Analyzing Preference Data With Local Privacy: Optimal Utility and Enhanced Robustness

Shaowei Wang, Xuandi Luo, Yuqiu Qian, Jiachun Du, Wenqing Lin, and Wei Yang, *Member, IEEE*

Abstract—Online service providers benefit from collecting and analyzing preference data from users, including both implicit preference data (e.g., watched videos of a user) and explicit preference data (e.g., ranking data over candidates). However, it brings ethical and legal issues of data privacy at the same time. In this paper, we study the problem of aggregating individual's preference data in the local differential privacy (LDP) setting. One naive approach is to add Laplace random noises, which however suffers from low statistical utility and is fragile to LDP-specific poisoning attacks. Therefore, we propose a novel mechanism to improve the utility and the robustness simultaneously: the *additive mechanism*. The additive mechanism randomly outputs a subset of candidates with a probability proportional to their total scores. For preference data with Borda rule over d items, its mean squared error bound is optimized from $O(\frac{d^5}{n\epsilon^2})$ to $O(\frac{d^4}{n\epsilon^2})$, and its maximum poisoning risk bound is reduced from $+\infty$ to $O(\frac{d^2}{n\epsilon})$. We also theoretically investigate minimax lower bounds of ϵ-LDP preference data aggregation, and prove the error rate of $O(\frac{d^4}{n\epsilon^2})$ is optimal for the Borda rule. Experimental results validate that our proposed approaches averagely reduce estimation error by 50% and are more robust to adversarial poisoning attacks.

Index Terms—Data aggregation, differential privacy, minimax error, local privacy, data poisoning, preference data

1 Introduction

Preference data represents rich behaviours and interests information about online service users. Exposed with a list of candidate items (e.g., products in E-commerce, online videos, and webpage links), a user may selectively buy/click some of them. Considering that the user's monetary and time consumption may vary among these items, the preference data reflects the vary weights over items. In the preference data, a numerical score is assigned to each item according to its preferential position. For example, in a simple form of preference data: set-valued data, the selected items are assigned with 1 while other items are assigned with 0. Intuitively, any scoring strategy that are non-increasing is reasonable in preference data. Besides these implicit behavioural preference information aforementioned, users may also be explicitly surveyed to rank over candidates (e.g., in electronic votings). In such preferential ranking data, one prevalent scoring strategy is the Borda rule [1], which assigns score of $(d-i)$ to the i-th item, where d is the number of total candidates.

Aggregating and analyzing such preference/ranking data (e.g., in Table 1) to identify popular items helps online service providers to improve quality of services, but the risk of data privacy violation is becoming a non-negligible ethical/legal issue (w.r.t. the GDPR regulation [2] regulation and the CCPA act [3]). Various streams of works have contributed to privacy preservation for user data aggregation, such as cryptographic techniques (e.g., secure multi-party computation [4]) and perturbation techniques (e.g., centralized differential privacy [5]). However, cryptographic-based techniques have efficiency issues for large-scale data aggregation systems with millions of users, is fragile to collusion between the service provider and other users. The centralized differential privacy relies on the existence of a trustful data curator for all users, which is usually a unpractical consumption in distributed computer networks.

Researchers has proposed the local version of differential privacy (ϵ-LDP) [6] for data privacy in distributed networking (e.g., mobile cloud computing) environments. It sanitizes the user data locally on the user's side, and ensures up to $\exp(\epsilon)$ distinguishability on outputting probabilities no matter what the true data a user holds. Comparing privacy preserving ways mentioned above, LDP is information-theoretically rigid, computationally efficient and operationally flexible. The user has full controllability during the privacy preserving procedure without the trust of any parties; the service provider is also tolerable to users' unsynchronized opt-out, withdrawal and modification actions on contributing data. These advantages make LDP the best fit for large-scale user data aggregation under the recently enacted regulations of GDPR and CCPA. As a result, LDP has been widely adopted both in academia and in industry (e.g., Google [7], Microsoft [8], Apple [9], and SAP [10]).

- *Shaowei Wang and Xuandi Luo are with the Institute of Artificial Intelligence and Blockchain, Guangzhou University, Guangzhou, Guangdong 510006, China. E-mail: wangsw@gzhu.edu.cn, 2112006166@e.gzhu.edu.cn.*
- *Yuqiu Qian, Jiachun Du, and Wenqing Lin are with the Tencent Games, Shenzhen, Guangdong 518054, China. E-mail: {yuqiuqian, kevinjcdu, edwlin}@tencent.com.*
- *Wei Yang is with the University of Science and Technology of China, Hefei, Anhui 230052, China. E-mail: qubit@ustc.edu.cn.*

Manuscript received 27 February 2021; revised 20 May 2022; accepted 9 September 2022. Date of publication 19 September 2022; date of current version 21 June 2023.
This work was supported in part by the National Natural Science Foundation of China under Grants 62102108, 62102107, 62072132, and 62002074, in part by the Natural Science Foundation of Guangdong Province of China under Grant 2022A1515010061, and in part by Guangzhou Basic and Applied Basic Research Foundation under Grants 202201010194 and 622191-098.
(Corresponding author: Shaowei Wang.)
Recommended for acceptance by N. Zhang.
Digital Object Identifier no. 10.1109/TKDE.2022.3207486

TABLE 1
An Example of Aggregating Preference Data From 4 Users over 5 Candidate Items $\{Video_1, Video_2, ..., Video_5\}$ With the Borda Scoring Rule

	$Video_1$	$Video_2$	$Video_3$	$Video_4$	$Video_5$
$v^{(1)}$	2	3	4	1	0
$v^{(2)}$	0	4	3	1	2
$v^{(3)}$	0	3	2	1	4
$v^{(4)}$	4	3	1	0	2
Total Score	6	13	10	3	8
Average Score	1.5	3.25	2.5	0.75	2.0

For simplicity, users are assumed to be exposed with the same list of items. When the expose list is not the same, similar score aggregation and privacy preserving procedures can also applied, as the list is known by the service provider and hence is non-private.

One naive way to realize LDP is injecting Laplace random noises [11]. After representing the preference data in a vector form (e.g., $v^{(1)} = [2,\ 3,\ 4,\ 1,\ 0]$ in Table 1), Laplace noises of scale $\frac{\epsilon}{\Delta}$ are independently added to each score in a preference data, where ϵ is the privacy level and Δ is the maximum absolute difference between any two ranking data (see Lemma 1). Therefore, the preference data $v^{(1)}$ could become:

$$\tilde{v}^{(1)} = [26.6,\ -45.2,\ 6.3,\ -7.3,\ -1.5],$$

by choosing $\epsilon = 1.0$. It is easy to observe that the (unbiased) private view $\tilde{v}$ might far deviate from the true data v by such a way. Specifically, the expected deviation of the private view here is $\mathbb{E}[|\tilde{v} - v|_2^2] = \frac{2d\Delta^2}{\epsilon^2}$, which indicates preserving privacy comes at the cost of *data utility*, and improving the data utility under LDP is the central focus in the current literature [6], [12], [13], [14], [15].

In this work, we observe that privacy preserving also comes at the cost of *system robustness*. That is, the adversarial attackers can manipulate the aggregation result with data fraud/poisoning, especially when the privacy level gets higher. We identify two types of LDP-specific robustness attacks and summarize them in the following subsection.

1.1 Privacy Induced Attacks

According to the attacker's ability of bypassing the privacy preserving procedure, LDP induced attacks can be classified into data amplification attack and view disguise attack.

Data Amplification Attack. When the attacker is unable to skip the privacy preserving procedure, he(she) contributes fraud raw data. The effect of fraud data on the final aggregation result might be amplified due to the randomness nature of privacy preservation. A more rigid level of privacy preservation (i.e., smaller ϵ value) means the private view will have more randomness, and have more magnitude. Take the former Laplace approach as an example, in the non-private setting (e.g., $\epsilon = +\infty$), the magnitude of one preference data is $|v| = \sum_{j\in[1,d]}(d-j) = \frac{d(d-1)}{2} = 10$. Meanwhile the maximum possible magnitude of a private view $\tilde{v}$ becomes infinite, since Laplace random noises are unbounded. The expected magnitude of the private view grows with the privacy level as:

$$\mathbb{E}[|\tilde{v}|] = \frac{\Delta}{\epsilon} \cdot \frac{e^{\epsilon/\Delta} - e^{(1-d)\epsilon/\Delta}}{e^{\epsilon/\Delta} - 1} + \frac{d(d-1)}{2}.$$

A larger magnitude of an private view implies a larger impact on the aggregation result, hence LDP amplifies an atttacker's deconstructive power by contributing a adversary raw data.

View Disguise Attack. Another scenario is that an attacker has direct control on the private view sent to the aggregator. In this case, the adversary will be able to disguise a malicious private view as an ordinary (randomized) one, and thus make constructive/deconstructive changes to the aggregation result. The domain of private views is broader than the true data and grows with the level of privacy, hence an adversary's constructive/deconstructive power becomes larger. For example, in the non-private setting, the domain of a preference data is bounded by $[0, d-1]^d$, while in the Laplace approach, the domain of the private view is scaled to $[-\infty, +\infty]^d$. Even though the system can filter out private views which are extremely unlikely to be observed, the filtered domain $[-\tilde{\Theta}(\frac{1}{\epsilon}), +\tilde{\Theta}(\frac{1}{\epsilon})]$ still grows with the level of privacy preservation (see Section 3.3 for details). Consequently, it becomes easier for an attacker to manipulate the aggregation result from the (filtered) domain of private views, comparing to that in the non-private setting.

1.2 Our Contributions

With the adoption of LDP in real-world applications, these privacy-induced attacks may pose severe security threatens to consequent decision makings in data aggregation systems. As a remedy, we propose novel mechanisms to improve utility and robustness simultaneously. The main contributions of this paper are summarized as follows:

I. We identify *robustness issues* of LDP data aggregation systems, and categorize them into *data amplification attack* and *view disguise attack*, based on the adversary's controllability over the privacy preserving procedure. The data amplification attack captures an adversary's deconstructive power on the aggregation result by contributing fraudulent raw data. The view disguise attack captures an adversary's constructive/deconstructive power on the aggregation result by directly disguising private views. Formal quantified metrics are defined (in Section 2) to measure the power of adversarial attacks and the robustness of local private aggregation systems.

II. We thoroughly analyze Laplace mechanism for local private preference data aggregation (in Section 3), including sensitivity bounds with arbitrary scoring rules, error bounds of estimated scores (i.e., $\frac{d^5}{4n\epsilon^2}$), and maximum manipulation risk bounds (i.e., $+\infty$) under attacks.

III. We propose *the additive mechanism* (in Section 4), which samples a subset of candidates according to the summation of their scores. For preference data with Borda scores, the additive mechanism has estimation error bounds of $O(\frac{d^4}{n\epsilon^2})$, and expected/maximum manipulation risk bounds both at $O(\frac{d^2}{n\epsilon})$.

IV. We theoretically investigate the minimax lower bounds of ϵ-LDP preference data aggregation (in Section 5), provide lower bounds for arbitrary scoring rules. Specifically for the Borda scores, we prove that the additive mechanism is utility optimal.

TABLE 2
List of Notations

Notation	Description
$\mathbf{A}$	The set of candidates/items
d	The number of candidates/items
n	The number of users
$\mathbf{w}$	A scoring rule's score vector
$v^{(i)}$	The preference data of user i
D_v	The set of all possible permutations of $\mathbf{w}$
$\tilde{v}^{(i)}$	The view of the preference data $v^{(i)}$
$\mathsf{D}_{\tilde{v}}$	The set of all possible private views
θ	Average scores of candidates.
$\tilde{\theta}$	Estimator of average scores.
ϵ	The privacy budget

V. We discuss the interaction between utility, robustness, truthfulness and indistinguishability in the local private data aggregation (in Section 8). Quantified relations between estimation error bound and manipulation risk bound are built, which implies that optimizing utility generally benefits robustness.

VI. Experiments on extensive aggregation scenarios are conducted (in Section 6) to validate proposed mechanisms. Compared with existing approaches, estimation errors and manipulation risks are both significantly reduced.

2 Preliminaries

This section introduces definitions of preference data, local differential privacy, and the model of private data aggregation. The metrics of utility and robustness are also formally defined here. Notations throughout the paper are summarized in Table 2.

2.1 Preference/Ranking Data

A ranking π is a linear ordering over all candidates $\mathbf{A} = \{A_1, A_2, ..., A_d\}$. In a positional scoring rule, the j-th candidate π_j in a ranking data is assigned by a score of w_j. For reasonable positional scoring rules, the score vector $\mathbf{w} = \{w_1, w_2, ..., w_d\}$ is non-increasing. Included categorical/set-valued data as special cases, examples of score vector for popular scoring rules (with 5 candidates) are as follows:

- Borda: $\{4,\ 3,\ 2,\ 1,\ 0\}$;
- Nauru: $\{1/1,\ 1/2,\ 1/3,\ 1/4,\ 1/5\}$;
- Plurality: $\{1,\ 0,\ 0,\ 0,\ 0\}$;
- Anti-plurality: $\{1,\ 1,\ 1,\ 1,\ 0\}$;
- k-Approval (k = 2): $\{1,\ 1,\ 0,\ 0,\ 0\}$.

For the simplicity of reference, we rewrite the user i's ranking data $\pi^{(i)}$ as numerical preferential scores (namely preference data) on each candidate: $v^{(i)} = [v_1^{(i)}, v_2^{(i)}, ..., v_d^{(i)}]$, where $v_j^{(i)}$ is the score of candidate A_j.

2.2 Local Differential Privacy

LDP ensures bounded distinguishability in outputs for any two possible inputs, hence blocks adversaries from inferring much information from outputs. Let D_π denotes the domain of ranking data, which represents all possible orderings over candidates $\mathbf{A}$, let M denotes a randomized mechanism, and D_M denotes the output domain of the mechanism, LDP can be formally stated in Definition 1.

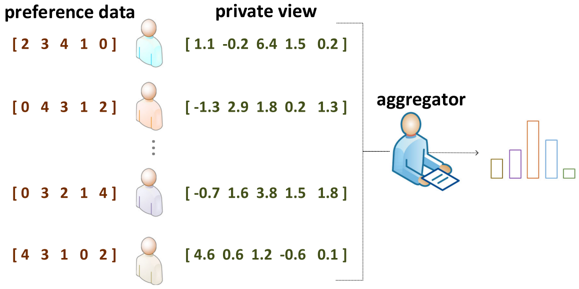


Fig. 1. Demonstration of preference data aggregation with local privacy.

Definition 1 (ϵ-LDP). *A randomized mechanism* M *satisfies local* ϵ*-differential privacy iff for any possible ranking pair* $\pi, \pi' \in \mathsf{D}_\pi$, *and any possible output* $t \in \mathsf{D}_\mathsf{M}$,

$$\mathbb{P}[\mathsf{M}(\pi) = t] \le e^\epsilon \cdot \mathbb{P}[\mathsf{M}(\pi') = t]$$

holds.

Here the parameter ϵ is called the privacy budget, which controls the level of privacy preservation. Practical values for ϵ range between $[0.01, 3.0]$.

2.3 Aggregation Model

Consider n users $N = \{1, 2, ..., n\}$, each user i holds a ranking data $\pi^{(i)}$ (or a preference data $v^{(i)}$). For the purpose of privacy preservation, the user i sanitizes $\pi^{(i)}$ to get the private view $\tilde{v}^{(i)}$ by running an ϵ-LDP mechanism M locally and independently. The private view $\tilde{v}^{(i)}$ from a meaningful mechanism is an estimator of the true preference data $v^{(i)}$, hence the aggregator could estimate the actual average scores $\theta = \frac{1}{n}\sum v^{(i)}$ by:

$$\tilde{\theta} = \frac{1}{n}\sum \tilde{v}^{(i)}. \tag{1}$$

Fig. 1 demonstrates the procedures of local private preference data aggregation. In adversarial environments, the aggregator may filter out some potential malicious private views.

2.4 Performance Metrics

Utility Metrics. Estimators of average scores $\tilde{\theta}$ given by different mechanisms have varied accuracy, here we use two typical utility metrics:

- Mean squared error: $\mathrm{err}_{\mathrm{MSE}} = \mathbb{E}[|\tilde{\theta} - \theta|_2^2]$;
- Total variation error: $\mathrm{err}_{\mathrm{TVE}} = \mathbb{E}[|\tilde{\theta} - \theta|_1]$;

Since preference data aggregation tasks usually aim to determine a ranking over candidates, we also measure the utility with the ranking correlation coefficient Kendall's tau [16] (denoted as $\mathrm{score}_{\mathrm{KT}}$), *Accuracy of Winner* ($\mathrm{accuracy}_{\mathrm{AOW}}$) and *Loss of Winner* ($\mathrm{err}_{\mathrm{LOW}}$). Let $j_{max} = \arg\max_{j\in[1,d]}\theta_j$ denote the candidate's index with maximum average score in true average scores θ, and $\tilde{j}_{max} = \arg\max_{j\in[1,d]}\tilde{\theta}_j$ denote the winning candidate's index in the estimated average scores $\tilde{\theta}$, we adopt following two utility metrics:

- Accuracy of winner: $\mathrm{accuracy}_{\mathrm{AOW}} = \mathbb{E}[\tilde{j}_{max} = j_{max}]$;
- Loss of winner: $\mathrm{err}_{\mathrm{LOW}} = \mathbb{E}[\theta_{j_{max}} - \theta_{\tilde{j}_{max}}]$;

Robustness Metrics. To measure an attacker's manipulation power on the aggregation result, we define two robustness metrics over the private view $\tilde{v}$ as:

- Expected magnitude: $\text{risk}_{\text{EM}} = \mathbb{E}[\frac{|\tilde{v}|_1}{n}]$.
- Maximum magnitude: $\text{risk}_{\text{MM}} = \max_{\tilde{v} \in \mathsf{D}_{\tilde{v}}} \frac{|\tilde{v}|_1}{n}$;

These two metrics measure the expected/maximum absolute difference that one private view could contribute. Specifically, the risk_{EM} measures the expected manipulation power by contributing one extra adversarial data (i.e., the data amplification attack), and the risk_{MM} measures the largest manipulation power by contributing one extra adversarial view (i.e., the view disguise attack).

3 Laplace Mechanism

We start with the classical Laplace mechanism for local private preference data aggregation.

3.1 Design

For numerical values like preference data, the Laplace mechanism is the most popular approach to achieve (local) differential privacy. The scale of the Laplace random noises is calibrated to the sensitivity Δ of the preference data [11] as in Algorithm 1. Lemma 1 gives exact bound of the sensitivity Δ for all positional scoring rules, shows that the sensitivity is the total difference between descent and ascent score vectors: $\sum_{j \in [1,d]} |w_j - w_{d-j+1}|$.

Lemma 1. *For any position rules with non-increasing score vector* $\mathbf{w}$*, the sensitivity of preference data is*

$$\Delta = \max_{v, v' \in \mathsf{D}_v} |v - v'|_1 = \sum_{j \in [1,d]} |w_j - w_{d-j+1}|.$$

Algorithm 1. Laplace Mechanism

```
Input: A preference data $v \in \mathsf{D}_v$, privacy budget $\epsilon$ and the score vector $\mathbf{w}$ of the scoring rule.
Output: An unbiased private view $\tilde{v} \in \mathbb{R}^m$ that satisfies $\epsilon$-LDP.
▷ Compute sensitivity
$\Delta \leftarrow \sum_{j \in [1,d]} |w_j - w_{d-j+1}|$
▷ Randomization by adding Laplace noises
for $j \leftarrow 1$ to $d$ do
    $\tilde{v}_j \leftarrow v_j + Lap(\frac{\Delta}{\epsilon})$
end for
return $\tilde{v} = \{\tilde{v}_1, \tilde{v}_2, ..., \tilde{v}_d\}$
```

3.2 Utility Analysis

The utility bound of the average score estimator given by Laplace mechanism is analyzed in Theorem 1, proof of which is a simple application of Laplace random variables' variance formulation. The estimation error bound is square to the sensitivity $\sum_{j \in [1,d]} |w_j - w_{d-j+1}|$. For preference data with Borda scoring rule, we have the sensitivity at $\Theta(d^2)$, hence $err_{MSE} = \Theta(\frac{d^5}{n\epsilon^2})$. The power factor of 5 on the number of candidates d implies that the Laplace mechanism introduces much noise, hence we seek optimized approaches in the following sections.

Theorem 1. *The mean squared error* $\mathbb{E}[|\tilde{\theta} - \theta|_2^2]$ *of the Laplace mechanism is:*

$$\text{err}_{\text{MSE}} = \frac{2d \cdot \left(\sum_{j \in [1,d]} |w_j - w_{d-j+1}|\right)^2}{n\epsilon^2}.$$

3.3 Robustness Analysis

We show risks under data amplification attack and view disguise attack in Theorem 2.

Theorem 2. *The risks of Laplace mechanism under adversarial attacks are:*

$$\text{risk}_{\text{MM}} = +\infty; \ \text{risk}_{\text{EM}} = \sum_{j \in [1,d]} \frac{\Delta}{\epsilon} \exp\left(\frac{-|w_j|\epsilon}{\Delta}\right) + |w_j|.$$

It can be seen that the maximum possible risk of the Laplace mechanism is infinite and the expected risk grows linearly with $\frac{1}{\epsilon}$. Therefore, imposing a stringent level of privacy harms the robustness of the aggregation result. One possible solution to restrict the unlimited maximum possible risk is filtering out private views that are extremely unlikely observed. For example, we may define an allowable output area (with threshold probability β) as:

$$\tilde{\mathsf{D}}_p = \{\tilde{v} \mid \tilde{v} \in \mathbb{R}^d, \ \Pr[\tilde{v}|v] \geq \beta \text{ for some } v \in \mathsf{D}_v\}.$$

For Laplace mechanism, it is equivalent to:

$$\{\tilde{v} \mid \tilde{v} \in \mathbb{R}^d, \ |\tilde{v} - v| \leq \frac{\Delta(\log(1/\beta) + d\log(\Delta/\epsilon))}{\epsilon} \text{ for some } v \in \mathsf{D}_v\}.$$

Thus even if we can filter out outliers of private views, the volume of the allowable output area, which determines maximum possible risks, still grows with $\frac{1}{\epsilon}$.

4 Additive Mechanism

The utility/robustness performance of the Laplace mechanism largely depends on the sensitivity $\Delta_{\mathbf{w}} = \sum_{j \in [1,d]} |w_j - w_{d-j+1}|$. Without much modification, we can also adapt ϵ-LDP multi-dimensional categorical/numerical mechanisms (e.g., [17], [18], [19], [20]) for the preference data. They cast a multi-dimensional problem to an easier one-dimensional one via data-dependent sampling (i.e., sample each candidate j with a probability proportional to $|v_j|$ in [17], [18]) or data-independent sampling (i.e., uniform-randomly select one dimension in [19], [20]), then employ well-studied one-dimensional mechanisms (e.g., unary encoding, local hash[21], subset [22] or piecewise mechanism [19]). Rooted in their sample-then-randomize nature, given that the err_{MSE} lower bound for one-dimensional estimation over [0, 1] is $O(\frac{1}{n\epsilon^2})$ [23], the data-dependent sampling-based mechanisms incur $O(\frac{d \cdot (\sum_{j \in [1,d]} |w_j|)^2}{n\epsilon^2})$ errors; the data-independent sampling-based mechanisms incur $O(\frac{d^2 |w_1 - w_d|^2}{n\epsilon^2})$ errors. As for robustness, the $risk_{EM}$ of state-of-the-art mechanisms (i.e., local hash [21], subset [22] or piecewise mechanism [19]) for one-dimensional estimation over [0,1] is $O(\frac{1}{\epsilon})$, therefore the data-dependent sampling-based mechanisms introduce $O(\frac{d \sum_{j \in [1,d]} |w_j|}{\epsilon})$ risk; the data-independent sampling mechanisms introduce $O(\frac{d|w_1-w_d|}{\epsilon})$ risks. The data-independent sampling mechanisms could improve upon the Laplace mechanism or data-dependent sampling mechanisms, but they simply treat v^i as $[w_d, w_1]^d$ and ignore the permutation structure of the preference data, which has identical ℓ_p-norm for every $v^i \in \mathsf{D}_v$. As opposed to the above sampling-then-randomize paradigm, here we propose an end-to-end approach that utilizes the fixed ℓ_p-norm property: the additive mechanism.

4.1 Design

Let $\mathbf{A}^k = \{S \mid S \subseteq \mathbf{A} \text{ and } |S| = k\}$ denote the set of candidate subsets of size of k, let $\mathbf{w}^k_{max} = \sum_{j\in[1,k]} w_j$ denote the maximum total weights of one subset, let $\mathbf{w}^k_{min} = \sum_{j\in[[d-k+1,d]]} w_j$ denote the minimum total weights of one subset, the detail of the additive mechanism is presented in Definition 2.

As its name, the additive mechanism randomly responses with a subset of candidates S with a probability linear to their total scores $\sum_{A_{j'}\in S} v_{j'}$. The mechanism is a novel mutant of the popular exponential mechanism for achieving differential privacy. The exponential mechanism responses with a probability proportional to the **exponential** of candidates' scores, while the additive mechanism responses with a probability proportional to the **additive** summation of candidates' scores. Consequently, we can derive an unbiased estimation of average scores. Take the preference data with Borda rule $v = \{3, 4, 0, 2, 1\}$ as an example, when $k = 1$, the additive mechanism outputs $\{A_1\}, \{A_2\}, \{A_3\}, \{A_4\}, \{A_5\}$ with probability $\frac{3(e^\epsilon-1)+1}{\Phi}, \frac{4(e^\epsilon-1)+1}{\Phi}, \frac{1}{\Phi}, \frac{2(e^\epsilon-1)+1}{\Phi}, \frac{e^\epsilon}{\Phi}$ respectively.

Definition 2 (Additive Mechanism). *In an ϵ-LDP aggregation system, where the candidates are $\mathbf{A}$ and the scored vector is $\mathbf{w}$, take preference data v as input, the additive mechanism randomly outputs an $S \in \mathbf{A}^k$ according to following probability design:*

$$\Pr[S|v] = \frac{\sum_{A_{j'}\in S} v_{j'} - \mathbf{w}^k_{min}}{\mathbf{w}^k_{max} - \mathbf{w}^k_{min}} \cdot \frac{e^\epsilon - 1}{\Phi} + \frac{1}{\Phi},$$

where the normalizer factor is $\Phi = \binom{d}{k} \frac{\frac{k}{n}(e^\epsilon-1)\sum_{j\in[1,d]} w_j - e^\epsilon \mathbf{w}^k_{min} + \mathbf{w}^k_{max}}{\mathbf{w}^k_{max} - \mathbf{w}^k_{min}}$. The estimator of the preference data v is ($j \in [1, d]$):

$$\tilde{v}_j = a_k \cdot [C_j \in S] - b_k,$$

where $a_k = [\sum_{j'\in[1,d]} w_{j'}(e^\epsilon - 1) - \frac{d}{k} e^\epsilon \mathbf{w}^k_{min} + \frac{d}{k}\mathbf{w}^k_{max}] \frac{d-1}{(d-k)(e^\epsilon-1)}$, *and* $b_k = [\frac{(k-1)(e^\epsilon-1)}{d-1} \sum_{j'\in[1,d]} w_{j'} - e^\epsilon \mathbf{w}^k_{min} + \mathbf{w}^k_{max}] \frac{d-1}{(d-k)(e^\epsilon-1)}$

Theorem 3. *The additive mechanism satisfies ϵ-LDP.*

Proof. Since $\tilde{v}$ is mapped from S, to prove the private view $\tilde{v}$ satisfies ϵ-LDP, it's enough to show that the intermediate view S satisfies ϵ-LDP.

First we need to prove $\Pr[S|v]$ is a valid probability distribution, that is $\Pr[S|v] \geq 0.0$ and $\sum_{S\in\mathbf{A}^k} \Pr[S|v] = 1.0$ hold for any input $v \in \mathsf{D}_v$. Since $\sum_{A_{j'}\in S} v_{j'} \geq \mathbf{w}^k_{min}$, we have $\Phi > 0$ and hence $\Pr[S|v] \geq 0.0$. Now consider $\sum_{S\in\mathbf{A}^k} \Pr[S|v]$, we have:

$$\begin{aligned}
&\frac{\binom{d}{k}}{\Phi} + \sum_{S\in\mathcal{C}^k} \sum_{C_{j'}\in S} \frac{v_{j'} - \mathbf{w}^k_{min}}{\mathbf{w}^k_{max} - \mathbf{w}^k_{min}} \cdot \frac{e^\epsilon - 1}{\Phi} \\
&= \frac{\binom{d}{k}}{\Phi} + \binom{d-1}{k-1} \sum_{C_{j'}\in\mathcal{C}} \frac{v_{j'} - \mathbf{w}^k_{min}}{\mathbf{w}^k_{max} - \mathbf{w}^k_{min}} \cdot \frac{e^\epsilon - 1}{\Phi} \\
&= \frac{\binom{d}{k}}{\Phi} + \binom{d-1}{k-1} \frac{\sum_{j\in[1,d]} w_j - \mathbf{w}^k_{min}}{\mathbf{w}^k_{max} - \mathbf{w}^k_{min}} \cdot \frac{e^\epsilon - 1}{\Phi} \\
&= \frac{1}{\Phi}\binom{d}{k} \cdot \left(\frac{\frac{k}{d}\sum_{j\in[1,d]} \mathbf{w}_j - \mathbf{w}^k_{min}}{\mathbf{w}^k_{max} - \mathbf{w}^k_{min}} \cdot (e^\epsilon - 1) + 1 \right) \; = \; 1.
\end{aligned}$$

Second for any paired inputs $v, v' \in \mathsf{D}_v$ and any output values $S \in \mathbf{A}^k$, we have:

$$\begin{aligned}
\frac{\Pr[S|v]}{\Pr[S|v']} &\leq \frac{\max_{S\in\mathbf{A}^k} \Pr[S|v]}{\min_{S\in\mathbf{A}^k} \Pr[S|v']} \\
&\leq \frac{\frac{\mathbf{w}^k_{max} - \mathbf{w}^k_{min}}{\mathbf{w}^k_{max} - \mathbf{w}^k_{min}} \cdot \frac{e^\epsilon-1}{\Phi} + \frac{1}{\Phi}}{\frac{\mathbf{w}^k_{min} - \mathbf{w}^k_{min}}{\mathbf{w}^k_{max} - \mathbf{w}^k_{min}} \cdot \frac{e^\epsilon-1}{\Phi} + \frac{1}{\Phi}} \\
&\leq e^\epsilon.
\end{aligned}$$

□

Lemma 2. *The private view $\tilde{v}$ given by the additive mechanism is an unbiased estimation of the preference data v.*

After giving formal ϵ-LDP guarantee and unbiasedness guarantee of the additive mechanism in Theorem 3 and Lemma 2 respectively, we turn to consider an efficient implementation of the additive mechanism. By decomposing subsets in $\mathbf{A}^k$ into $d - k + 1$ groups according to the top-most rank of candidates in a subset, then randomly choose one group and transform to a sub-problem of selecting $k - 1$ weighted options as in Algorithms 2 and 3. The computational complexity of the above recursive procedure is $O(d \cdot k)$ and hence is efficient for large scale deployments.

Algorithm 2. Additive Mechanism

```
Input: A ranking data $\pi$, privacy budget $\epsilon$, scoring vector $\mathbf{w}$ and parameter $k$.
Output: An unbiased private view $\tilde{v} \in \mathbb{R}^d$ that satisfies $\epsilon$-LDP.
▷ Select $k$ ranking positions
for $j \in [1, d]$ do
    ▷ Compute weights of presence for a ranking position
    $z_j \leftarrow \frac{w_j - \mathbf{w}^k_{min}/k}{\mathbf{w}^k_{max} - \mathbf{w}^k_{min}} \cdot (e^\epsilon - 1) + \frac{1}{k}$
end for
$T \leftarrow additive\_select(d, k, \mathbf{z})$
▷ Deriving unbiased estimator
$S \leftarrow \{\pi_j \mid j \in T\}$
for $j \in [1, d]$ do
    $\tilde{v}_j \leftarrow [C_j \in S] \cdot a_k - b_k$
end for
return $\tilde{v} = \{\tilde{v}_1, \tilde{v}_2, ..., \tilde{v}_d\}$
```

4.2 Utility Analysis

The estimation error bound of the additive mechanism is given in Theorem 4. The formulation of the bound has a dependence on the score vector of a scoring rule. Under the plurality rule (a.k.a the categorical preference data), the

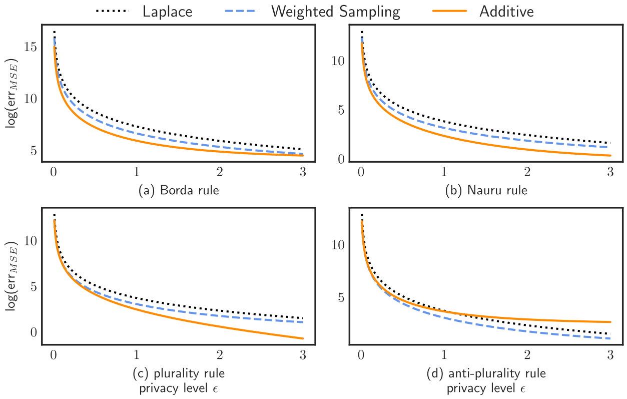


Fig. 2. Theoretical mean squared estimation error of Laplace, weighted sampling and additive mechanism with Borda, Nauru, plurality and anti-plurality scoring rules over 5 candidates.

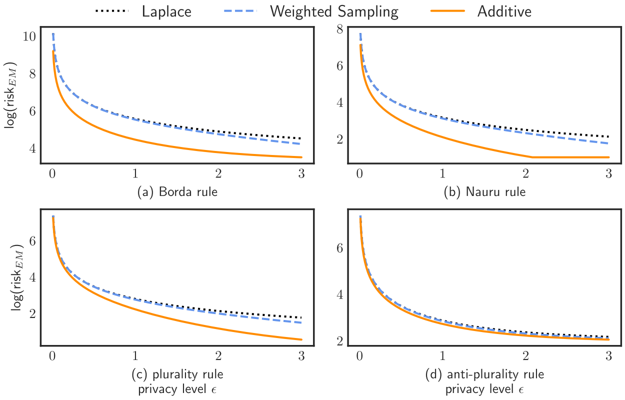


Fig. 3. Theoretical expected magnitude risks of Laplace, weighted sampling and additive mechanism with Borda, Nauru, plurality and anti-plurality rules over 5 candidates.

additive mechanism degrades to the optimal ϵ-LDP mechanism for categorical data aggregation: the subset mechanism [21]. Under Borda rule, we have optimal mechanism $\sum_{j'\in[1,d]} \hat{w}_{j'} = O(d^2)$ when $\epsilon = O(1)$, hence the mean squared error bound is $O(\frac{d^4}{\epsilon^2})$. As a comparison, the mean squared error bounds of the Laplace mechanism and weighted sampling mechanism are $O(\frac{d^5}{\epsilon^2})$.

For better illustration, we depict numerical bound of the additive mechanism for Borda rule in Fig. 2, along with a comparison with the Laplace and weighed sampling mechanism. The numerical results show average 75% error reduction compared to the Laplace mechanism, and average 40% error reduction compared to the weighted sampling mechanism.

Theorem 4. *The mean squared error* $\mathbb{E}[|\tilde{\theta} - \theta|_2^2]$ *of the additive mechanism is bounded as follows:*

$$\frac{\left(\sum_{j'\in[1,d]} \hat{w}_{j'}\right)^2 - \sum_{j'\in[1,d]} \hat{w}_{j'}^2}{n(e^\epsilon - 1)^2},$$

where $\hat{w}_j = w_j(e^\epsilon - 1) - e^\epsilon w_d + w_1$.

Proof. The parameter $k = 1$ is near to optimal for many voting rules except extremal cases of score vector **w** (e.g., plurality voting). Hence to characterize the usefulness performance of additive mechanism, we only need to analyze the case when $k = 1$. Given that $a_1 = \sum_{j'\in[1,d]} w_{j'} - \frac{e^\epsilon d}{e^\epsilon - 1} w_d + \frac{d}{e^\epsilon - 1} w_1$, $b_1 = -\frac{e^\epsilon}{e^\epsilon - 1} w_d + \frac{1}{e^\epsilon - 1} w_1$, and $\Pr[A_j \in S|v] = \frac{v_j(e^\epsilon - 1) - e^\epsilon w_d + w_1}{\sum_{j'\in[1,d]} w_{j'}(e^\epsilon - 1) - e^\epsilon d w_d + d w_1}$. The variance of Bernoulli variable $[A_j \in S]$ is $\Pr[A_j \in S|v](1 - \Pr[A_j \in S|v])$, then the variance of $\tilde{v}_j = a_1[A_j \in S] - b_1$ is:

$$(a_1)^2 \Pr[A_j \in S|v](1 - \Pr[A_j \in S|v]).$$

Consequently, the total variance $\mathrm{E}[|\tilde{v} - v|_2^2]$ is $\frac{(\sum_{j'\in[1,d]} \hat{w}_{j'})^2 - \sum_{j'\in[1,d]} \hat{w}_{j'}^2}{(e^\epsilon - 1)^2}$. □

4.3 Robustness Analysis

The adversarial risks of additive mechanism are presented in Theorem 5. When applying parameter $k = 1$ for the preference data with Borda rule, we have $a_1 = O(\frac{d}{\epsilon})$ and $b_1 = O(\frac{d}{\epsilon})$, hence the risk bounds of $\mathrm{risk}_{\mathrm{MM}}$ and $\mathrm{risk}_{\mathrm{EM}}$ are both $O(\frac{d^2}{\epsilon})$, and the risk bound of $\mathrm{risk}_{\mathrm{DD}}$ is $O(\frac{d}{\epsilon})$. As a comparison, in the data-dependent sampling mechanisms, the risk bounds of $\mathrm{risk}_{\mathrm{MM}}$ and $\mathrm{risk}_{\mathrm{EM}}$ are $O(\frac{d^3}{\epsilon})$. Fig. 3 presents the numerical results of these risks in the additive mechanism with comparison to the Laplace and data-dependent weighted sampling mechanisms. In most cases, the additive mechanism reduces 70% expected magnitude.

Algorithm 3. Additive_select$(d, k, \mathbf{w})$

Input: The number of positions d, parameter k, and positions' weights **z**.

Output: k ranking positions $T \subseteq [1, d]$.

```
▷ Compute probabilities of $p_j = \Pr[min(T) = j]$
for $j \in [1, d - k + 1]$ do
    $p_j \leftarrow \binom{d-j}{k-1} \cdot (z_j + (\sum_{j'\in[j+1,d]} z_{j'} - z_j)\frac{k-1}{d-j})$
end for
▷ Select a minimum rank $j^*$
$j^* \leftarrow 0$
while $r \geq 0.0$ do
    $j^* \leftarrow j^* + 1$
    $r \leftarrow r - \frac{p_{j^*}}{\sum_{j\in[1,d]} p_{j^*}}$
end while
for $j \in [j^* + 1, d]$ do
    $z'_{j-j^*} = w_j + \frac{z_{j^*}}{k-1}$
end for
▷ Recursively select $k - 1$ ranking positions
$T' = additive\_select(d - j^*, k - 1, \mathbf{z}')$
return $T = \{j^*\} \cup \{j + j^* \mid j \in T'\}$
```

Theorem 5. *The manipulation risks of the additive mechanism are:*

$$\mathrm{risk}_{\mathrm{MM}} = \frac{k|a_k - b_k| + (d-k)|b_k|}{n}; \mathrm{risk}_{\mathrm{EM}} = \frac{k|a_k - b_k| + (d-k)|b_k|}{n}.$$

Proof. For any intermediate result of subset $S \in \mathbf{A}^k$, the corresponding private view $\tilde{v} = \{\tilde{v}_1, \tilde{v}_2, ..., \tilde{v}_d\}$ contains a number k of $a_k - b_k$ and a number $d - k$ of $-b_k$, hence both $\mathrm{risk}_{\mathrm{MM}}$ and $\mathrm{risk}_{\mathrm{EM}}$ are $\frac{|a_k - b_k| + (d-1)|b_k|}{n}$. □

5 Tight Error Bounds

This section theoretically gives lower error bounds of ϵ-LDP preference data aggregation, and further shows our proposed additive mechanism is optimal (under certain scoring rules). The lower bound is derived by applying the recently established local private version of Assouad's method [23] and utilizing the permutation structure of preference data domain.

5.1 Local Private Minimax Risks

Let $\mathcal{M}_\epsilon$ denote the set of all possible mechanisms $\mathsf{M} = \{M^1, ..., M^n\}$ that satisfy ϵ-LDP for all n users. Taking as input samples $\{x^1, x^2, ..., x^n\}$, some mechanisms $\mathsf{M} \in \mathcal{M}_\epsilon$ output a list of sanitized private views $\{z^1, z^2, ..., z^n\}$. If the estimator

$$\hat{\theta} = \hat{\theta}(\{x^1, x^2, ..., x^n\})$$

is computed on these private views while having no access to input samples $\{x^j\}_{j=1}^n$, the minimax squared error of ϵ-LDP mechanisms can be defined as:

$$\begin{aligned}&\mathfrak{M}_n(\theta(\mathcal{P}), \|\cdot\|_2^2, \epsilon)\\ &:= \inf_{\mathsf{M}\in\mathcal{M}_\epsilon} \inf_{\hat{\theta}} \sup_{P\in\mathcal{P}} \mathbb{E}_{p,\mathsf{M}}[\|\hat{\theta}(z^1, z^2, ..., z^n) - \theta(P)\|_2^2],\end{aligned}$$

which is the minimum possible worst-case (w.r.t. the $\mathcal{P}$) error for any ϵ-LDP mechanisms (in the $\mathcal{M}_\epsilon$).

The Assouad's method [24] is a useful lower bounding tool with multiple hypothesis testings. First, it defines a hypercube $\mathcal{V} = \{-1, 1\}^d$ ($d \in \mathbb{N}$), and then definesa family of distributions $\{P_\nu\}_{\nu\in\mathcal{V}}$ indexed by the hypercube (each P_ν is defined on one common space). We say that the distribution family induces a 2δ-Hamming separation for the squared loss $\|\cdot\|_2^2$ if there exists a vertex mapping (a function $\kappa : \theta(\mathcal{P}) \mapsto \{-1, 1\}^d$) satisfying:

$$\|\theta, \theta(P_\nu)\|_2^2 \geq 2\delta \sum_{j=1}^{d} \mathbf{1}\{[\kappa(\theta)]_j \neq \nu_j\}.$$

The nature first uniform-randomly chooses a vector $V \in \{-1, 1\}^d$, then we assume the samples $\{\mathbf{x}^1, ..., \mathbf{x}^n\}$ are drawn from the distribution P_ν given $V = \nu$. These samples are taken as input into ϵ-LDP mechanisms $\mathbf{K}$. The work [23] gives a private version of Assouad's method for ϵ-LDP estimation problems as follows.

Lemma 3 (Private Assouad bound [23]). *Let* $P_{+j} = \frac{1}{2^{d-1}} \sum_{\nu:\nu_j=1} P_\nu$ *and* $P_{-j} = \frac{1}{2^{d-1}} \sum_{\nu:\nu_j=-1} P_\nu$*, we have*

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

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

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

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

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

5.2 Lower Bounds for Preference Data

In this part, we proceed to prove that any ϵ-LDP mechanism M must has worst-case mean squared error $O(\frac{d\sum_{j=1}^d (w_j - w_{d+1-j})^2}{n\epsilon^2})$ on estimating the average scores.

Theorem 6. *For the preference data aggregation problem with scoring rule* $\mathbf{w} \in \mathcal{R}^d$*, for any* ϵ*-LDP mechanism, there exists a universal constant* $c > 0$ *such that for all* $\epsilon \in [0, 1]$,

$$\mathfrak{M}_n(\theta(\mathcal{P}), \|\cdot\|_2^2, \epsilon) \geq c \cdot \min\left\{\frac{1}{d}, \frac{d\sum_{j=1}^d (w_j - w_{d+1-j})^2}{n\epsilon^2}\right\}$$

Proof. Follow conventional lower bounding procedures, the sketch of our proof is as follows:

1. The first step is a standard reduction from estimation to a multiple binary hypothesis testing problem. Here for preference data, each binary hypothesis test is to decide whether rank j or $d + 1 - j$ is prevalent on one candidate.
2. After constructing appropriately separated binary hypothesis tests, we then control the testing accuracy in the binary testing problem, and finally applies Lemma 3 for bounding estimation error.

Our first step is to define a large set of distributions, and these distributions are well hamming-separated for the squared error. Assuming that the number of candidates d is even, we denote a hypercube as $\mathcal{B} = \{-1, 1\}^{d/2}$. For a preference data v, here we consider special cases where $v_{j'}$ is either w_j or w_{d+1-j} for some $j \in [1, d/2]$, and the corresponding $v_{d+1-j'}$ is either w_j or w_{d+1-j} for the same j. As a valid preference data v, when j' is enumerated from 1 to $d/2$, the j is also enumerated from 1 to $d/2$, here we denote the corresponding bijection function from j to j' as $q : \{1, d/2\} \longrightarrow \{1, d/2\}$. Fix a parameter $\delta \in [0, 1]$, for any $b \in \mathcal{B}$, we can then define a probability distribution P_b of the preference data as follows (for $j \in [1, d/2]$):

$$v_{q(j)} = \begin{cases} w_j, & \text{with probability } \frac{1+b_j}{2}; \\ w_{d+1-j}, & \text{with probability } \frac{1-b_j}{2}. \end{cases}$$

$$v_{d+1-q(j)} = \begin{cases} \mathbf{w}_j, & \text{with probability } \frac{1-b_j}{2}; \\ \mathbf{w}_{d+1-j}, & \text{with probability } \frac{1+b_j}{2}. \end{cases}$$

Consequently, the expected average score under distribution p_b is then:

$$\theta_{q(j)} = \frac{w_j + w_{d+1-j}}{2} + b_j \frac{w_j - w_{d+1-j}}{2},$$

$$\theta_{d+1-q(j)} = \frac{w_j + w_{d+1-j}}{2} - b_j \frac{w_j - w_{d+1-j}}{2}.$$

For any average score estimator $\hat{\theta}$ that is agnostic of bijection function q, by defining $\hat{b}_j = sign(\hat{\theta}_{q(j)} - \frac{w_j + w_{d+1-j}}{2})$, we have lower bound on separation:

$$||\hat{\theta} - \theta_{q(j)}||_2^2 \geq \frac{\delta^2 \sum_{j'=1}^{d} (w_j - w_{d+1-j})^2}{4d} \sum_{j=1}^{d} [[\hat{b}_j \neq b_j]].$$

We now proceed to derive the maximum possible discrepancy $F_{\mathbb{B}_\infty(\mathcal{X}^s),\mathcal{P}}$ between induced marginal distributions P_{+j} and P_{-j}, hence control the accuracy of binary hypothesis testing about b_j.

By construction, P_{+j} is a joint distribution of $d/2$ independent binomial distributions:

$$\Pi_{a=1}^{d/2} \left[\frac{1}{4}\mathbf{1} + \frac{\delta}{4} \left[e_{j \bmod 2}^\top - e_{j \bmod 2}^\top \right]^\top \{\lfloor j/2 \rfloor = a\} \right] \in \boldsymbol{\Delta}_4,$$

and similarly for P_{-j}, where $e_j \in \{0,1\}^2$ denote the j-th standard basis vector.

Due to the interleaving structure of the $d/2$-dimensional distribution P_{+j} and P_{+j}, for any $\gamma \in [-1,1]^d$, we have:

$$\sum_{a=1}^{d/2} \sum_{j=1}^{2} \left(\int_{\mathcal{X}^m} \gamma(x)(dP_{+j}(x) - dP_{-j}(x)) \right)^2 \leq 4\delta^2,$$

due to fact that assigning $\gamma \in [-1,1]^{2d}$ according to one of the dimension maximizes the overall integral discrepancy.

Combining the previous results on hamming separation and marginal distribution's discrepancy, according to local private version of Assouad bound in Lemma 3, we have:

$$\max_{b \in \mathcal{B}} \mathbb{E}_{P_b}[||\hat{\theta} - \theta_b||_2^2] \geq \frac{\delta^2 \sum_{j=1}^{d} (w_j - w_{d+1-j})^2}{16} [1 - (2n(e^\epsilon - 1)^2 \delta^2)^{\frac{1}{2}}].$$

Further choose the parameter δ^2 at $\min\{1, d/(8n(e^\epsilon - 1)^2)\}$ or $\min\{2/d, 1/(8n(e^\epsilon - 1)^2)\}$, we have the lower bound as:

$$\mathfrak{M}_n(\theta(\mathcal{P}), ||\cdot||_2^2, \epsilon) \geq \min\left\{ \frac{1}{4d}, \frac{d \sum_{j=1}^{d} (w_j - w_{d+1-j})^2}{128n(e^\epsilon - 1)^2} \right\}.$$

□

The estimation lower bound is highly binded to the design of scoring rule $\mathbf{w}$, and the the error term $\sum_{j=1}^{d} (w_j - w_{d+1-j})^2$ is the squared diameter of the preference data domain.

We here compare the lower bound with the upper error bound of the proposed Additive mechanism, and conclude that the Additive mechanism is minimax error optimal for the Borda rule. Recall that the mean squared error upper bound of the additive mechanism scales with $(\sum_{j' \in [1,d]} \hat{w}_{j'})^2 - \sum_{j' \in [1,d]} \hat{w}_{j'}^2$ $(\hat{w}_j = w_j(e^\epsilon - 1) - e^\epsilon w_d + w_1.)$, while the minimax lower bound scales with $d \sum_{j=1}^{d} (w_j - w_{d+1-j})^2$. For the Borda rule, we have $d \sum_{j=1}^{d} (w_j - w_{d+1-j})^2 = \Theta(d^4)$, hence the additive mechanism matches the minimum possible worst-case error bound (i.e., the minimax error bound) in Theorem 6 and thus achieves an optimal rate of convergence.

TABLE 3
Enumeration of Experiment Settings, the Values in the Bold Format are the Default Settings

Parameter	Enumerated values
scoring rule	**Borda**, Nauru
number of candidates d	$4, \mathbf{8}, 16, 32$
normal users n	$1000, \mathbf{10000}, 1000000$
adversarial data n'	$n \cdot 0.1\%, \quad n \cdot 1\%, \quad n \cdot 5\%$
adversarial views n''	$n \cdot 0.1\%, \quad n \cdot 1\%, \quad n \cdot 5\%$
privacy budget ϵ	$0.01, 0.1, 0.2, 0.4, 0.8, 1.0, 1.5, 2.0, 3.0$

6 Experiments

We now evaluate the utility and robustness performance of the proposed additive mechanism with optimal parameter k^*, and compare it with the *Laplace* mechanism [11], the data-dependent sampling-based approach [17], [25] with naive magnitude $|v_j|$, and optimized *Weighted Sampling* approach that sampling with relative magnitude $|v_j - \min(\mathbf{w})|$. The sampled candidate is then sanitized with an optimal categorical ϵ-LDP mechanism: the subset mechanism [22], which dominates the RAPPOR [7], binary randomized response [23], unary encoding and local hash [21] in all settings (w.r.t. choices of d or ϵ). We also compare it with the state-of-the-art data-independent sampling-based approach for multi-dimensional numerical data $[w_d, w_1]^d$: the piecewise mechanism [19].

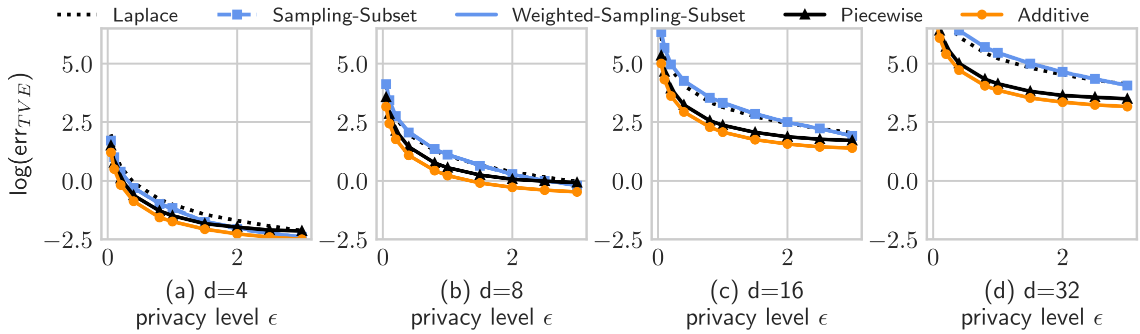
6.1 Settings

Datasets. In order to thoroughly assess the performance of mechanisms in extensive settings, we use synthetic datasets with diverse parameters. In each simulation, each candidate A_j is assigned with a uniform random scale $\alpha_j \in [0.0, 1.0)$. Each user's numerical preference $\beta_{(i,j)}$ on candidate A_j is an independent uniform random value $r_{(i,j)} \in [0.0, 1.0)$ multiplied by the scale $\alpha_j \in [0.0, 1.0)$, then the user's ranking on candidates is determined by $\beta_{(i,j)}$. In these simulations, the number of candidates d ranges from 4 to 32, the number of users n ranges from 1000 to 1 000 000.

Adversarial data in the simulation of data amplification attack are uniform-randomly selected from the domain D_v. Their number n' ranges from $n \cdot 0.1\%$ to $n \cdot 5\%$.

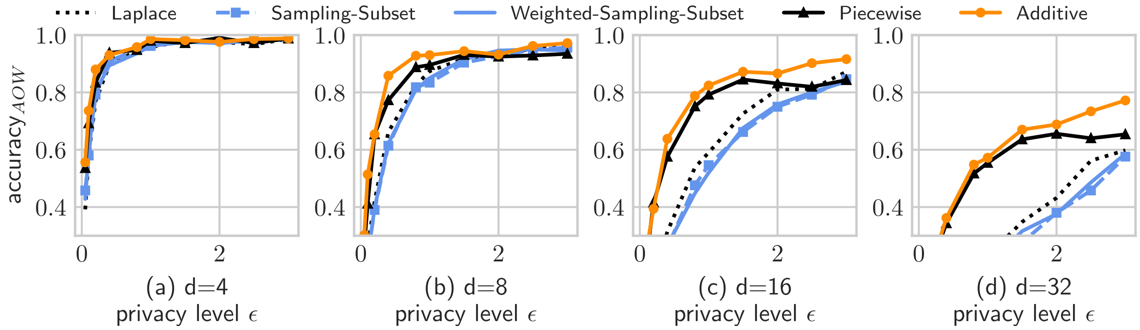
Adversarial private views in the simulation of view disguise attack are generated so that the 2nd-rank candidate A_{j_2} (in the non-adversarial aggregation result) benefits most. That is, we assume the adversary has the prior knowledge of 1st and 2nd ranked candidates, and each adversarial private view $\tilde{v}$ has maximum $\tilde{v}_{j_2} - \tilde{v}_{j_1}$ among the domain of private view. Specifically for the Laplace mechanism that the private view's domain is $[-\infty, +\infty]^d$, we use 95% confidence interval of the Laplace distribution as filtered domain, and assign $\tilde{v}_{j_2} = \log(\frac{1}{1-0.95})\Delta + w_1$, $\tilde{v}_{j_1} = -\log(\frac{1}{1-0.95})\Delta + w_d$. The number of adversarial private views n'' in simulations ranges from $n \cdot 0.1\%$ to $n \cdot 5\%$.

Information about simulation parameters is summarized in Table 3. Our experiments focus on the most popular Borda and Nauru scoring rules, as the additive mechanism is theoretically proved to be utility optimal for Plurality/



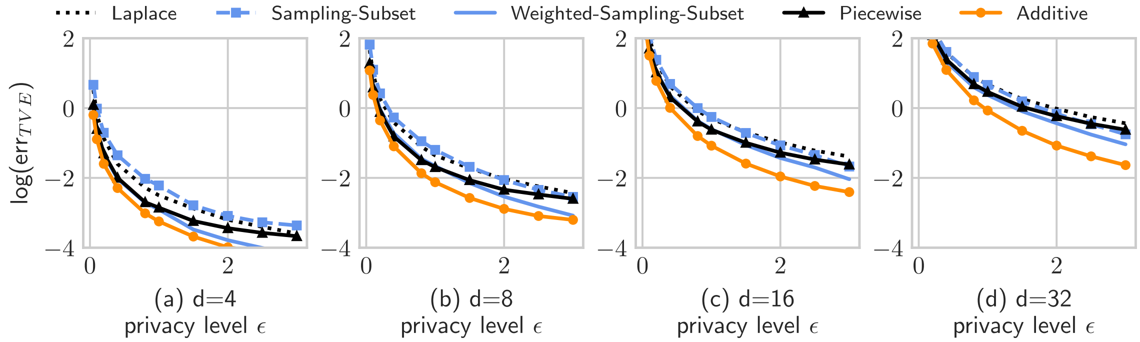
(a) d=4 privacy level ϵ (b) d=8 privacy level ϵ (c) d=16 privacy level ϵ (d) d=32 privacy level ϵ

Fig. 4. Total variation error under Borda rule over $4, 8, 16, 32$ candidates with 10000 users.



(a) d=4 privacy level ϵ (b) d=8 privacy level ϵ (c) d=16 privacy level ϵ (d) d=32 privacy level ϵ

Fig. 5. Accuracy of winner under Borda rule over $4, 8, 16, 32$ candidates with 10000 users.



(a) d=4 privacy level ϵ (b) d=8 privacy level ϵ (c) d=16 privacy level ϵ (d) d=32 privacy level ϵ

Fig. 6. Total variation error under Nauru rule over $4, 8, 16, 32$ candidates with 10000 users.

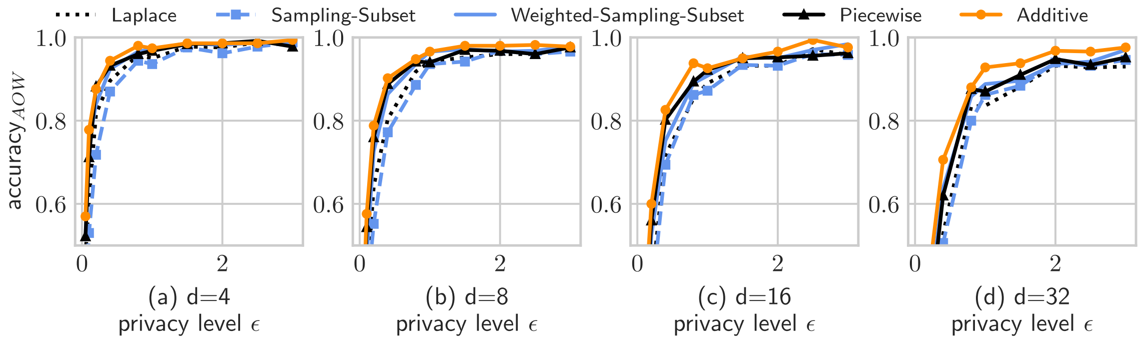
Anti-plurality rules. Every experimental result is the average value of 400 repeated simulations.

Evaluation Metrics. We use utility metrics in Section 2.4 to evaluate the performance of mechanisms in the non-adversarial and adversarial settings. Since the mean squared error and the robustness metrics of mechanisms are theoretically analyzed and numerically compared in former sections, their results are omitted.

6.2 Non-Adversarial Results

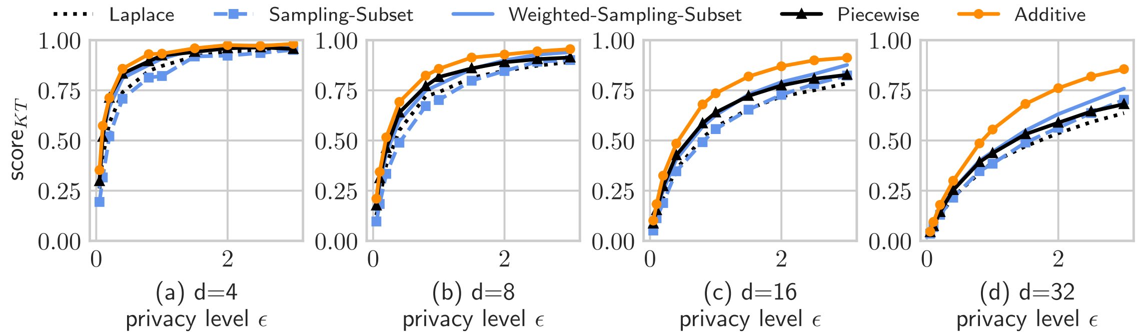
Varying Number of Candidates. Simulated with $n = 10000$ users, the experimental results under the Borda/Nauru rule with varying number of candidates are presented in Figs. 4, 5, 6, 7, and 8 respectively. When compared to the Laplace mechanism, the additive mechanism averagely reduces $\mathrm{err}_{\mathrm{TVE}}$ by 50%. The performance discrepancy between the weighted sampling and additive mechanism grows with the number of candidates, which confirms our theoretical analyses of mean squared error bounds. The piecewise mechanism and the additive mechanism have comparable performances under the Borda rule. While under the Nanru rule, the additive mechanism outperforms the piecewise mechanism by about 50% on $\mathrm{err}_{\mathrm{TVE}}$ and about 0.15 on $\mathrm{score}_{\mathrm{KT}}$. This implies that treating preference data as a permutation (in the additive mechanism) is more appropriate than treating it as $[w_d, w_1]^d$ (in the piecewise mechanism), especially when $\sum_{i=[1,d]}(w_i - w_d)$ is much smaller than $d \cdot (\mathbf{w}_1 - \mathbf{w}_d)$ (e.g., with the Nauru rule).

Varying Number of Users. Simulated with $d = 8$ candidates, experimental results with $n = 1000$ users are demonstrated in Figs. 13 and 15, experimental results with $n = 100\,000$ users are showed in Figs. 14 and 16. Comparing them with results on $n = 10000$ users, it is observed that increasing the number of users improves the utility performance significantly. The additive mechanism is still practical when there are relatively few users



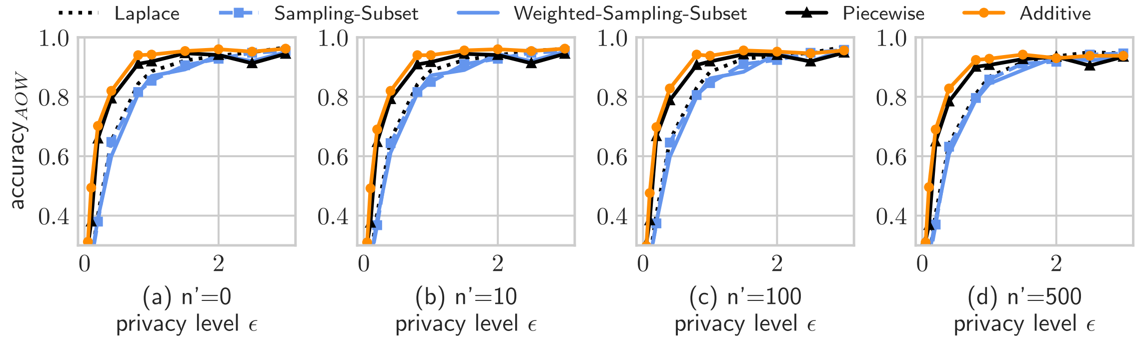
(a) d=4 privacy level ϵ (b) d=8 privacy level ϵ (c) d=16 privacy level ϵ (d) d=32 privacy level ϵ

Fig. 7. Accuracy of winner under Nauru rule over $4, 8, 16, 32$ candidates with 10000 users.



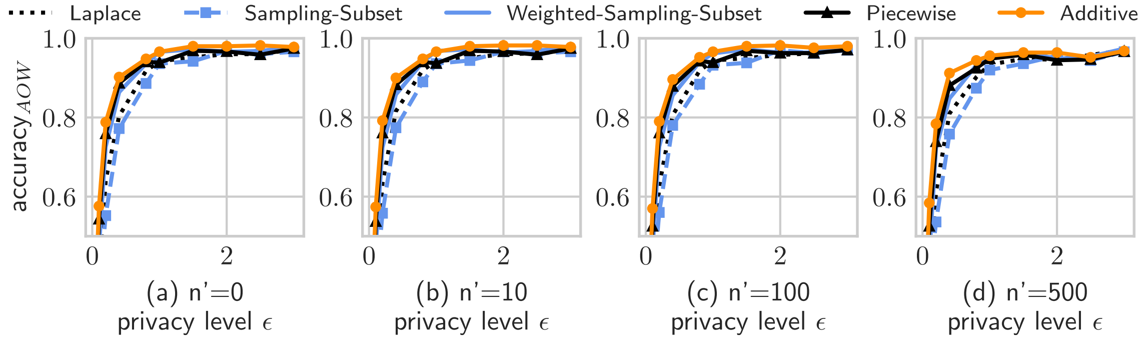
(a) d=4 privacy level ϵ (b) d=8 privacy level ϵ (c) d=16 privacy level ϵ (d) d=32 privacy level ϵ

Fig. 8. Kendall's tau under Nauru rule over $4, 8, 16, 32$ candidates with 10000 users.



(a) n'=0 privacy level ϵ (b) n'=10 privacy level ϵ (c) n'=100 privacy level ϵ (d) n'=500 privacy level ϵ

Fig. 9. Accuracy of winner under Borda rule with 10000 honest users and $n' = 0, 10, 100, 500$ adversarial data.



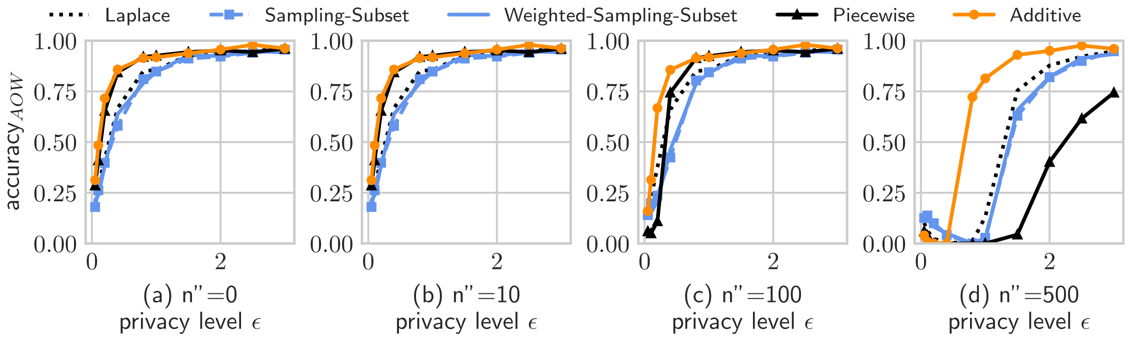
(a) n'=0 privacy level ϵ (b) n'=10 privacy level ϵ (c) n'=100 privacy level ϵ (d) n'=500 privacy level ϵ

Fig. 10. Accuracy of winner under Nauru rule with 10000 honest users and $n' = 0, 10, 100, 500$ adversarial data.

(e.g., $n = 1000$), and achieve more 75% accuracy when $\epsilon \geq 1.0$. When the number of users is 100 000, all mechanisms achieve nearly 100% accuracy of winner and 0.8 Kendall's tau correlation even when the privacy budget is relatively low (e.g., $0.2 < \epsilon < 1.0$).

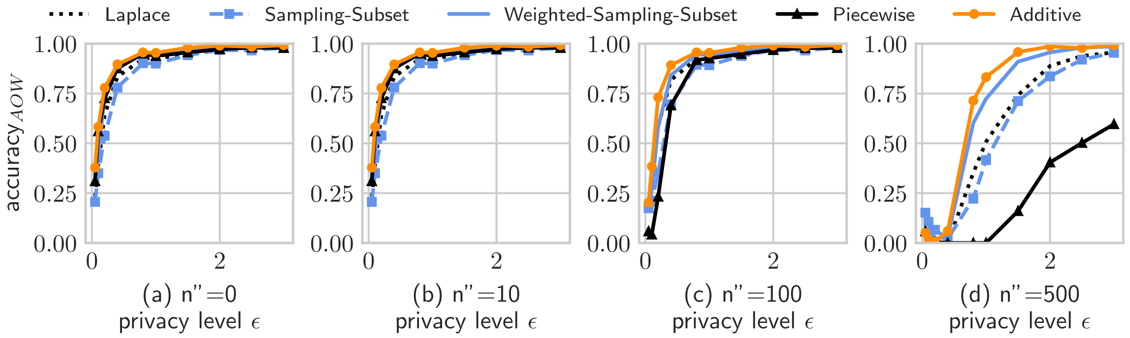
6.3 Data Amplification Attack

Simulated with $d = 8$ candidates and $n = 10000$ benign users, the results under Borda/rule with extra $n' = 0, 10, 100, 500$ adversarial users are presented in Figs. 9 and 10 respectively. Results show that less than 1% adversarial



(a) n''=0 privacy level ϵ (b) n''=10 privacy level ϵ (c) n''=100 privacy level ϵ (d) n''=500 privacy level ϵ

Fig. 11. Accuracy of winner under Borda rule with 10000 honest users and $n'' = 0, 10, 100, 500$ adversarial views.



(a) n''=0 privacy level ϵ (b) n''=10 privacy level ϵ (c) n''=100 privacy level ϵ (d) n''=500 privacy level ϵ

Fig. 12. Accuracy of winner under Nauru rule with 10000 honest users and $n'' = 0, 10, 100, 500$ adversarial views.

data has few effects impacts on the aggregated result, but more than 5% adversarial data will significantly harm the utility of the result. The additive mechanism outperforms other approaches in all adversarial settings with fraudulent data.

6.4 View Disguise Attack

Simulated with $d = 8$ candidates and $n = 10000$ benign users, the experimental results under Borda/Nauru rule with extra $n'' = 0, 10, 100, 500$ adversarial private views are presented in Figs. 11 and 12 respectively. Results show that less than 0.1% adversarial views won't have effective impacts on the aggregated result, but more than 1% adversarial views will significantly decrease the utility of the result. The additive mechanism outperforms other approaches in all adversarial settings with disguised private views. Compared with the results under data amplification attacks, the aggregation result is more sensitive to view disguise attacks, and the piecewise mechanism is extremely fragile to the view disguise attacks. It is also observed that the utility of the additive mechanism drops much slower than other approaches when number of adversarial views gets larger, this implies the additive mechanism is more robust to view disguise attacks.

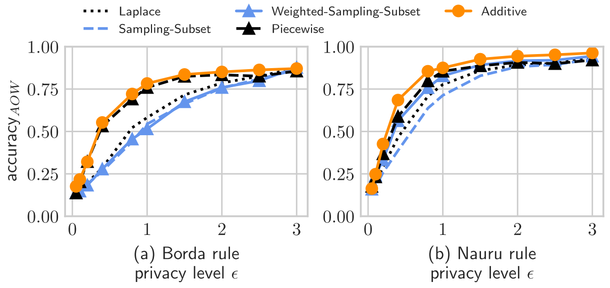
7 Related Work

Security requirements in data aggregation systems cover many aspects, such as privacy, verifiability and robustness. This section reviews some representative works on the privacy, anonymity, and robustness in the area of data aggregation, then retrospects recent works on privacy preserving data analyses within the differential privacy framework.

7.1 Security in Data Aggregation Systems

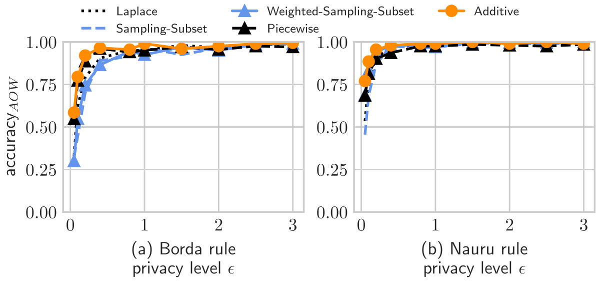
7.1.1 Privacy/Anonymity

Since the seminal work of Chaum [26], plenty of cryptographic schemes have contributed to keeping users or (and) user data secret in electronic aggregation systems. Schemes based on homomorphic encryption operate on encrypted user data to compute the summation/average without



(a) Borda rule privacy level ϵ (b) Nauru rule privacy level ϵ

Fig. 13. Accuracy of winner under Borda and Nauru rules over 8 candidates with 1000 users.



(a) Borda rule privacy level ϵ (b) Nauru rule privacy level ϵ

Fig. 14. Accuracy of winner under Borda and Nauru rules over 8 candidates with 100000 users.

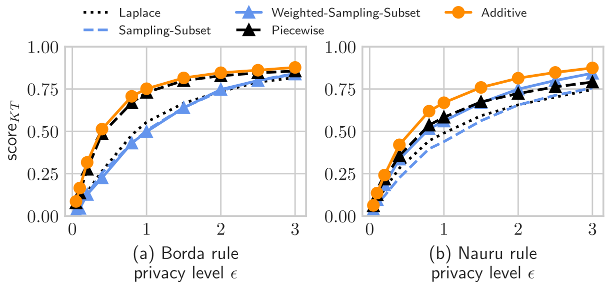


Fig. 15. Kendall's tau of under Borda and Nauru rules over 8 candidates with 1000 users.

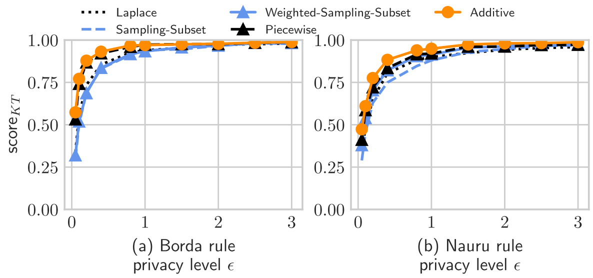


Fig. 16. Kendall's tau of Borda and Nauru rules over 8 candidates with 100000 users.

knowing plain-text of true data (e.g., ElGamal encryption in [27], [28], Paillier encryption in [29]), hence keeps data private from the aggregator or adversaries. Further combining cryptographic techniques with anonymous channels (e.g, the mixnet in [30], [31], [32], [33]) that randomly shuffles a bundle of messages from users, user data (or ciphertexts) are then unlinkable to source users.

Another line of works that could be employed for privacy preserving aggregation systems is data perturbation, which uses techniques of generalization (e.g., k-anonymity [34] in [35]) and randomization (e.g., Gaussian noise adding [36], randomized response [37] and differential privacy [5]) to hide the exact value of each user's true data or the aggregated result. Compared to cryptographic techniques providing computational secrecy and anonymity, data perturbation approaches are often much more efficient. Among them, classic privacy notions and techniques like k-anonymity and Gaussian noise adding have shown to be risky for adversaries with prior knowledge [38], [39].

7.1.2 Robustness

Consider strategic behaviors in aggregation systems like data manipulation, fraud and bribery, many works have contributed to finding counter-measures for various scoring rules. One approach is putting restrictions on users' preference. Specifically, works of [40], [41] show that aggregation with single peak preference and quasilinear preference is truthful and non-manipulatable.

Another anti-manipulation approach is ensuring the computational hardness of finding constructive/deconstructive manipulation strategies (e.g., in [42], [43]). However, for preference data with positional scoring rules considered in this work, there exist simple greedy algorithms finding strategic preference data that manipulate the result in polynomial time [44]. There are also some works propose to randomize the aggregation process (e.g., sampling users at random) for mitigating manipulation attacks, but the utility of the aggregation result will be severely harmed [45]. As a comparison, this work introduces randomness to preference data for the purpose of privacy preserving, it is demonstrated that local differential privacy helps to defend against data manipulation but makes the aggregation result more vulnerable to fraudulent data (see Section 8.1 for discussion).

7.2 Local Private Data Aggregation

Originated from the differential privacy [5] in the centralized setting, when defining neighboring datasets as any pairs of values individuals may hold, differential privacy is preserved in the local setting [6] (LDP). Because of the solidness of privacy guarantee and flexibility for deployment, LDP has gained massive attention from both industry and academy. Giant internet service providers are collecting user preference (e.g., browser's homepage) and usage records (e.g., typed words) from their users in the local differential privacy manner, such as Google [7], [46], Apple [47], [48], and Microsoft [8]. Research works have explored local private data analyses and modeling tasks on various kinds of data, such as distribution estimation on categorical data [12], [22], [49] and set-valued data [17], [50], [51], joint distribution estimation and frequent itemset mining on multidimensional data [46], [52], [53], [54], [55], mean estimation on numerical data [6], [19]. There are also theoretical contributions to give lower error bounds of local private analyses (e.g., in [6], [51], [56], [57], [58], [59], [60]) with categorical, set-valued data or other data with certain l_1-norm and l_2-norm. However, the lower bounds for preference data (with general but fixed score vectors) have not been investigated.

It's worth noting that under some specific scoring rules, such as plurality/k-approval rules (see Section 2), the preference data can be directly seen as categorical or set-valued data and then processed with existing approaches (e.g., in [12], [22], [50], [61]). This work intends to deal with preference data with arbitrary design of scoring. The local private preference data aggregation problem can also be cast as the multi-dimensional mean estimation problem, one approach to which is adding Laplace noises to every score (see detail in Section 3), another is first randomly sampling one (data-independent) candidate without knowing the value of preference data and then adding Laplace noises to the candidate's points (e.g., in [19], [62], [63]).

Recent works [64], [65] explore pair-wise or partial-wise comparison for ranking data aggregation, but have no convergence guarantees, while this work focuses on preference data with scoring rules and provides theoretical utility/robustness guarantees.

Robustness of Locally Private Mechanisms. Existing works on local differential privacy focus mostly on the utility aspect, some of which may also consider computational and communication efficiency. This work instead calls for attention to the robustness aspect in local private data analyses, which is severe in real-world systems (e.g., the RAPPOR of Google [7] and iOS/macOS data collection of Apple [9]) where there are malicious and adversarial clients. The preliminary version of this work [66] initializes the study of data poisoning attacks on locally private mechanisms. The data amplification/view disguise attacks in this work also apply to other ϵ-LDP mechanisms. Latterly, [67], [68] and [69] study these ϵ-LDP attacks on distribution/heavy-hitter estimation protocols. The recent work [70] considers poisoning attacks/defenses on key-value analyzing protocols.

8 Discussion

In this section, we first review the relation between utility, robustness and level of indistinguishability, and then build numerical connections between utility and robustness.

8.1 Robustness versus Indistinguishability

Considering the data amplification attack, according to the definition of ϵ-LDP, the probabilistic outputs are at most e^{ϵ}-distinguishable regardless of the input of manipulated/ true data. Therefore, a more rigid level of privacy protection in ϵ-LDP has the advantage of limiting an adversary's *constructive/directive* power. Thus the resulting scores are less led by preferences of the manipulated data. However, when an adversary could falsely contribute an extra data to the aggregation system, a more rigorous level of privacy means a larger magnitude of its private view, thus the expected amount of scores an adversary added to the resulting scores is amplified. As shown in the robustness analyses of the Laplace mechanism and our proposed mechanisms, the expected magnitude of private view is linear to $\frac{1}{\epsilon}$, hence the *deconstructive* power of a possible adversary gets larger due to the noises injecting for privacy preservation. Specifically for the additive mechanism, it has a desirable property that the maximum possible magnitude of a private view is equal to the expected magnitude, which does not hold in the Laplace and weighted sampling mechanisms.

Now consider the view disguise attack that an adversary directly sends a fraud private view. In this case, the power of the adversary is closely related to the private view's domain range, from which the adversary could choose a value to destroy or to reform the final result. As discussed, the private view's domain of classical Laplace mechanism spreads to $\mathbb{R}^d$, which is reduced to $[-\Theta(\frac{1}{\epsilon}), \Theta(\frac{1}{\epsilon})]^d$ in the additive mechanism. These results suggest that a higher level of privacy preservation empowers a higher ability for an adversary to manipulate the aggregation result.

8.2 Utility Versus Robustness

Consider the robustness metric of expected magnitude risk_{EM} and the utility metric of mean squared error err_{MSE}, we have:

$$\mathrm{risk}_{EM} \leq \frac{\sqrt{d \cdot n \cdot \mathrm{err}_{MSE}} + \sum_{j\in[1,d]} |\mathbf{w}_j|}{n},$$

which is derived according to the convexity of the square root. This inequality implies that a mechanism with better utility performance usually has better robustness performance. Therefore, improving utility performance during the design of an ϵ-LDP mechanism often helps improving robustness.

9 Conclusion

Considering adversarial behaviors existed in real-world private data aggregation systems, this work pays attention to both the utility and robustness aspects of privacy preserving mechanisms. Adversarial behaviors tailed for the local privacy setting are classified into data amplification attack and view disguise attack, which are then quantitatively measured by their manipulation power over the aggregation result.

In the context of preference data aggregation with local privacy, we derived minimax error bounds and proposed an optimized mechanism: the Additive mechanism, to improve utility and robustness upon the classical Laplace mechanism and sampling-based mechanisms. Besides theoretical analyses showing a factor of d (or d^2) reduction in estimation-error/manipulation-risk bounds, the performance improvements are further validated by extensive experiments in both non-adversarial and adversarial scenarios.

This article also discusses subtle relations among utility, robustness and indistinguishability, and calls for further researches demystifying interactions between these fundamental requirements in real-world data aggregation systems.

References

[1] D. Black, R. A. Newing, I. McLean, A. McMillan, and B. L. Monroe, *The theory of committees and elections*, Cambridge, U.K.: Cambridge Univ., 1958.

[2] P. Voigt and A. Von dem Bussche, "The EU general data protection regulation (GDPR)," *A Practical Guide*, 1st Ed., Berlin, Germany: Springer, 2017.

[3] E. Goldman, "An introduction to the california consumer privacy act (CCPA)," *Santa Clara Univ. Legal Stud. Res. Paper*, 2020.

[4] D. Chaum, C. Crépeau, and I. Damgard, "Multiparty unconditionally secure protocols," in *Proc. 20th Annu. ACM Symp. Theory Comput.*, 1988, pp. 11–19.

[5] C. Dwork, "Differential privacy," *Encyclopedia of Cryptography and Security*, Berlin, Germany: Springer, 2011, pp. 338–340.

[6] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Local privacy and statistical minimax rates," in *Proc. IEEE 54th Annu. Symp. Found. Comput. Sci.*, 2013, pp. 429–438.

[7] Ú. Erlingsson, V. Pihur, and A. Korolova, "Rappor: Randomized aggregatable privacy-preserving ordinal response," in *Proc. ACM SIGSAC Conf. Comput. Commun. Secur.*, 2014, pp. 1054–1067.

[8] B. Ding, J. Kulkarni, and S. Yekhanin, "Collecting telemetry data privately," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2017, pp. 3574–3583.

[9] J. Tang, A. Korolova, X. Bai, X. Wang, and X. Wang, "Privacy loss in apple's implementation of differential privacy on macos 10.12," 2017, *arXiv:1709.02753*.

[10] S. Kessler, J. Hoff, and J.-C. Freytag, "SAP HANA goes private: From privacy research to privacy aware enterprise analytics," *Proc. VLDB Endowment*, vol. 12, no. 12, pp. 1998–2009, 2019.

[11] C. Dwork, F. McSherry, K. Nissim, and A. Smith, "Calibrating noise to sensitivity in private data analysis," in *Proc. Theory Cryptogr. Conf.*, 2006, pp. 265–284.

[12] P. Kairouz, K. Bonawitz, and D. Ramage, "Discrete distribution estimation under local privacy," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 2436–2444.

[13] M. Bun, J. Nelson, and U. Stemmer, "Heavy hitters and the structure of local privacy," in *Proc. Symp. Princ. Database Syst.*, 2018, pp. 435–447.

[14] H. Husain, B. Balle, Z. Cranko, and R. Nock, "Local differential privacy for sampling," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2020, pp. 3404–3413.

[15] M. Yang, L. Lyu, J. Zhao, T. Zhu, and K.-Y. Lam, "Local differential privacy and its applications: A comprehensive survey," 2020, *arXiv:2008.03686*.

[16] M. G. Kendall, *Rank Correlation Methods*. Berlin, Germany: Springer, 1948.

[17] Z. Qin, Y. Yang, T. Yu, I. Khalil, X. Xiao, and K. Ren, "Heavy hitter estimation over set-valued data with local differential privacy," in *ACM SIGSAC Conf. Comput. Commun. Secur.*, 2016, pp. 192–203.

[18] X. Gu, M. Li, Y. Cheng, L. Xiong, and Y. Cao, "{PCKV }: Locally differentially private correlated { Key-Value} data collection with optimized utility," in *Proc. USENIX Secur.*, 2020, pp. 967–984.

[19] N. Wang et al., "Collecting and analyzing multidimensional data with local differential privacy," in *Proc. IEEE Int. Conf. Data Eng.*, 2019, pp. 638–649.

[20] Q. Ye, H. Hu, X. Meng, H. Zheng, K. Huang, C. Fang, and J. Shi, "PrivKVM*: Revisiting key-value statistics estimation with local differential privacy," *IEEE Trans. Dependable Secure Comput.*, early access, Aug. 27, 2021, doi: 10.1109/TDSC.2021.3107512.

[21] T. Wang, J. Blocki, N. Li, and S. Jha, "Locally differentially private protocols for frequency estimation," in *Proc. 26th {USENIX} Secur. Symp.*, 2017, pp. 729–745.

[22] M. Ye and A. Barg, "Optimal schemes for discrete distribution estimation under locally differential privacy," *IEEE Trans. Inf. Theory*, vol. 64, no. 8, pp. 5662–5676, Aug. 2018.

[23] J. C. Duchi, M. I. Jordan, and M. J. Wainwright, "Minimax optimal procedures for locally private estimation," *J. Amer. Statist. Assoc.*, vol. 113, no. 521, pp. 182–201, 2018.

[24] B. Yu, "Assouad, fano, and le cam," in *Festschrift for Lucien Le Cam*. Berlin, Germany: Springer, 1997, pp. 423–435.

[25] Y. Kawamoto and T. Murakami, "Differentially private obfuscation mechanisms for hiding probability distributions," 2018, *arXiv:1812.00939*.

[26] D. L. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Commun. ACM*, vol. 24, no. 2, pp. 84–90, 1981.

[27] Y. Mu and V. Varadharajan, "Anonymous secure e-voting over a network," in *Proc. IEEE 14th Annu. Comput. Secur. Appl. Conf.*, 1998, pp. 293–299.

[28] M. Hirt, "Receipt-free k-out-of-l voting based on elgamal encryption," in *Towards Trustworthy Elections*, Berlin, Germany: Springer, 2010, pp. 64–82.

[29] Z. Xia, S. A. Schneider, J. Heather, and J. Traoré, "Analysis, improvement, and simplification of prêt à voter with paillier encryption," in *Proc. Conf. Electron. Voting Technol.*, 2008, pp. 1–15.

[30] C. Park, K. Itoh, and K. Kurosawa, "Efficient anonymous channel and all/nothing election scheme," in *Proc. Workshop Theory Appl. Cryptographic Techn.*, 1993, pp. 248–259.

[31] M. Abe, "Universally verifiable mix-net with verification work independent of the number of mix-servers," in *Proc. Int. Conf. Theory Appl. Cryptographic Techn.*, 1998, pp. 437–447.

[32] B. Lee, C. Boyd, E. Dawson, K. Kim, J. Yang, and S. Yoo, "Providing receipt-freeness in mixnet-based voting protocols," in *Proc. Int. Conf. Informat. Secur. Cryptol.*, 2003, pp. 245–258.

[33] P. Bulens et al., "Running mixnet-based elections with helios," in *Proc. Electron. Voting Technol. Workshop Trustworthy Elections*, 2011, pp. 6–6.

[34] L. Sweeney, "K-anonymity: A model for protecting privacy," *Int. J. Uncertainty, Fuzziness Knowl.-Based Syst.*, vol. 10, no. 05, pp. 557–570, 2002.

[35] Q. Zhao and Y. Liu, "E-voting scheme using secret sharing and k-anonymity," in *Proc. Int. Conf. Broadband Wirel. Comput., Commun. Appl.*, 2016, pp. 893–900.

[36] C. Dwork, K. Kenthapadi, F. McSherry, I. Mironov, and M. Naor, "Our data, ourselves: Privacy via distributed noise generation," in *Proc. Annu. Int. Conf. Theory Appl. Cryptographic Techn.*, 2006, pp. 486–503.

[37] S. L. Warner, "Randomized response: A survey technique for eliminating evasive answer bias," *J. Amer. Statist. Assoc.*, vol. 60, no. 309, pp. 63–69, 1965.

[38] H. Kargupta, S. Datta, Q. Wang, and K. Sivakumar, "On the privacy preserving properties of random data perturbation techniques," in *Proc. IEEE Int. Conf. Des. Mater.*, 2003, pp. 99–106.

[39] N. Li, T. Li, and S. Venkatasubramanian, "T-closeness: Privacy beyond k-anonymity and l-diversity," in *Proc. IEEE Int. Conf. Data Eng.*, 2007, pp. 106–115.

[40] M. Dummett and R. Farquharson, "Stability in voting," *Econometrica: J. Econometric Soc.*, vol. 29, no. 1, pp. 33–43, 1961.

[41] E. Ephrati et al., "Multi-agent planning as a dynamic search for social consensus," in *Proc. Int. Joint Conf. Artif. Intell.*, 1993, pp. 423–429.

[42] J. J. Bartholdi III, C. A. Tovey, and M. A. Trick, "How hard is it to control an election?," *Math. Comput. Modelling*, vol. 16, no. 8/9, pp. 27–40, 1992.

[43] P. Faliszewski, E. Hemaspaandra, and L. A. Hemaspaandra, "How hard is bribery in elections?," *J. Artif. Intell. Res.*, vol. 35, pp. 485–532, 2009.

[44] J. J. Bartholdi, C. A. Tovey, and M. A. Trick, "The computational difficulty of manipulating an election," *Social Choice Welfare*, vol. 6, no. 3, pp. 227–241, 1989.

[45] A. Gibbard et al., "Manipulation of schemes that mix voting with chance," *Econometrica*, vol. 45, no. 3, pp. 665–681, 1977.

[46] G. Fanti, V. Pihur, and Ú. Erlingsson, "Building a RAPPOR with the unknown: Privacy-preserving learning of associations and data dictionaries," *Proc. Annu. Privacy Enhancing Technol. Symp.*, vol. 2016, no. 3, pp. 41–61, 2016.

[47] A. G. Thakurta et al., "Emoji frequency detection and deep link frequency," Jul. 11 2017, U.S. Patent 9,705,908.

[48] A. G. Thakurta et al., "Learning new words," Mar. 28 2019, U.S. Patent App. 16/159,473.

[49] Z. Li, T. Wang, M. Lopuhaä-Zwakenberg, N. Li, and B. Škoric, "Estimating numerical distributions under local differential privacy," in *Proc. ACM SIGMOD Int. Conf. Manage. Data*, 2020, pp. 621–635.

[50] S. Wang, L. Huang, Y. Nie, P. Wang, H. Xu, and W. Yang, "Privset: Set-valued data analyses with locale differential privacy," in *Proc. IEEE INFOCOM Conf. Comput. Commun.*, 2018, pp. 1088–1096.

[51] S. Wang, Y. Qian, J. Du, W. Yang, L. Huang, and H. Xu, "Set-valued data publication with local privacy: Tight error bounds and efficient mechanisms," *Proc. VLDB Endowment*, vol. 13, pp. 1234–1247, 2020.

[52] X. Ren, C.-M. Yu, W. Yu, S. Yang, X. Yang, and J. McCann, "High-dimensional crowdsourced data distribution estimation with local privacy," in *Proc. IEEE Int. Conf. Comput. Informat. Technol.*, 2016, pp. 226–233.

[53] X. Ren et al., "LoPub: High-dimensional crowdsourced data publication with local differential privacy," *IEEE Trans. Inf. Forensics Secur.*, vol. 13, no. 9, pp. 2151–2166, Sep. 2018.

[54] G. Cormode, T. Kulkarni, and D. Srivastava, "Answering range queries under local differential privacy," *Proc. VLDB Endowment*, vol. 12, pp. 1126–1138, 2019.

[55] J. Yang, T. Wang, N. Li, X. Cheng, and S. Su, "Answering multi-dimensional range queries under local differential privacy," *Proc. VLDB Endowment*, vol. 14, pp. 378–390, 2020.

[56] P. Kairouz, S. Oh, and P. Viswanath, "Extremal mechanisms for local differential privacy," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 492–542.

[57] J. Ullman, "Tight lower bounds for locally differentially private selection," 2018, *arXiv:1802.02638*.

[58] J. Duchi and R. Rogers, "Lower bounds for locally private estimation via communication complexity," in *Proc. Conf. Learn. Theory*, 2019, pp. 1161–1191.

[59] D. Wang and J. Xu, "Tight lower bound of sparse covariance matrix estimation in the local differential privacy model," *Theor. Comput. Sci.*, vol. 815, pp. 47–59, 2020.

[60] A. Rohde et al., "Geometrizing rates of convergence under local differential privacy constraints," *Ann. Statist.*, vol. 48, no. 5, pp. 2646–2670, 2020.

[61] Z. Yan, J. Liu, and S. Liu, "Dpwevote: Differentially private weighted voting protocol for cloud-based decision-making," *Enterprise Informat. Syst.*, vol. 13, no. 2, pp. 236–256, 2019.

[62] T. T. Nguyên, X. Xiao, Y. Yang, S. C. Hui, H. Shin, and J. Shin, "Collecting and analyzing data from smart device users with local differential privacy," 2016, *arXiv:1606.05053*.

[63] Q. Ye, H. Hu, X. Meng, and H. Zheng, "PrivKV: Key-value data collection with local differential privacy," in *Proc. IEEE Conf. Symp. Secur. Privacy*, 2019, pp. 317–331.

[64] J. Yang, X. Cheng, S. Su, R. Chen, Q. Ren, and Y. Liu, "Collecting preference rankings under local differential privacy," in *Proc. IEEE Int. Conf. Data Eng.*, 2019, pp. 1598–1601.

[65] Z. Yan, G. Li, and J. Liu, "Private rank aggregation under local differential privacy," 2019, *arXiv:1908.04486*.

[66] S. Wang et al., "Aggregating votes with local differential privacy: Usefulness, soundness versus indistinguishability," 2019, *arXiv:1908.04920*.

[67] X. Cao, J. Jia, and N. Z. Gong, "Data poisoning attacks to local differential privacy protocols," 2019, *arXiv:1911.02046*.

[68] X. Cao, J. Jia, and N. Z. Gong, "Data poisoning attacks to local differential privacy protocols," in *Proc. USENIX Secur.*, 2021, pp. 947–964.

[69] A. Cheu, A. Smith, and J. Ullman, "Manipulation attacks in local differential privacy," in *Proc. IEEE Symp. Secur. Privacy*, 2021, pp. 883–900.

[70] Y. Wu, X. Cao, J. Jia, and N. Z. Gong, "Poisoning attacks to local differential privacy protocols for Key-Value data," in *Proc. USENIX Secur.*, 2022, pp. 519–536.

Shaowei Wang received the PhD degree in computer science from the University of Science and Technology of China (USTC), in 2019. He is an associate professor with the Institute of Artificial Intelligence and Blockchain, Guangzhou University. His research interests include data privacy and federated learning.

Xuandi Luo is currently working toward the master degree with the Guangzhou University. His research focuses on differential privacy and deep learning.

Yuqiu Qian received the PhD degree in computer science from the University of Hong Kong, in 2019. She is currently a senior researcher with Tencent in Shenzhen, China. Her research interests include distributed computing, recommendation systems, and machine learning.

Jiachun Du received the PhD degree in mathematic from the University of Science and Technology of China (USTC), in 2007. He is currently a senior researcher with Tencent in Shenzhen, China. His research interests include recommendation systems, and game data mining.

Wenqing Lin received the PhD degree in computer science from Nanyang Technological University, in 2015. He is currently a senior researcher with Tencent in Shenzhen, China. His research interests include graph databases and data mining.

Wei Yang (Member, IEEE) received the PhD degree in computer science from the University of Science and Technology of China (USTC), in 2007. He is an associate professor with the School of Computer Science and Technology, USTC. His research interests include information security, quantum information, and human-computer interaction.