$\sup_{\gamma \in \mathbb{B}_\infty(X^*)} \sum_{i=1}^d \left( \int_{X^*} \gamma(x) (dP_{\nu + j}(x) - dP_{\nu - j}(x))^2 \right)^{\frac{1}{2}}.$$

We consider numerical vectors that can be decomposed into $s$ buckets, each bucket has $\frac{d}{2}$ indexes with only one non-zero entry. We then define a hypercube of length $d$ and construct a class of $2^{\frac{2s^2}{\epsilon^2}}$-hamming separated probability distributions. Guided by Lemma 1, we bound the maximum possible marginal distance $F_{\mathbb{B}_\infty(X^*)}$ under the value of $\frac{2s^2}{\epsilon^2}$. Theorem 1 gives the final lower bounds for the problem of local private numerical vector mean estimation.

**Theorem 1.** For the numerical vector aggregation problem, for any $\epsilon$-LDP mechanism, there exists a universal constant $c > 0$ such that for all $\epsilon \in [0, 1]$,

$$\mathcal{M}_n(\theta(P), \| \cdot \|_2^2, \epsilon) \geq c \cdot \min\{\frac{s^2}{d}, \frac{ds}{n\epsilon^2}\}. $$

To understand the minimax rate, we can consider the non-private error rate of decomposable numerical vector aggregation, which is $E[\|\theta - \theta(P)\|_2^2] \leq \sum_{i=1}^d E[\|\theta_i - \theta_i(P)\|_2^2] \leq \frac{ds}{d}$. Thus the enforcement of local $\epsilon$-LDP causes the effective sample size decreasing from $n$ to $n\epsilon^2/d$.

### 4 Optimal Mechanism

Let $j_+$ and $j_-$ denote events that the $j$-th element of $x$ (i.e., $x_j$) equals to $-1$ and $1$ respectively, a numerical vector $x$ could be represented in the set form as:

$$X_\nu = \{j- \mid j \in [1, d] \text{ and } x_j = -1\} \cup \{j+ \mid j \in [1, d] \text{ and } x_j = 1\}.$$

Existing works on $\epsilon$-LDP numerical vector aggregation can be categorized into two types, they do dimension sampling in a data-agnostic manner (e.g., the PrivKV in [Ye et al., 2019; Sun et al., 2019]) or a data-dependent manner (e.g., the PCKV in [Gu et al., 2020]).

**The PrivKV Mechanism.** The seminal work [Ye et al., 2019] on $\epsilon$-LDP key-value data aggregation propose to firstly randomly sample a dimension (from the domain of keys) $j \in [1, d]$, then applies an $\epsilon$-LDP categorical mechanism on the corresponding (key,value) pair that takes a value from $\{(j, 0), (j, 1), (j, -1)\}$, where $(j, 0)$ means that the key is empty in the key-value data. Essentially, the PrivKV mechanism is equivalent to dividing the population of $n$ into $d$ groups, and each group is employed to estimate $\{j_+ \in Y_\nu \} \text{ and } \{j_- \in Y_\nu \}$ for each $j \in [1, d]$ with privacy budget $\epsilon$. Since the minimax lower error bound for estimating frequencies on population of $n'$ with privacy budget $\epsilon$ and domain size $d'$ is $\Theta(\frac{n'}{\epsilon^2 d'})$ [Duchi et al., 2018], the estimation error of $\{j_+ \in Y_\nu \} \text{ and } \{j_- \in Y_\nu \}$ is hence $\Theta(\frac{n'}{\epsilon^2 d'})$, as $n' = \frac{n}{d}$ and $d' = 3$. Therefore, its total estimation error of frequencies or mean values of $d$-dimensional vector is $O(\frac{d^2}{\epsilon^2 n})$. It has a gap of $\frac{2}{3}$ from the optimal error rate in Theorem 1. Similar methodology and result also hold for the following-up works in [Sun et al., 2019; Liu et al., 2021].

**The PCKV Mechanism.** The work of [Gu et al., 2020] proposes to sample one key from existing $s$ keys in a key-value data. Afterwards, an $\epsilon$-LDP categorical mechanism is applied to the corresponding 1-sparse numerical vector, which is equivalent to categorical data with domain size of around $2d$. Recall that the minimax lower error bound for estimating
The normalization factor is $Y(s)$ with size $s$.

The paradigm of dimension sampling and categorical randomization fails to achieve optimal statistical rate for $\epsilon$-LDP numerical vector aggregation. Therefore, we consider randomizing the numerical vector as a whole with the exponential mechanism [McSherry and Talwar, 2007], and propose the Collision mechanism.

If defining event domain as: $Y = \{1, \ldots, n\}$, let $Y_X$ as a subset of $Y$ with size $s$. Define the output domain as $Z = \{1, 2, \ldots, t\}$, the Collision mechanism probabilistically outputs one item $z \in Z$. The outputting probabilities are based on whether each item has collision with hashed events in $Y_X$. The Collision mechanism is formally given in Definition 3.

**Definition 3** ($(d, s, e, t)$-Collision Mechanism). Given a random-chosen hash function $H : Y \mapsto Z$, take an $s$-sparse numerical vector $Y_X \subseteq Y$ as input, the Collision mechanism probabilistically outputs an element $z \in Z$ according to following probability design:

$$
P[z|X] = \begin{cases} 
\frac{e^y}{\Omega} & \text{if } \exists y \in Y_X, z = H(y); \\
\frac{(1 - e^{-y})}{\Omega} & \text{if } \forall y \in Y_X, z \neq H(y); \\
\frac{t - \Omega(y)}{\Omega} & \text{otherwise}.
\end{cases}
$$

The normalization factor is $\Omega = s \cdot e^r + t - s$. An unbiased estimator of indicator $[j_b \in Y_X]$ for $b \in \{-1, 1\}$ and $j \in [1, d]$ is $(s \geq 2)$:

$$
[j_b \in Y_X] = \frac{[H(j_b) = z] - 1/t}{e^r/\Omega - 1/t}.
$$

The privacy guarantee of the mechanism is given in Proposition 1, which is obvious as $\|H(y)\| = n$ for $y \in Y_X$. The utility-optimality guarantee of the mechanism is given in Theorem 2. For $\epsilon = O(1)$, its computational complexity is bounded by $t^* \approx s + 2s - 1 + s \cdot e^r = O(s)$, and communication complexity is $\log_d(2s - 1 + s \cdot e^r) = O(\log s)$.

**Proposition 1.** The $(d, s, \epsilon, t)$-Collision mechanism in Definition 3 satisfies $\epsilon$-LDP for numerical vector data.

**Theorem 2.** Given privacy budget $\epsilon = O(1)$, with optimal choice of the parameter $t^*$, the mean estimation error of $(d, s, \epsilon, t)$-Collision mechanism for numerical vector is $O(\frac{ds}{ne^r})$.

**Proof.** Since $[H(j_b) = z]$ are Bernoulli random variables, we have the mean squared error:

$$
\Var[\tilde{X}] \leq 2 \sum_{j=1}^{d} \sum_{b \in \{-1, 1\}} \Var[[j_b \in Y_X]]
$$

$$
\leq \frac{2}{n} \cdot \frac{s \cdot e^r/\Omega(1 - e^r/\Omega) + (2d - s) \cdot 1/t(1 - 1/t)}{(e^r/\Omega - 1/t)^2}.
$$

Taking the previous formula as a function of continuous $t$, actually the function is convex when $d \geq t \geq s$. Choosing approximate optimal $t^*$ at around $2s - 1 + s \cdot e^r$, we then have (with $e^r \approx e + 1$):

$$
\Var[\tilde{X}] \leq 2d \cdot \Theta(s^3) + \epsilon \cdot \Theta(s^3)
$$

$$
\leq O\left(\frac{ds}{ne^2}\right).
$$

**5 Privacy Amplification in Shuffle Model**

When semi-trusted shufflers lie between users and the aggregator, the aggregator only observes the multi-set of private views $\{z_1, z_2, \ldots, z_n\}$, thus the privacy level of some LDP mechanisms is amplified (w.r.t. CDP). For binary domain of private views $Z = \{0, 1\}$ generated from randomized response [Cheu et al., 2019; Balle et al., 2019] show the observed frequencies of $z \in Z$ satisfies $\left(\sqrt{14\ln(2/\delta)} - s e^r/\Omega - 1\right)$-CDP, since about $2n(1 - 1/s + 1/t)$ users response uniform randomly, they contributed $B\left(\frac{2n(1 - 1/s + 1/t)}{s e^r/\Omega - 1}, \frac{1}{2}\right)$ Binomial random noises to the frequencies.

Similarly, in the Collision mechanism with private view domain $Z = \{0, 1, \ldots, t\}$, about $\frac{n}{2} - \frac{1}{s e^r/\Omega - 1}$ users response uniform randomly, they contributed $B\left(\frac{2n(1 - 1/s + 1/t)}{s e^r/\Omega - 1}, \frac{1}{2}\right)$ Binomial random noise to each frequency of $z \in Z$. As a result, the Collision mechanism also satisfies $\left(\sqrt{14\ln(2/\delta)} - s e^r/\Omega - 1\right)$-CDP. The formal guarantee of privacy amplification of the Collision mechanism is presented in Theorem 3.

**Theorem 3.** In the shuffle model, the $(d, s, \epsilon, t)$-Collision mechanism satisfies $\left(\sqrt{14\ln(2/\delta)} - s e^r/\Omega - 1\right)$-CDP.

**Proof.** Since private views with hash functions are randomly shuffled and the final estimator is derived from them, to prove the final estimator is $\left(\sqrt{14\ln(2/\delta)} - s e^r/\Omega - 1\right)$-CDP, it is enough to show that the distribution of observed private views with hash functions satisfies CDP. Further because the hash functions are randomly chosen from some universe, we only need to show the frequency distribution of observed private views satisfies CDP. The frequency distribution is noised by $N \sim B(n - 1, \frac{t}{s e^r/\Omega - 1})$ users, each user contributed with uniform-random $t$-multinomial distribution in the output domain. Then, according to the tail bounding result on this noise distribution with multiplicative Chernoff bound or Bennett’s inequality (the privacy blanket theorem in [Balle et al., 2019]), such a noise distribution satisfies $\sqrt{14\ln(2/\delta)} - s e^r/\Omega - 1\right)$-CDP with probability at least $1 - \delta$ when $n \geq \frac{2t^2(e^r/\Omega - 1)}{14\ln(2/\delta) - s e^r/\Omega - 1} + 1$. 

With optimal choice of $t \approx 2s - 1 + s \cdot e^r$ in LDP, the privacy amplification bound is thus independent of the domain size $d$. Alternatively, when the privacy level in the CDP is given as $\epsilon$, then $\omega = s \cdot e^r + t - s = \frac{e^r(n - 1)}{14\ln(2/\delta)}$, then
the numerical vector estimation error bound is a function of $t$ with constant $\Omega$ as:

$$Var[\hat{X}] \leq \frac{(2d-s)(t-1)}{t^2} \frac{\Omega(\xi^2-s^2)(2s^2+3s)}{12s^2}.$$ 

Choosing $t \approx 4+\Omega+s+\sqrt{12(7s^2-8)+s^2-16s+16}$ approximately minimizes the error, which is also independent of the domain size $d$. With the optimal choice of $t$ in terms of CDP, when the numerical vector is highly sparse (i.e. $d \gg s$) and the user population is large, we have the asymptotic estimation error in the shuffle setting as:

$$Var[\hat{X}] = O\left(\frac{ds^2\ln 1/\delta}{n^2\epsilon^2} \cdot \max\{1, \frac{14s\ln 2/\delta}{\epsilon^2(n-1)}\}\right),$$

which nearly matches the optimal error rate in the centralized differential privacy setting (the sensitivity is $2s$).

6 Experiments

The statistical efficiency of the proposed Collision mechanism for $\epsilon$-LDP numerical vector aggregation is evaluated in this section. Competing mechanisms include the PrivKV mechanism [Ye et al., 2019], the PCKV mechanism with general randomized response as the base randomizer (denoted as PCKV-GRR), and the PCKV mechanism with unary encoding as the base randomizer [Gu et al., 2020] (denoted as PCKV-UE). Since the performances of all these mechanisms are data-independent, it is enough to utilize synthetic datasets for fair evaluation. The parameters of synthetic datasets are listed as follows (default values are in bold form), covering most cases encountered in real-world applications:

i. Number of users $n$: 10,000, 100,000.

ii. Dimension $d$: 256, 1024.

iii. Sparsity parameter $s$: 4, 8, 16, 32.

iv. Privacy budget $\epsilon$: 0.001, 0.01, 0.1, 0.2, 0.4, 0.8, 1.0, 1.5, 2.0, 2.5, 3.0.

During each simulation, the numerical vector of each user is independent-randomly generated, the non-zero entries are uniform-randomly selected from $d$ dimensions, and each dimension has an equal probability of being $-1$ or $1$.

6.1 Evaluation Metric

Previous theoretical results focus on the mean squared error: $\sum_{j \in [1,d]} ||X_j - \hat{X}_j||^2$. Here we evaluate mechanisms with metrics on frequency estimators of $\{Y_j \in X\}$ (including TVE and MAE), which are basic statistics for both the unconditional and conditional mean estimation in Equation (1) and Equation (2). The total variation error (TVE) is defined as:

$$TVE = \sum_{j \in [1,d], b \in \{-1,1\}} ||Y_j - \hat{Y}_j||_1,$$

and the maximum absolute error (MAE) is defined as:

$$MAE = \max_{j \in [1,d], b \in \{-1,1\}} ||Y_j - \hat{Y}_j||_1.$$

Since the $\frac{1}{2}$-scaled frequencies lie in the $d$ dimensional probability simplex, the estimated frequencies are projected into the $\Delta_d$-simplex as in [Wang and Carreira-Perpinán, 2013]. All experimental results are the mean natural logarithm value of 10 repeated simulations.

6.2 Effects of Sparsity $s$

Assume that there are $n = 100,000$ users, and the dimension is $d = 256$. When the number of non-zero entries in numerical vectors varies from 4 to 32, the TVE/MAE error results are presented in Figure 1 and Figure 2 respectively. The PCKV-UE mechanism improves upon the PrivKV in the extreme sparse cases, but for other cases (e.g., $s = 32$), the PCKV-UE and the PrivKV mechanism have similar performances. The Collision mechanism outperforms all competing mechanisms in all cases significantly, and averagely reduces more than 60% errors. As the sparsity parameter $s$ gets larger, the performance gaps get larger.

6.3 Effects of Dimension $d$

Assume that there are $n = 100,000$ users, but the dimension now becomes $d = 1024$. When the number of non-zero entries in numerical vectors still varies from 4 to 32, the results of TVE and MAE are shown in Figure 3 and Figure 4 respectively. Compared to cases of $d = 256$ (i.e. TVE results in Figure 1 and MAE results in Figure 2), it is easy to observe that the TVE/MAE value grows with around $\sqrt{d}$.
6.4 Effects of Number of Users $n$

Assume that there are only $n = 10,000$ users, and the dimension is $d = 256$. When the number of non-zero entries in numerical vectors varies from 4 to 32, the results of TVE and MAE are listed in Figure 5 and Figure 6 respectively. Compared to the case of $n = 100,000$ (i.e., Figure 1 and Figure 2), the TVE/MAE value is about $\sqrt{100000}/10000$ times larger (i.e. decreases with around $\sqrt{n}$).

6.5 After Shuffling

Considering the privacy budget is amplified in the shuffle model, typically when the number of users is $n = 100000$, privacy budget in CDP is $\epsilon_c = 0.5$ and $\delta = 1/n$, the local privacy budget in the Collision mechanism is scaled to around $2.0$ with optimal $t$. In these regions with large local privacy budget, the performance of the Collision is far better than other approaches. Besides, the privacy amplification of PrivKV/PCKV scales poorly with dependence on the domain size, thus when CDP budget $(\epsilon_c, \delta)$ is given and the domain is relatively large, the performance gap grows.

6.6 Experimental Summary

Through experimental evaluation, we can conclude that the Collision mechanism outperforms existing approaches in all cases. Their performance gaps also support our previous theoretical analysis on error bounds (MAE errors usually have magnitude proportional to the root of mean squared error).

7 Conclusion

Within the framework of distributed differential privacy, this paper studied the problem of numerical vector statistical estimation, which has its applications in federated learning and key-value data aggregation. We provided tight minimax error bounds of $O\left(\frac{d}{\epsilon_n^2}\right)$ for local differential private mean estimation on numerical vectors. Our proof relies on a novel decomposition technique for data domain with sparse structure and an application of the local private version of Assouad methods. Given that existing approaches are suffering gaps form the optimal error bounds, we further design an optimal mechanism for the problem, and then give an efficient implementation with linear computation/communication complexity. To further break the error bounds, we consider numerical vector estimation in the shuffled differential privacy, and show the proposed mechanism has the advantages of domain-independent privacy amplification and near-optimal utility. Experimental results show averagely 60% error reduction of the optimal mechanism when compared with current approaches.

Acknowledgements

We would like to thank anonymous reviewers for their helpful comments. This work is supported by National Key Project of China (No. 2020YFB1005700), National Natural Science Foundation of China for Joint Fund Project (No. U1936218), Project of Guangzhou University (No. RQ2021006), and the Pazhou lab, Guangzhou, China.
References


