(Locally) Differentially Private Combinatorial Semi-Bandits

Xiaoyu Chen [* 1] Kai Zheng [* 1 2] Zixin Zhou [3] Yunchang Yang [4] Wei Chen [5] Liwei Wang [1 4]

Abstract

In this paper, we study Combinatorial Semi-Bandits (CSB) that is an extension of classic Multi-Armed Bandits (MAB) under Differential Privacy (DP) and stronger Local Differential Privacy (LDP) setting. Since the server receives more information from users in CSB, it usually causes additional dependence on the dimension of data, which is a notorious side-effect for privacy preserving learning. However for CSB under two common smoothness assumptions (Kveton et al., 2015; Chen et al., 2016), we show it is possible to remove this side-effect. In detail, for B_∞-bounded smooth CSB under either ε-LDP or ε-DP, we prove the optimal regret bound is $\Theta(\frac{mB_\infty^2 \ln T}{\Delta \varepsilon^2})$ or $\tilde{\Theta}(\frac{mB_\infty^2 \ln T}{\Delta \varepsilon})$ respectively, where T is time period, Δ is the gap of rewards and m is the number of base arms, by proposing novel algorithms and matching lower bounds. For B_1-bounded smooth CSB under ε-DP, we also prove the optimal regret bound is $\tilde{\Theta}(\frac{mKB_1^2 \ln T}{\Delta \varepsilon})$ with both upper bound and lower bound, where K is the maximum number of feedback in each round. All above results nearly match corresponding non-private optimal rates, which imply there is no additional price for (locally) differentially private CSB in above common settings.

1. Introduction

Stochastic Multi-Armed Bandits (MAB) (Bubeck et al., 2012) is a fundamental problem in machine learning with wide applications in real world. In stochastic MAB, there is an unknown underlying distribution over $[0, 1]^m$ for m base arms and a learner (or called a server) interacts with the environment for T rounds. At each round, the environment draws random rewards from the distribution for m base arms. At the same time, the learner chooses one of m base arms based on previously collected information, and receives the reward of chosen arm. The goal of the learner is to minimize the regret, measured as the difference between the reward of best fixed base arm and the learner's total reward in expectation. Multi-Armed Bandits has been used in recommendation systems, clinical trial, etc. However, many of these applications rely heavily on users' sensitive data, which raise great concerns about data privacy. For example, in recommendation systems, observations at each round represent some preferences of the user over the recommended item set, which is the personal information of user t and should be protected.

Since first proposed in 2006, Differential Privacy (DP) (Dwork et al., 2006) has become a gold-standard in privacy preserving machine learning (Dwork & Roth, 2014). We say an algorithm protects differential privacy if there is not much difference between outputs of this algorithm over two datasets with Hamming distance 1 (see Section 2 for the rigorous definition in the streaming setting). For ε-differentially private stochastic Multi-Armed Bandits, there has already been extensive studies (Mishra & Thakurta, 2015; Tossou & Dimitrakakis, 2016; Sajed & Sheffet, 2019). Based on classic non-private optimal UCB algorithm (Auer et al., 2002), as well as the tree-based aggregation technique to calculate private summation (Dwork et al., 2010), both Mishra & Thakurta (2015) and Tossou & Dimitrakakis (2016) designed algorithms under DP guarantee but with sub-optimal guarantee [1]. Recently, Sajed & Sheffet (2019) proposed a complex algorithm based on non-private Successive Elimination (Even-Dar et al., 2002) and sparse vector technique (Dwork & Roth, 2014) to achieve the optimal $\mathcal{O}(\frac{m \ln T}{\varepsilon \Delta})$ regret bound, where Δ is the minimum gap of rewards, and it matches both the non-private lower bound (Lai & Robbins, 1985) and the differentially private lower bound (Shariff & Sheffet, 2018) in common parameter regimes.

However, stochastic MAB is the simplest model for sequen-

---

[1]In fact, (Tossou & Dimitrakakis, 2016) achieved a better utility bound but under a weaker privacy guarantee compared with common differential privacy in the streaming setting.

tial decision making with uncertainty. There are many problems in real world that have a combinatorial nature among multiple arms and maybe even non-linear reward functions, such as online advertising, online shortest path, online social influence maximization, etc, which can be modeled via Combinatorial Semi-Bandits (CSB) (Chen et al., 2013; 2016; Lattimore & Szepesvári, 2018). In CSB, the learner chooses a super arm which is a set of base arms instead of a single base arm in MAB, and then observes the outcomes of the chosen arms as the feedback, and receive a reward determined by the chosen arms' outcomes. The reward can be a non-linear function in terms of these observations. Since many applications modeled via CSB also have issues about privacy leakage, in this paper, we study how to design private algorithms for Combinatorial Semi-Bandits under two common assumptions about non-linear rewards: B_∞-bounded smoothness and B_1-bounded smoothness (see section 2 for definitions.), which contain social influence maximization and linear CSB as important examples respectively (Kveton et al., 2015; Chen et al., 2016; Wang & Chen, 2017).

Main Difficulty: Compared with simple stochastic MAB, it is more difficult to design differentially private algorithms for CSB, due to its large action space and non-linear rewards. Though each super arm in CSB can be regarded as a base arm in stochastic MAB, a straightforward implementation of differentially private algorithms for stochastic MAB will lead to a dependence over the size of decision set for super arms, which can be exponentially large in terms of m. Besides above two differences, we receive observations of a set of base arms contained in the chosen super arm at each round, instead of a single base arm in MAB. Denote the maximum cardinality of a super arm as K, which means the sensitive data collected at each round is roughly in a K-dimensional L_∞ ball.

However, protecting differential privacy usually causes an additional dependence on the dimension of data for utility guarantee compared with corresponding non-private result, which is a notorious side-effect of DP, such as in differentially private empirical risk minimization (ERM) (Bassily et al., 2014), bandits linear optimization (Agarwal & Singh, 2017), online convex optimization and bandits convex optimization (Thakurta & Smith, 2013), etc. On one hand, in some cases such as differentially private ERM (Bassily et al., 2014), this additional dependence on the dimension is unavoidable. On the other hand, some researchers show it is possible to eliminate this side-effect if there are some extra structures, such as assumptions about restricted strong convexity, parameter set in L_1 norm, or generalized linear model with data bounded in L_2 norm, etc (Kifer et al., 2012; Smith & Thakurta, 2013; Jain & Thakurta, 2014; Talwar et al., 2015). In general, it is unclear whether it is possible to eliminate the side-effect about dimensional dependence

brought by privacy protection, let alone that our CSB setting does not have any extra structure mentioned above.

Besides, compared with differential privacy that admits the server to collect users' true data, local differential privacy (LDP) is a much stronger notion of privacy, which requires protecting data privacy before collection. Thus LDP is more practical and user-friendly compared with DP (Cormode et al., 2018). Intuitively, learning under LDP guarantee is more difficult as what we collect is already noisy. Moreover, eliminating the side-effect on the dimension is also more difficult under LDP guarantee even when we have some extra assumptions. For example, there are some negative results for locally differentially private sparse mean estimation (Duchi et al., 2016).

Our Contributions: Given above discussions, it seems hard to obtain nearly optimal regret for CSB under DP and much stronger LDP guarantee. Somewhat surprisingly, without any additional structure assumption such as sparsity, we show that it is indeed possible to achieve nearly optimal regret bound, by designing private algorithms with theoretical upper bounds and proving corresponding lower bounds in each case. Our upper bounds (nearly) match both our private lower bounds and non-private lower bounds (see Table 1 for an overview, where Δ is some gap defined in Section 3, $\mathcal{O}(\cdot)$ represents the upper bound, Θ represents both the upper bound and lower bound, and for $\tilde{\mathcal{O}}, \tilde{\Theta}$, we hide the poly-logarithmic dependence such as $\ln T, \ln m$). The main contributions of this paper are summarized as the follows:

(1) For B_∞-bounded smooth CSB under ε-LDP and ε-DP, we propose novel algorithms with regret bounds $\mathcal{O}(\frac{mB_\infty^2 \ln T}{\varepsilon^2 \Delta})$ and $\tilde{\mathcal{O}}(\frac{mB_\infty^2 \ln T}{\varepsilon \Delta})$ respectively, and prove nearly matching lower bounds;

(2) For B_1-bounded smooth CSB under ε-DP, we propose an algorithm with regret bound $\tilde{\mathcal{O}}(\frac{mKB_1^2 \ln T}{\varepsilon \Delta})$ and nearly matching lower bound.

In Section 2, we provide some backgrounds in Combinatorial Semi-Bandits and (Local) Differential Privacy. Then in Section 3 and Section 4, we study both upper and lower bounds for (locally) differentially private B_∞-bounded smooth and B_1-bounded smooth CSB respectively. Finally, we conclude our main results in Section 5.

1.1. Other Related Work

Besides differentially private stochastic MAB, there are also some works considering adversarial MAB with DP guarantee (Thakurta & Smith, 2013; Tossou & Dimitrakakis, 2017; Agarwal & Singh, 2017). Later, Shariff & Sheffet (2018) study contextual linear bandits under a relaxed definition of DP called *Joint Differential Privacy*. Compared with DP, bandits learning with LDP guarantee is paid less attention to.

Problem	ε-LDP	ε-DP	Non-Private Result	
B_∞-Smooth CSB	$\Theta(\frac{mB_\infty^2 \ln T}{\varepsilon^2 \Delta})$	$\tilde{\Theta}(\frac{mB_\infty^2 \ln T}{\varepsilon \Delta})$	$\Theta(\frac{mB_\infty^2 \ln T}{\Delta})$	(Chen et al., 2016; Wang & Chen, 2017)
B_1-Smooth CSB	$\mathcal{O}(\frac{mK^2 B_1^2 \ln T}{\varepsilon^2 \Delta})$	$\tilde{\Theta}(\frac{mK B_1^2 \ln T}{\varepsilon \Delta})$	$\Theta(\frac{mK B_1^2 \ln T}{\Delta})$	(Kveton et al., 2015; Wang & Chen, 2017)

Table 1. Summary of Our Results for Private CSB. Θ represents matching upper bounds and lower bounds. $\mathcal{O}$ represents upper bounds. Our lower bound in DP setting is actually in an additive form, see Theorem 9. Here, we write it in a multiplicative form for simplicity, which is natural in common parameter regimes.

Only Gajane et al. (2018) study stochastic MAB under LDP guarantee. Recently, Basu et al. (2019) investigate relations about several variants of differential privacy in MAB setting, and prove some lower bounds. For non-private Combinatorial Semi-Bandits, there is an extension of study (György et al., 2007; Chen et al., 2013; 2016; Kveton et al., 2015; Combes et al., 2015; Wang & Chen, 2017; 2018).

2. Preliminaries

Now we detail the concrete setting studied in this paper.

2.1. Combinatorial Semi-Bandits

In a Combinatorial Semi-Bandits (CSB), there are m base arms (denote $[m] = \{1, 2, \ldots, m\}$), and a predefined decision set $\mathcal{S} \subset 2^m$, each element of which is a subset of $[m]$ with at most K base arms and is called a super arm or an action, i.e. $|S| \leqslant K$ for any $S \in \mathcal{S}$ and $|\cdot|$ represents the cardinality of a set. $\mathcal{D}$ is an underlying unknown distribution supported on $[0,1]^m$ with expectation $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_m)$. There are T rounds in total. At each round, the player chooses a super arm $S_t \in \mathcal{S}$, and the environment draws a fresh random outcome $\boldsymbol{X}_t = (X_{t,1}, \ldots, X_{t,m})$ from $\mathcal{D}$ independently of any other variables. Then the player receives a reward $R_t = R(S_t, \boldsymbol{X}_t)$ and observes the feedback $\{(i, X_{t,i}) | i \in S_t)\}$. We assume the reward function $R(\cdot, \cdot)$ satisfies following assumptions, which are common in either real applications or previous literature (Chen et al., 2016; Wang & Chen, 2018), such as Linear CSB, social influence maximization.

Assumption 1. *There exists a reward function $r_{\boldsymbol{\mu}}(S)$ such that $\mathbb{E}[R(S, \boldsymbol{X})] = r_{\boldsymbol{\mu}}(S)$ for any $S \in \mathcal{S}$, where the expectation is over the randomness of outcome $\boldsymbol{X}$ and $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{X}]$.*

Under above assumption, define $\mathrm{opt}_{\boldsymbol{\mu}} = \max_{S \in \mathcal{S}} r_{\boldsymbol{\mu}}(S)$ as the optimal reward if we know $\boldsymbol{\mu}$ in advance.

Assumption 2 (B_p-bounded smoothness). *There exists a constant B_p, such that for arbitrary super arm S, and two mean vectors $\boldsymbol{\mu}, \boldsymbol{\mu}'$, there is $|r_{\boldsymbol{\mu}}(S) - r_{\boldsymbol{\mu}'}(S)| \leqslant B_p \|\boldsymbol{\mu}_S - \boldsymbol{\mu}'_S\|_p)$, where $\boldsymbol{\mu}_S$ represents the truncated vector of $\boldsymbol{\mu}$ on subset S.*

Assumption 3 (Monotonicity). *For any $\boldsymbol{\mu}, \boldsymbol{\mu}'$ such that*

$\boldsymbol{\mu} \leqslant \boldsymbol{\mu}'$ *(element-wise compare), we have* $r_{\boldsymbol{\mu}}(S) \leqslant r_{\boldsymbol{\mu}'}(S)$.

Intuitively, Assumptions 2 and 3 are about the smoothness and monotonicity of expected reward function $r_{\boldsymbol{\mu}}(\cdot)$, which are critical to deal with non-linear rewards $r_{\boldsymbol{\mu}}(S)$.

In this paper, we mainly consider two norms: L_∞ norm $\|\cdot\|_\infty$ and L_1 norm $\|\cdot\|_1$. Important examples that satisfy B_∞-bounded smoothness include social influence maximization and Probabilistic maximum coverage bandit (Chen et al., 2013). For B_1-bounded smooth CSB, online shortest path and online maximum spanning tree are typical applications (Wang & Chen, 2018). Obviously, Linear combinatorial semi-bandits is B_1-bounded smooth. We regard B_∞ and B_1 as constants in the whole paper. Apparently, B_1-bounded smoothness is a weaker assumption compared with B_∞-bounded smoothness, and we have the following fact:

Fact 1. *Suppose a reward function is B_∞-bounded smooth, then it is also B_1-bounded smooth with $B_1 = B_\infty$. On the contrary, suppose a reward function is B_1-bounded smooth, then it is B_∞-bounded smooth with $B_\infty = K B_1$.*

For many combinatorial problems such as MAX-CUT, Minimum Weighted Set Cover etc, there are only efficient approximation algorithms. Therefore, it is natural to model them as a general approximation oracle defined as below:

Definition 1. *For some $\alpha, \beta \leqslant 1$, (α, β)-approximation oracle is an oracle that takes an expectation vector $\boldsymbol{\mu}$ as input, and outputs a super arm $S \in \mathcal{S}$, such that $\Pr[r_{\boldsymbol{\mu}}(S) \geqslant \alpha \cdot \mathrm{opt}_{\boldsymbol{\mu}}] \geqslant \beta$. Here α is the approximation ratio and β is the success probability of the oracle.*

With approximation oracle, we should then consider corresponding approximation regret as we can only solve offline problem approximately:

Definition 2. *(α, β)-approximation regret of a CMAB algorithm $\mathcal{A}$ after T rounds using an (α, β)-approximation oracle under the expectation vector $\boldsymbol{\mu}$ is defined as* $\mathrm{Reg}_{\boldsymbol{\mu}, \alpha, \beta}(T) := T \cdot \alpha\beta \cdot \mathrm{opt}_{\boldsymbol{\mu}} - \mathbb{E}\left[\sum_{t=1}^{T} r_{\boldsymbol{\mu}}(S_t)\right]$.

2.2. (Local) Differential Privacy

Now we give definitions of DP and LDP, as well as a basic building block.

Definition 3 (Differential Privacy (Dwork et al., 2006; Jain et al., 2012)). *Let $D = \langle x_1, x_2, \ldots, x_T \rangle$ be a sequence of data with domain $\mathcal{X}^T$. Let $\mathcal{A}(D) = Y$, where $Y = \langle y_1, y_2, \ldots, y_T \rangle \in \mathcal{Y}^T$ be T outputs of the randomized algorithm $\mathcal{A}$ on input D. $\mathcal{A}$ is said to preserve ε-differential privacy, if for any two data sequences D, D' that differ in at most one entry, and for any subset $U \subset \mathcal{Y}^T$, it holds that*

$$\Pr(\mathcal{A}(D) \in U) \le e^{\varepsilon} \cdot \Pr(\mathcal{A}(D') \in U).$$

Compared with DP, Local Differential Privacy (LDP) is a stronger notion of privacy than DP, see Kasiviswanathan et al. (2011); Duchi et al. (2013). Since LDP requires to encrypt each user's data to protect privacy before collection, there is no need to define corresponding streaming version. Here we adopt the LDP definition given in (Bassily & Smith, 2015).

Definition 4 (LDP). *A mechanism $\mathcal{A} : \mathcal{X} \to \mathcal{Y}$ is said to be ε-local differential private or ε-LDP, if for any $x, x' \in \mathcal{X}$, and any (measurable) subset $U \subset \mathcal{Y}$, there is*

$$\Pr(\mathcal{A}(x) \in U) \le e^{\varepsilon} \cdot \Pr(\mathcal{A}(x') \in U).$$

To protect ε-LDP, the most commonly used method is Laplacian mechanism. Suppose the output domain $\mathcal{Y}$ of an algorithm $\mathcal{A}$ is bounded by a d-dimensional L1 ball with radius R, Laplacian mechanism just injects a d-dimensional random noise to the true output $\mathcal{A}(x)$, and each entry of noise is sampled from $\mathrm{Lap}(R/\varepsilon)$ independently [2]. It is easy to prove the Laplacian mechanism guarantees ε-LDP (Dwork & Roth, 2014).

3. B_∞-Bounded Smooth CSB with Privacy Guarantee

Since learning under LDP is much more difficult compared with DP, we mainly consider how to design an optimal algorithm for B_∞-Bounded Smooth CSB under ε-LDP guarantee. As we can see, based on our observation for locally differentially private CSB, it is then easy to obtain results for differentially private CSB.

As a warm-up, we show that a simple mechanism can achieve non-trivial regret with LDP guarantee, but the dependence on dimension K is sub-optimal. Next, we design an improved version with optimal utility bound, and the matching lower bound is proved in Subsection 3.3.

---

[2] $\mathrm{Lap}(b)$ represents The Laplace distribution centered at 0 with scale b, and its *p.d.f* is $\mathrm{Lap}(x|b) = \frac{1}{2b} \exp(-\frac{|x|}{b})$. The corresponding variance is $2b^2$.

Algorithm 1 CUCB-LDP1

1: **Input:** Privacy budgets ε, δ
2: **Initialize:** $\forall i \in [m], T_{0,i} = 0$, empirical mean $\tilde{\mu}_0(i) = 0$.
3: **for** $t = 1, 2, \ldots$ **do**
4: $\quad \forall i, \bar{\mu}_{t-1}(i) = \min\{\tilde{\mu}_{t-1}(i) + 4\sqrt{\frac{2K \ln T}{\varepsilon^2 T_{t-1,i}}}, 1\}$ [3]
5: $\quad$ Play $S_t = \mathrm{Oracle}(\bar{\mu}_{t-1})$ if $\bar{\mu}_{t-1} \ge 0$ else $\forall S \in \mathcal{S}$
6: $\quad$ User generates outcome $X_{t,i}$ for $i \in S_t$, and sends $X_{t,i} + z_{t,i}$ to the server, where $z_{t,i} \sim \mathrm{Lap}(K/\varepsilon)$
7: $\quad$ Server updates $T_{t,i} = T_{t-1,i} + 1, \tilde{\mu}_{t,i} = \frac{T_{t-1,i}\tilde{\mu}_{t-1,i}+X_{t,i}+z_{t,i}}{T_{t,i}}$, for $i \in S_t$, and keep others unchanged.
8: **end for**

3.1. A Straightforward Algorithm with Sub-Optimal Guarantee

Our private algorithm is based on previous non-private CSB algorithm, Combinatorial UCB (CUCB) (Chen et al., 2013; 2016). Though the reward function is non-linear in terms of super arm S and we only have access to some approximation oracle, which make our setting more complicated compared with previous private stochastic MAB (Mishra & Thakurta, 2015; Tossou & Dimitrakakis, 2016; Sajed & Sheffet, 2019), we show that the most straightforward method described in Algorithm 1 (denoted as CUCB-LDP1), i.e. using Laplacian mechanism with respect to each user's data before collection, is enough to guarantee LDP and corresponding regret.

The key observation is that, the mean estimation of each base arm lies at the core of CUCB algorithm, and adding a Laplacian noise with respect to each observation causes additional variance to these estimations, which can be handled by relaxed upper confidence bounds. Injecting noise to the reward is used both in Tossou & Dimitrakakis (2017) and Agarwal & Singh (2017) for differentially private adversarial MAB. The idea about relaxed UCB also appears before for differentially private stochastic MAB (Mishra & Thakurta, 2015), whereas we study more general locally differentially private CSB with non-linear reward and approximation oracle. Given the Laplacian mechanism, the privacy guarantee of Algorithm 1 is obvious:

Theorem 1. *Algorithm 1 guarantees ε-LDP.*

Before stating the regret bound, we define some necessary notations. We say a super arm S is bad if $r_{\boldsymbol{\mu}}(S) < \alpha \cdot \mathrm{opt}_{\boldsymbol{\mu}}$, and denote the set of bad super arms as $\mathcal{S}_B := \{S \in \mathcal{S} | r_{\boldsymbol{\mu}}(S) < \alpha \cdot \mathrm{opt}_{\boldsymbol{\mu}}\}$. For any base arm $i \in [m]$, define

$$\Delta_{\min}^i := \alpha \cdot \mathrm{opt}_{\boldsymbol{\mu}} - \max\{r_{\boldsymbol{\mu}}(S) | S \in \mathcal{S}_B, i \in S\}, \quad (1)$$

$$\Delta_{\max}^i := \alpha \cdot \mathrm{opt}_{\boldsymbol{\mu}} - \min\{r_{\boldsymbol{\mu}}(S) | S \in \mathcal{S}_B, i \in S\}, \quad (2)$$

---

[3] If a denominator is 0, we define corresponding constant as $+\infty$.

and $\Delta := \min_{i \in [m]} \Delta^i_{\min}$.

Now, we state the utility guarantee of Algorithm 1:

Theorem 2. *Under B_∞-bounded smoothness and monotonicity assumptions, the regret of Algorithm 1 is upper bounded by*

$$Reg_{\mu,\alpha,\beta}(T) \leqslant \mathcal{O}\left(\sum_{i \in [m], \Delta^i_{\min} > 0} \frac{K^2 B^2_\infty \ln T}{\varepsilon^2 \Delta^i_{\min}} \right). \quad (3)$$

Compared with corresponding non-private CUCB that achieves $\mathcal{O}\left(\sum_{i \in [m], \Delta^i_{\min} > 0} \frac{B^2_\infty \ln T}{\Delta^i_{\min}} \right)$ regret (Chen et al., 2013; 2016), one can see the regret bound of Algorithm 1 has an extra multiplicative factor $\frac{K^2}{\varepsilon^2}$, which is the price we pay for protecting LDP. According to our lower bound proved in Subsection 3.3, the dependence on the privacy parameter ε is optimal. However the additional term K^2 brought by privacy protection is undesirable and will hurt final performance for large K. In the next subsection, we show how to eliminate this additional K^2 factor.

3.2. An Improved Algorithm with the Best Guarantee

Compared with the previous studies that try to eliminate the side-effect of dimension brought by privacy protection under either *sparsity* or *low complexity* assumptions (Jain & Thakurta, 2014; Talwar et al., 2015; Zheng et al., 2017), in our general CSB setting, the information at each round is contained in a K-dimensional L_∞ ball, and we do not have any sparsity assumption, which makes the additional K^2 factor seem unavoidable.

Somewhat surprisingly, after a careful analysis, we find that there is some redundant information implicitly even without any sparsity assumption. In detail, in the analysis of Algorithm 1, the instant regret of choosing super arm S_t at round t is controlled by the largest mean estimation error among all base arms in S_t, which implies that we do not need to require all the observation of base arms in S_t of user t to update corresponding empirical means. Instead, we only use the observation of least pulled base arm in S_t to update its empirical mean and keep others unchanged, as it is the weakest one in S_t and causes largest estimation error. Since the user only sends the information of one entry to server now, it is enough to add noise in $\mathcal{O}(1/\varepsilon)$ order to protect it, which then gets rids of the annoying additional K^2 factor in the regret guarantee. Denote this variant as CUCB-LDP2, as shown in Algorithm 2.

Again, the privacy guarantee follows directly from the classic Laplacian mechanism:

Theorem 3. *Algorithm 2 guarantees ε-LDP.*

Since we condense the information required from each user significantly, which is reduced from K observations to one

---

Algorithm 2 CUCB-LDP2

1: **Input:** Privacy budgets ε, δ
2: **Initialize:** $\forall i \in [m], T_{0,i} = 0$, empirical mean $\tilde{\mu}_0(i) = 0$.
3: **for** $t = 1, 2, \dots$ **do**
4:   $\forall i, \bar{\mu}_{t-1}(i) = \min\{\tilde{\mu}_{t-1}(i) + 4\sqrt{2\frac{\ln T}{\varepsilon^2 T_{t-1,i}}}, 1\}$
5:   Play $S_t = \text{Oracle}(\bar{\mu}_{t-1})$ if $\bar{\mu}_{t-1} \geqslant 0$ else $\forall S \in \mathcal{S}$
6:   User generates outcome $X_{t,i}$ for $i \in S_t$, and sends $X_{t,I_t} + z_{t,I_t}$ to the server, where $I_t = \text{argmin}_{i \in S_t} T_{t-1,i}, z_{t,I_t} \sim \text{Lap}(1/\varepsilon)$
7:   Server updates $T_{t,I_t} = T_{t-1,I_t} + 1, \tilde{\mu}_{t,I_t} = \frac{T_{t-1,I_t}\tilde{\mu}_{t-1,I_t} + X_{t,I_t} + z_{t,I_t}}{T_{t,I_t}}$, and keep others unchanged.
8: **end for**

---

observation, now we can inject less noise and prove a much better regret bound compared with the guarantee of Algorithm 1:

Theorem 4. *Under B_∞-bounded smoothness and monotonicity assumptions, the regret of Algorithm 2 is upper bounded by*

$$Reg_{\mu,\alpha,\beta}(T) \leqslant \mathcal{O}\left(\sum_{i \in [m], \Delta^i_{\min} > 0} \frac{B^2_\infty \ln T}{\varepsilon^2 \Delta^i_{\min}} \right) \quad (4)$$

Compared with the non-private theoretical guarantee, theorem 4 implies that we can achieve optimal locally differentially private B_∞-bounded smooth CSB without any additional price paid for privacy protection, which is a bit surprising given the previous work about (locally) differentially private learning. See section A in the supplementary materials for the proof of theorem 4.

Multi-Armed Bandits (MAB) is a special case of CSB, where $\mathcal{S} = \{e_i | i \in [m]\}$ and $K = 1$. In this case, our Algorithms 1 and 2) are exactly the same, and we obtain an algorithm for MAB under ε-LDP with regret bound $\mathcal{O}(\sum_{i \neq i^*} \frac{\ln T}{\Delta_i \varepsilon^2})$, where i^* is the optimal base arm, and Δ_i is the gap between arm i and optimal arm i^*. Apparently, this regret bound is also optimal given the LDP lower bound $\Omega(\sum_{i \neq i^*} \frac{\ln T}{\varepsilon^2 \Delta_i})$ proved in Basu et al. (2019) and non-private lower bound $\Omega(\sum_{i \neq i^*} \frac{\ln T}{\Delta_i})$ (Bubeck et al., 2012).

Finally, if one wants to protect ε-DP rather than ε-LDP, based on the same observation as above, we can simply use the tree-based aggregation technique (Dwork et al., 2010) with respect to the least pulled base arm to calculate its empirical mean estimation with DP guarantee. Since the tree-based aggregation technique injects much less noise compared with Algorithm 2 designed for LDP, it is not hard to prove that this variant for DP can achieve regret bound

$\tilde{\mathcal{O}}(\frac{mB_\infty^2 \ln T}{\varepsilon\Delta})$.[4]

3.3. Lower Bounds

In this subsection, we prove the regret lower bound for locally private CSB problem with B_∞-bounded smoothness. Like previous work (Kveton et al., 2015; Wang & Chen, 2017), we only consider lower bound with exact oracle, i.e. $\alpha = \beta = 1$.

First we define a class of algorithms that we are interested in:

Definition 5. *An algorithm is called consistent if for any suboptimal super arm S, the number of times S is chosen by the algorithm is subpolynomial in T for any stochastic CSB instance, i.e. $\mathbb{E}\left[N_S(T)\right] \leq o(T^p)$ for any $0 < p < 1$.*

Our lower bound is derived for the consistent algorithm class, which is natural for the stochastic CSB and has been used for lower bound analysis in many previous results (Lattimore & Szepesvári, 2018; Basu et al., 2019; Lai & Robbins, 1985; Kveton et al., 2015).

Our analysis focuses on CSB instances where the suboptimality gap Δ of any super arms are equal. Since general CSB problem is harder than CSB problem with equal suboptimality gap (The latter problem can be reduced to the former), our lower bound can be directly applied to general CSB class, with Δ replaced with $\Delta_{\min}^i$ for each base arm i.

Theorem 5. *For any m and K, and any Δ satisfying $0 < \Delta/B_\infty < 0.35$, the regret of any consistent ε-locally private algorithm π on the CSB problem with B_∞-bounded smoothness is bounded from below as*

$$\liminf_{T\to\infty} \frac{Reg(T)}{\log T} \geq \frac{B_\infty^2(m-1)}{64(e^\varepsilon-1)^2\Delta}$$

Specifically, for $0 < \varepsilon \leq 1/2$, the regret is at least

$$\liminf_{T\to\infty} \frac{Reg(T)}{\log T} \geq \frac{B_\infty^2(m-1)}{128\varepsilon^2\Delta}$$

The lower bound shows that Algorithm 2 achieves optimal regret with respect to all the parameters of the CSB instance. The proof of the theorem is an almost direct reduction from private MAB. Previous result (Theorem 2 in Basu et al. (2019)) shows that the regret for any consistent ε-locally private algorithm for MAB is at least $\Omega\left(\frac{m\ln T}{\varepsilon^2\Delta}\right)$. Since any MAB instance is a special case of CSB with $B_\infty = 1$, the regret lower bounds for stochastic CSB with $B_\infty = 1$ follows directly by reduction. For general CSB problem with B_∞-bounded smoothness, we consider a similar instance

---

[4]The proof for this result is actually a combination of techniques used in this subsection and what we will use in subsection 4.2, hence omitted.

with the reward of each arm in MAB instance multiplied by B_∞. See Section B in the supplementary materials for the detailed analysis. For B_∞-bounded smooth CSB under DP setting, using nearly the same technique, it is not hard to prove that the corresponding lower bound is $\Omega(\frac{mB_\infty^2 \ln T}{\varepsilon\Delta})$.

4. B_1-Bounded Smooth CSB with Privacy Guarantee

4.1. B_1-Bounded Smooth CSB under LDP

Though our proposed Algorithm 2 is already optimal for B_∞-bounded smooth CSB, if we use it for B_1-bounded smooth CSB such as important linear CSB to protect ε-LDP, we will obtain its regret bound in order $\mathcal{O}\left(\sum_{i\in[m],\Delta_{\min}^i>0} \frac{K^2 B_1^2 \ln T}{\varepsilon^2\Delta_{\min}^i}\right)$ due to Fact 1. However, the optimal non-private regret bound for B_1-bounded smooth CSB is $\Theta\left(\sum_{i\in[m],\Delta_{\min}^i>0} \frac{KB_1^2 \ln T}{\Delta_{\min}^i}\cdot\right)$ (Kveton et al., 2015; Wang & Chen, 2017), which implies a gap with our locally differentially private upper bound. Is it possible to eliminate this additional K just like in the previous locally differentially private B_∞-bounded smooth CSB? First we prove a lower bound for B_1-Bounded Smooth CSB under LDP guarantee. Our result under B_1-bounded smoothness assumption can be applied to linear CSB problem by setting $B_1 = 1$.

Theorem 6. *For any m and K such that m/K is an integer, and any Δ satisfying $0 < \Delta/(B_1K) < 0.35$, the regret of any consistent ε-locally private algorithm π on the CSB problem satisfying B_1-bounded smoothness is bounded from below as*

$$\liminf_{T\to\infty} \frac{Reg(T)}{\log T} \geq \frac{B_1^2(m-K)K}{64(e^\varepsilon-1)^2\Delta}$$

Specifically, for $0 < \varepsilon \leq 1/2$, the regret is at least

$$\liminf_{T\to\infty} \frac{Reg(T)}{\log T} \geq \frac{B_1^2(m-K)K}{128\varepsilon^2\Delta}$$

We borrow the hard instance from Kveton et al. (2015) to prove the lower bound. Consider a K-path semi-bandit problem with m base arms. The feasible super arms are m/K paths, each containing base arm $(i-1)K+1, (i-1)K+2, ..., iK$ for $i \in \{1, ..., m/K\}$. The reward of pulling super arm S is B_1 times the sum of the weight $\tilde{w}_i$ for $i \in S$. The weights $\tilde{w}_i$ of the different base arms in the same super arm are identical, while the weights in the different paths are i.i.d sampled. Denote the best super arm as S^*, The weight of each base arm is a Bernoulli random variable with mean:

$$\bar{w}(i) = \begin{cases} 0.5 & i \in S^* \\ 0.5 - \Delta/(B_1K) & \text{otherwise} \end{cases}$$

We use the general canonical bandit model (Lattimore & Szepesvári, 2018) to prove above theorem. See Section C in the supplementary materials for the detailed proof.

Though we can only prove a lower bound of $\Omega(\frac{mKB_1^2 \ln T}{\varepsilon^2 \Delta})$ in the same order as corresponding non-private optimal guarantee, we conjecture our lower bound is loose and the right lower bound is $\Omega(\frac{mK^2 B_1^2 \ln T}{\varepsilon^2 \Delta})$. In other words, maybe there is indeed some side-effect for utility guarantee about the dimension K if we hope to protect LDP. Intuitively, for B_1 bounded smooth CSB, we may have to update all arms in a played super arm for the regret guarantee (instead of only one arm as we did for B_∞ bounded smooth CSB), and this makes the privacy protection harder with an extra factor of K.

Since *differential privacy* is a relatively weaker notion compared with LDP, there may be some hope to further improve the regret bound if we focus on the guarantee of DP. In next two subsections, we show it is indeed true, by designing an ε-differentially private algorithm with regret bound $\tilde{\mathcal{O}}\left(\sum_{i \in [m], \Delta_{\min}^i > 0} \frac{KB_1^2 \ln^2 T}{\Delta_{\min}^i} + \frac{mKB_1 \ln^3 T}{\varepsilon}.\right)$, and proving a nearly matching lower bound.

4.2. Upper Bound under DP

Compared with LDP, in which case the learning algorithm (or the server) can only receives noisy information, DP only has some restriction for the output of an algorithm, and the server has authority to collect true data. Thus, it is possible to inject much less noise under DP setting via an economic allocation of privacy budget ε.

We use tree-based aggregation scheme (Dwork et al., 2009; Chan et al., 2011) to protect ε-DP in our algorithm, which is an effective method in releasing private continual statistics over a data stream and frequently used in previous work, such as stochastic MAB (Mishra & Thakurta, 2015; Tossou & Dimitrakakis, 2016), Online Convex Optimization (Thakurta & Smith, 2013). Consider a data stream $(X_1, X_2, ..., X_T)$ where $X_i \in [0, 1]$. In each step t, the algorithm receives data X_t, and needs to output the sum $\bar{X}_t = \sum_{i=1}^{t} X_i$, while insuring that the output sequence $(\bar{X}_1, \bar{X}_2, ..., \bar{X}_T)$ are ε-differentially private. Tree-based mechanism solves this problem in an elegant way with a binary tree. Each leaf node denotes data X_t received in step t. Each internal node calculates the sum of data in the leaf nodes rooted at it. Notice that one only needs access to $\lceil \log t \rceil$ nodes and sums up the values on them in order to calculate $\bar{X}_t$. Using the Laplacian mechanism, previous results have shown that adding i.i.d $\mathrm{Lap}(\|X\|_1 \log T / \varepsilon)$ to each node ensures ε-differential privacy for the scheme as stated in the following lemma:

Lemma 1 (Dwork et al. (2010); Chan et al. (2011)). *Tree-based aggregation scheme with i.i.d* $\mathrm{Lap}(\|X\|_1 \log T / \varepsilon)$

Algorithm 3 CUCB-DP

1: **Input:** Privacy budgets ε, δ.
2: **Initialize:** $\forall i \in [m], T_{0,i} = 0$, empirical mean $\tilde{\mu}_0(i) = 0$.
3: **for** $t = 1$ **to** T **do**
4: $\quad \forall i, \bar{\mu}_{t-1}(i) = \min\{\tilde{\mu}_{t-1}(i) + \sqrt{\frac{4 \ln(mT)}{T_{t-1,i}}} + \frac{12K \ln^3 T}{T_{t-1,i}\varepsilon}, 1\}$
5: $\quad$ Play $S_t = \mathrm{Oracle}(\bar{\mu}_{t-1})$ if $\bar{\mu}_{t-1} \geqslant 0$ else $\forall S \in \mathcal{S}$
6: $\quad$ User generates outcome $X_{t,i}$ for $i \in S_t$, and sends $X_{t,i}$ to the server
7: $\quad$ Server updates base arms in S_t: $\tilde{\mu}_{t,i} = \frac{TreeBasedAggregation(\{X_{\tau,i} | \tau \in [t], i \in S_\tau\})}{T_{t,i}}$, $T_{t,i} = T_{t-1,i} + 1$, and keeps others unchanged
8: **end for**

---

noise added to each node is ε-differentially private.

In our CSB setting, we store a vector $\boldsymbol{X_t}$ with support at most K in the leaf nodes of step t. Each internal node calculates the sum of $\boldsymbol{X_t}$ in the leaf nodes rooted at it. For each node, we add i.i.d $\mathrm{Lap}(2K \log T / \varepsilon)$ noise to each dimension of the vector stored on the node to guarantee ε-DP (See Algorithm 3). Based on Lemma 1, we have

Theorem 7. *Algorithm 3 guarantees ε-DP.*

In Algorithm 3, when we need to estimate the mean weight μ_i based on the previous outcome $X_{t,i}$, we add additional Laplace noise to the sum of $X_{t,i}$ due to tree-based aggregation scheme. Note that the number of Laplace noises added (the number of nodes we access to) is only logarithmic. This means that the additional confidence bound due to Laplace noise is only $\tilde{\Theta}(1/T_{t-1,i})$ for base arm i when it is pulled for $T_{t-1,i}$ times. Compared with the original bound for the sub-Gaussian noise which is of order $\tilde{\Theta}(\sqrt{1/T_{t-1,i}})$, the additional bound for Laplace noise enjoys better dependence on $T_{t-1,i}$. This helps us to separate the term of Δ and ε in the regret via delicate analysis, and finally derive a nearly optimal bound in the additive form.

Theorem 8. *Under B_1-bounded smoothness and monotonicity assumptions, the regret of Algorithm 3 is upper bounded by*

$$Reg_{\mu,\alpha,\beta}(T) \leqslant \mathcal{O}\left(\sum_{i \in [m], \Delta_{\min}^i > 0} \frac{KB_1^2 \ln^2 T}{\Delta_{\min}^i}\right) + \mathcal{O}\left(\frac{mKB_1 \ln^3 T \ln\left(\frac{B_1 K \ln T}{\Delta_{\max}\varepsilon}\right)}{\varepsilon}\right).$$

Note when privacy parameter ε is regarded as a constant which is common in real applications, the second term in the right hand side of above inequality is nearly dominated

by the first term, which is almost the optimal regret bound in non-private B_1-bounded smooth CSB. Thus by relaxing LDP to DP, we have shown that it is possible to eliminate the side-effect on dimension induced by privacy protection and nearly match corresponding non-private optimal bound $\mathcal{O}(\sum_{i\in[m],\Delta_{\min}^i>0} \frac{KB_1^2\ln T}{\Delta_{\min}^i})$.

Before proving Theorem 8, we present the following lemma. This lemma gives an upper bounds on the sub-optimal gap in round t, which helps to treat the Δ term and ε term separately. We refer readers to Section D of the supplementary materials for the proof of Lemma 2.

Lemma 2. *Suppose $\Delta_{S_t} = \alpha r_\mu(S_\mu^*) - r_\mu(S_t)$. Denote*

$$F_t = \left\{\Delta_{S_t} \leq B_1 \sum_{i\in S_t}\left(4\sqrt{\frac{\ln T}{T_{t-1,i}}} + \frac{24K\ln^3 T}{T_{t-1,i}\varepsilon}\right)\right\}.$$

Then the regret for Algorithm 3 is bounded by

$$Reg_{\mu,\alpha,\beta}(T) \leq \sum_{t\in[T]}\Delta_{S_t}\mathbf{1}\{F_t\} + 3\sum_{i\in[m]}\Delta_{\max}^i \quad (5)$$

Now we are ready to prove Theorem 8.

Proof. (proof of Theorem 8) We mainly analyze the first term of the RHS in Inq. 5. Define $\hat{R}_T = \sum_{t\in[T]}\Delta_{S_t}\mathbf{1}\{F_t\}$. In step t, we consider the case that F_t happens. Define $\bar{\Delta}_{S_t} = \max_{i\in S_t}\Delta_{\min}^i$. Since $\Delta_{\min}^i \leq \Delta_{S_t}, \forall i \in S_t$, we have $\bar{\Delta}_{S_t} \leq \Delta_{S_t}$. Then we have

$$\Delta_{S_t} + \bar{\Delta}_{S_t} \leq 2\Delta_{S_t}$$

$$\leq 2B_1\sum_{i\in S_t}\left(4\sqrt{\frac{\ln T}{T_{t-1,i}}} + \frac{24K\ln^3 T}{T_{t-1,i}\varepsilon}\right)$$

that is,

$$\Delta_{S_t} \leq 2\Delta_{S_t} - \bar{\Delta}_{S_t}$$

$$\leq 2B_1\sum_{i\in S_t}\left(4\sqrt{\frac{\ln T}{T_{t-1,i}}} + \frac{24K\ln^3 T}{T_{t-1,i}\varepsilon} - \frac{\bar{\Delta}_{S_t}}{2B_1 K}\right)$$

$$\leq 2B_1\sum_{i\in S_t}\left(4\sqrt{\frac{\ln T}{T_{t-1,i}}} + \frac{24K\ln^3 T}{T_{t-1,i}\varepsilon} - \frac{\Delta_{\min}^i}{2B_1 K}\right)$$

Let $n_{\max}^i = \max\left\{\frac{256B_1^2 K^2\ln T}{(\Delta_{\min}^i)^2}, \frac{96B_1 K^2\ln^3 T}{\varepsilon\Delta_{\min}^i}\right\}$. Denote $\Delta_i(n) = 4B_1\sqrt{\frac{\ln T}{n}} + \frac{24B_1 K\ln^3 T}{n\varepsilon} - \frac{\Delta_{\min}^i}{2K}$. For base arm i, if $n \geq n_{\max}^i$, we have $\Delta_i(n) \leq 0$.

$$\hat{R}_T \leq \sum_{t\in[T]}\Delta_{S_t}\mathbf{1}\{F_t\}$$

$$\leq \sum_{t=1}^{T}\sum_{i\in S_t} 2\Delta_i(T_{i,t})\mathbf{1}\{F_t\}$$

$$\leq \sum_{i\in[m]}\sum_{n=1}^{n_{\max}^i} 2B_1\left(4\sqrt{\frac{\ln T}{n}} + \frac{24K\ln^3 T}{n\varepsilon}\right)$$

$$\leq \sum_{i\in[m]}\int_0^{n_{\max}^i} 8B_1\sqrt{\frac{\ln T}{n}}dn$$

$$+ \sum_{i\in[m]}\left(\frac{48B_1 K\ln^3 T}{\varepsilon} + \int_1^{n_{\max}^i}\frac{48B_1 K\ln^3 T}{n\varepsilon}dn\right)$$

$$\leq \sum_{i\in[m]} 16B_1\sqrt{\ln T\left(\frac{256B_1^2 K^2\ln T}{(\Delta_{\min}^i)^2} + \frac{96B_1 K^2\ln^3 T}{\varepsilon\Delta_{\min}^i}\right)} +$$

$$\sum_{i\in[m]}\left(\frac{48B_1 K\ln^3 T}{\varepsilon}\left(1 + \ln\left(\frac{256K^2 B_1^2\ln^4 T}{\varepsilon(\Delta_{\min}^i)^2}\right)\right)\right)$$

After simplifying the equation using basic inequalities such as $\sqrt{ab} \geq \frac{2}{1/a+1/b}$ and $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ $(a, b \geq 0)$, we can show that

$$\hat{R}_T \leq$$

$$\mathcal{O}\left(B_1^2 K\ln^2 T\sum_{i\in[m]}\frac{1}{\Delta_{\min}^i} + \frac{mB_1 K\ln^3 T\ln\left(\frac{B_1 K\ln T}{\Delta_{\max}\varepsilon}\right)}{\varepsilon}\right)$$

$\square$

4.3. Lower Bound under DP

In this subsection, we prove the lower bound for CSB algorithm under ε-DP. Similar with the result of LDP lower bound, we consider CSB algorithm with consistent property. The lower bound stated below implies that our algorithm 3 can achieve near-optimal regret regardless of logarithmic factors:

Theorem 9. *For any m and K such that $m \geq 2K$, and any Δ satisfying $0 < \Delta/(B_1 K) < 0.35$, the regret for any consistent CSB algorithm guaranteeing ε-DP is at least $\Omega\left(\frac{B_1^2 mK\ln T}{\Delta} + \frac{B_1 mK\ln T}{\varepsilon}\right)$.*

The theorem is proved in section E of the supplementary materials. We only sketch the proof here. Previous results have shown that for non-private stochastic linear CSB, the regret lower bound is at least $\Omega(\frac{mK\ln T}{\Delta})$. By slightly

modifying the hard instance, we can show that the regret lower bound for non-private CSB with B_1-bounded smoothness is $\Omega(\frac{B_1^2 mK \ln T}{\Delta})$. Since private CSB is strictly harder than non-private CSB (by reduction), the regret lower bound for private CSB is $\Omega(\frac{B_1 mK \ln T}{\Delta})$. We only need to prove that the regret lower bound for private CSB is $\Omega\left(\frac{B_1 mK \ln T}{\varepsilon}\right)$, from which we can prove that the regret lower bound is $\Omega\left(\max\left\{\frac{B_1^2 mK \ln T}{\Delta}, \frac{B_1 mK \ln T}{\varepsilon}\right\}\right) = \Omega\left(\frac{B_1^2 mK \ln T}{\Delta} + \frac{B_1 mK \ln T}{\varepsilon}\right)$.

Now we sketch the proof of $\Omega\left(\frac{B_1 mK \ln T}{\varepsilon}\right)$ term. Note a simple extension of Kveton et al. (2015) can only achieve $\Omega\left(\frac{B_1 m \ln T}{\varepsilon}\right)$ in our differentially private setting, which is not satisfactory. It is thus necessary to construct some new hard instance to prove Theorem 9.

To solve this problem, we design the following CSB problem as a special case of general CSB with B_1-bounded smoothness. Suppose there are m base arms, each associated with a weight sampled from Bernoulli distribution. These m base arms are divided into three sets, S^*, $\tilde{S}$ and $\bar{S}$. S^* contains m base arms, which build up the optimal super arm set. $\tilde{S}$ contains $K - 1$ "public" base arms for sub-optimal super arms. These arms are contained in all sub-optimal super arms. $\bar{S}$ contains $m - 2K + 1$ base arms. each base arm combined with $K - 1$ "public" base arms in $\tilde{S}$ builds up a sub-optimal super arm. Totally we have $m - 2K + 1$ sub-optimal super arms and one optimal super arm. The mean of the Bernoulli random variable associated to each base arm is defined as follow:

$$w(i) = \begin{cases} 0.5 & i \in S^* \\ 0.5 - \Delta/(B_1 K) & \text{otherwise} \end{cases}$$

The weights of base arms in $\tilde{S}$ are identical, while other weights are i.i.d sampled. The reward of pulling a super arm S is B_1 times the sum of weights of all base arm $i \in S$. As a result, the sub-optimality gap of each sub-optimal super arm is Δ. With the coupling argument in Karwa & Vadhan (2017), we can prove that $\mathbb{E}(N_S)$ is at least $\Omega(\frac{mKB_1 \ln T}{\varepsilon \Delta})$ for any sub-optimal super arm S with high probability. Since there are $\theta(m)$ sub-optimal super arm, we can reach the conclusion that the regret lower bound for private CSB is $\Omega\left(\frac{B_1 mK \ln T}{\varepsilon}\right)$.

5. Conclusion and Future work

In this paper, we study (locally) differentially private algorithm for Combinatorial Semi-Bandits under two common assumptions about reward functions. For B_∞-bounded smooth CSB under ε-LDP and ε-DP, we show the optimal

regret of these two settings are respectively $\Theta(\frac{mB_\infty^2 \ln T}{\varepsilon^2 \Delta})$ and $\tilde{\Theta}(\frac{mB_\infty^2 \ln T}{\varepsilon \Delta})$, by proving lower bounds and designing (nearly) optimal private algorithms. For relatively weaker B_1-bounded smooth CSB, if we are required to protect ε-DP instead of ε-LDP, we show the optimal regret is $\tilde{\Theta}(\frac{mKB_1^2 \ln T}{\varepsilon \Delta})$, and give a differentially private algorithm as well as a nearly matching lower bound. Moreover, above optimal performance in our (locally) differentially private CSB is nearly the same order as non-private setting (Kveton et al., 2015; Chen et al., 2016; Wang & Chen, 2017).

Our Algorithm 2 is applicable for locally private CSB with B_1-bounded smoothness, with a regret upper bound of $\mathcal{O}(\frac{mK^2 B_1^2 \ln T}{\Delta \varepsilon^2})$ in this setting. However, the regret lower bound we prove is just $\Omega(\frac{mKB_1^2 \ln T}{\Delta \varepsilon^2})$. We conjecture that our lower bound is loose and the Algorithm 2 is also near-optimal for locally private CSB with B_1-bounded smoothness. How to improve the lower bound is left as future work.

Recently, there are interesting results under Gini-weighted smoothness assumptions (Merlis & Mannor, 2019; 2020). Compared with general Lipschitz smoothness considered in this work, this is a more refined smoothness assumption, which leads to near optimal regret bounds with less dependence on the dimension K. Directly applying our algorithms to this setting will lead to an additional dependence on K. How to remove this additional price for privacy preserving, and how to prove the corresponding lower bounds, are interesting problems for future work.

Acknowledgements

References

Agarwal, N. and Singh, K. The price of differential privacy for online learning. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 32–40. JMLR. org, 2017.

Auer, P., Cesa-Bianchi, N., and Fischer, P. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

Bassily, R. and Smith, A. Local, private, efficient protocols for succinct histograms. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, pp. 127–135. ACM, 2015.

Bassily, R., Smith, A., and Thakurta, A. Private empirical risk minimization: Efficient algorithms and tight error bounds. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pp. 464–473. IEEE, 2014.

Basu, D., Dimitrakakis, C., and Tossou, A. Differential privacy for multi-armed bandits: What is it and what is its cost? *arXiv preprint arXiv:1905.12298*, 2019.

Bubeck, S., Cesa-Bianchi, N., et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5 (1):1–122, 2012.

Chan, T.-H. H., Shi, E., and Song, D. Private and continual release of statistics. *ACM Transactions on Information and System Security (TISSEC)*, 14(3):1–24, 2011.

Chen, W., Wang, Y., and Yuan, Y. Combinatorial multi-armed bandit: General framework and applications. In *International Conference on Machine Learning*, pp. 151–159, 2013.

Chen, W., Wang, Y., Yuan, Y., and Wang, Q. Combinatorial multi-armed bandit and its extension to probabilistically triggered arms. *The Journal of Machine Learning Research*, 17(1):1746–1778, 2016.

Combes, R., Shahi, M. S. T. M., Proutiere, A., et al. Combinatorial bandits revisited. In *Advances in Neural Information Processing Systems*, pp. 2116–2124, 2015.

Cormode, G., Jha, S., Kulkarni, T., Li, N., Srivastava, D., and Wang, T. Privacy at scale: Local differential privacy in practice. In *Proceedings of the 2018 International Conference on Management of Data*, pp. 1655–1658, 2018.

Duchi, J., Wainwright, M. J., and Jordan, M. I. Local privacy and minimax bounds: Sharp rates for probability estimation. In *Advances in Neural Information Processing Systems*, pp. 1529–1537, 2013.

Duchi, J., Wainwright, M., and Jordan, M. Minimax optimal procedures for locally private estimation. *arXiv preprint arXiv:1604.02390*, 2016.

Dwork, C. and Roth, A. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407, 2014.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In *Theory of cryptography*, pp. 265–284, Berlin, Germany, March 2006. Springer.

Dwork, C., Naor, M., Reingold, O., Rothblum, G. N., and Vadhan, S. On the complexity of differentially private data release: efficient algorithms and hardness results. In *Proceedings of the forty-first annual ACM symposium on Theory of computing*, pp. 381–390, 2009.

Dwork, C., Naor, M., Pitassi, T., and Rothblum, G. N. Differential privacy under continual observation. In *Proceedings of the forty-second ACM symposium on Theory of computing*, pp. 715–724. ACM, 2010.

Even-Dar, E., Mannor, S., and Mansour, Y. Pac bounds for multi-armed bandit and markov decision processes. In *International Conference on Computational Learning Theory*, pp. 255–270. Springer, 2002.

Gajane, P., Urvoy, T., and Kaufmann, E. Corrupt bandits for preserving local privacy. In *Algorithmic Learning Theory*, pp. 387–412, 2018.

György, A., Linder, T., Lugosi, G., and Ottucsák, G. The online shortest path problem under partial monitoring. *Journal of Machine Learning Research*, 8(Oct):2369–2403, 2007.

Jain, P. and Thakurta, A. G. (near) dimension independent risk bounds for differentially private learning. In *International Conference on Machine Learning*, pp. 476–484, 2014.

Jain, P., Kothari, P., and Thakurta, A. Differentially private online learning. In *Conference on Learning Theory*, pp. 24–1, 2012.

Karwa, V. and Vadhan, S. Finite sample differentially private confidence intervals. *arXiv preprint arXiv:1711.03908*, 2017.

Kasiviswanathan, S. P., Lee, H. K., Nissim, K., Raskhodnikova, S., and Smith, A. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011.

Kifer, D., Smith, A., and Thakurta, A. Private convex empirical risk minimization and high-dimensional regression. *Journal of Machine Learning Research*, 1(41):3–1, 2012.

Kveton, B., Wen, Z., Ashkan, A., and Szepesvari, C. Tight regret bounds for stochastic combinatorial semi-bandits. In *Artificial Intelligence and Statistics*, pp. 535–543, 2015.

Lai, T. L. and Robbins, H. Asymptotically efficient adaptive allocation rules. *Advances in applied mathematics*, 6(1): 4–22, 1985.

Lattimore, T. and Szepesvári, C. Bandit algorithms. *preprint*, 2018.

Lattimore, T. and Szepesvári, C. An information-theoretic approach to minimax regret in partial monitoring. *arXiv preprint arXiv:1902.00470*, 2019.

Merlis, N. and Mannor, S. Batch-size independent regret bounds for the combinatorial multi-armed bandit problem. *arXiv preprint arXiv:1905.03125*, 2019.

Merlis, N. and Mannor, S. Tight lower bounds for combinatorial multi-armed bandits. *arXiv preprint arXiv:2002.05392*, 2020.

Mishra, N. and Thakurta, A. (nearly) optimal differentially private stochastic multi-arm bandits. In *Proceedings of the Thirty-First Conference on Uncertainty in Artificial Intelligence*, pp. 592–601. AUAI Press, 2015.

Sajed, T. and Sheffet, O. An optimal private stochastic-mab algorithm based on optimal private stopping rule. In *International Conference on Machine Learning*, pp. 5579–5588, 2019.

Shariff, R. and Sheffet, O. Differentially private contextual linear bandits. In *Advances in Neural Information Processing Systems*, pp. 4296–4306, 2018.

Smith, A. and Thakurta, A. Differentially private model selection via stability arguments and the robustness of the lasso. *J Mach Learn Res Proc Track*, 30:819–850, 2013.

Talwar, K., Thakurta, A., and Zhang, L. Nearly optimal private lasso. In *Advances in Neural Information Processing Systems*, pp. 3025–3033, 2015.

Thakurta, A. G. and Smith, A. (nearly) optimal algorithms for private online learning in full-information and bandit settings. In *Advances in Neural Information Processing Systems*, pp. 2733–2741, 2013.

Tossou, A. C. and Dimitrakakis, C. Algorithms for differentially private multi-armed bandits. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

Tossou, A. C. Y. and Dimitrakakis, C. Achieving privacy in the adversarial multi-armed bandit. In *Thirty-First AAAI Conference on Artificial Intelligence*, 2017.

Wang, Q. and Chen, W. Improving regret bounds for combinatorial semi-bandits with probabilistically triggered arms and its applications. In *Advances in Neural Information Processing Systems*, pp. 1161–1171, 2017.

Wang, S. and Chen, W. Thompson sampling for combinatorial semi-bandits. In *International Conference on Machine Learning*, pp. 5101–5109, 2018.

Zheng, K., Mou, W., and Wang, L. Collect at once, use effectively: Making non-interactive locally private learning possible. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pp. 4130–4139. JMLR. org, 2017.

Appendices

A. Proof of Theorem 4

Theorem 4. *(Restate) For Algorithm 2, we have*

$$Reg_{\mu,\alpha,\beta}(T) \leqslant \mathcal{O}\left(\sum_{i\in[m],\Delta^i_{\min}>0} \frac{B^2_\infty \ln T}{\varepsilon^2 \Delta^i_{\min}}\right) \tag{6}$$

Proof. Suppose G_t denote the event that the oracle fails to produce an α-approximate answer with respect to the input vector in step t. We have $\mathbb{P}[G_t] \leq 1 - \beta$. The number of times G_t happens in expectation is at most $(1-\beta)T$. The cumulative regret in these steps is at most $R_{\text{fail}} \leq (1-\beta)T\Delta_{\max}$

Now we only consider the steps G_t doesn't happen. We maintain counters N_i in the proof, and denote its value in step t as $N_{t,i}$. The initialization of $N_{t,i}$ is the same as $T_{t,i}$, i.e. $N_{0,i} = 0$. In step t, if G_t doesn't happen, and the oracle selects a sub-optimal super arm, we increment N_{I_t} by one, i.e. $N_{t,I_t} = N_{t-1,I_t} + 1$, where $I_t = \arg\min_{i\in S_t} T_{t-1,i}$, otherwise we keep N_i unchanged. This indicates that $N_{t,i} \leq T_{t,i}$. Notice that if a sub-optimal super arm S_t is pulled in step t, exactly one counter N_{I_t} is incremented by one, and $I_t \in S_t$. As a result, we have:

$$
\begin{aligned}
Reg_{\mu,\alpha,\beta}(T) &\leq T\alpha\beta \operatorname{opt}_\mu - \mathbb{E}\sum_{t=1}^T r_\mu(S_t) \\
&\leq R_{\text{fail}} + T\alpha\beta \operatorname{opt}_\mu - \left(T\alpha \operatorname{opt}_\mu - \sum_{i\in[m],\Delta^i_{\min}>0}\sum_{j=1}^{N_{T,i}} \Delta_{i,j}\right) \\
&\leq \sum_{i\in[m],\Delta^i_{\min}>0}\sum_{j=1}^{N_{T,i}} \Delta_{i,j}
\end{aligned}
\tag{7}
$$

Here $\Delta_{i,j}$ denote the suboptimal gap $\alpha \cdot \operatorname{opt}_\mu - r(S_t)$ when N_i incremented from $j-1$ to j in a certain step t.

Now we only need to bound $N_{T,i}$ and $\Delta_{i,j}$. We denote the following event as $\Lambda_{t,i}$: For a fixed step $t \in T$ and a fixed base arm $i \in [m]$,

$$|\tilde{\mu}_t(i) - \mu_i| \leq 4\sqrt{\frac{2\ln T}{\varepsilon^2 T_{t,i}}}.$$

The noise in $\tilde{\mu}_t(i)$ comes from two parts: the Laplacian noise added for privacy and the randomness of $X_{t,i}$. For the first part, by Bernstein's Inequality over $T_{t,i}$ i.i.d Laplace distribution, the confidence bound is $2\sqrt{\frac{2\ln T}{\varepsilon^2 T_{t,i}}}$ with prob. at least $1 - 2/T^2$. For the second part, since $X_{t,i}$ is $[0,1]$ bounded, the confidence bound is $2\sqrt{\frac{2\ln T}{T_{t,i}}} \leq 2\sqrt{\frac{2\ln T}{\varepsilon^2 T_{t,i}}}$ with prob. at least $1 - 2/T^2$ by Hoeffding's inequality. This shows that $\Lambda_{t,i}$ happens with prob. $1 - 4/T^2$. By union bounds over all steps, $\Lambda_{t,i}$ happens for all t and i with prob. $1 - 4/T$. We denote this event as Λ.

Suppose Λ happens, we have $\mu(i) \leq \bar{\mu}_t(i) \leq \mu(i) + 4\sqrt{\frac{2\ln T}{\varepsilon^2 T_{t,i}}}$. If a sub-optimal arm S_t is pulled in step t. we have

$$
\begin{aligned}
\alpha r_\mu(S^*_\mu) - r_\mu(S_t) &\leq \alpha r_{\bar{\mu}_t}(S^*_\mu) - (r_{\bar{\mu}_t}(S_t) - B_\infty\|\bar{\mu}_t - \mu\|_\infty) \\
&\leq B_\infty\|\bar{\mu}_t - \mu\|_\infty \\
&\leq B_\infty(\|\bar{\mu}_t - \tilde{\mu}_t\|_\infty + \|\tilde{\mu}_t - \mu\|_\infty) \\
&\leq B_\infty 8 \max_{i\in S_t}\left\{\sqrt{\frac{2\ln T}{\varepsilon^2 T_{t-1,i}}}\right\} \\
&\leq B_\infty 8 \max_{i\in S_t}\left\{\sqrt{\frac{2\ln T}{\varepsilon^2 N_{t-1,i}}}\right\}
\end{aligned}
\tag{8}
$$

The first inequality is due to monotonicity and B_∞-bounded smoothness assumption. The second inequality is because the oracle returns S_t which satisfies $r_{\bar{\mu}_t}(S_t) \geq \alpha r_{\bar{\mu}_t}(S_\mu^*)$. The third inequality is due to the definition of $\bar{\mu}_t$ and the concentration bound for $\tilde{\mu}_t$. The last inequality is due to $N_{t,i} \leq T_{t,i}$.

Define $\bar{\Delta}_S = \max_{i \in S} \Delta_{\min}^i$. If $N_{t-1,i} > \frac{128B_\infty^2 \ln T}{\varepsilon^2 \bar{\Delta}_{S_t}^2}$ for any $i \in S_t$, we have $\alpha r_\mu(S_\mu^*) - r_\mu(S_t) < \max_{i \in S_t} \Delta_{\min}^i$ by Equ. 8. On the other hand, by the definition of $\Delta_{\min}^i$, $\alpha r_\mu(S_\mu^*) - r_\mu(S_t) = \alpha \operatorname{opt}_\mu - r_\mu(S_t) \geq \max_{i \in S_t} \Delta_{\min}^i$, which leads to a contradiction. This means that if sub-optimal arm S_t is pulled in step t, and S_t contains base arm i, the counter $N_{t-1,i}$ is at most $\frac{128B_\infty^2 \ln T}{\varepsilon^2 \bar{\Delta}_{S_t}^2} \leq \frac{128B_\infty^2 \ln T}{\varepsilon^2 (\Delta_{\min}^i)^2}$. That is, under high probability event Λ, the counter N_i is at most $\frac{128B_\infty^2 \ln T}{\varepsilon^2 \bar{\Delta}_{S_t}^2}$.

Besides, by Equ. 8, we know that $\Delta_{i,j} \leq 8B_\infty \sqrt{\frac{2 \ln T}{\varepsilon^2 j - 1}}$, since $N_{t-1,i}$ is the minimum counter in $\{N_{t-1,i}, i \in S_t\}$ and increments by one in step t.

Combining with Equ. 7, we have

$$
\begin{aligned}
Reg_{\mu,\alpha,\beta}(T) &\leq \sum_{i \in [m], \Delta_{\min}^i > 0} \sum_{j=1}^{N_{T,i}} \Delta_{i,j} \\
&\leq \sum_{i \in [m], \Delta_{\min}^i > 0} \sum_{j=1}^{N_{T,i}} 8B_\infty \sqrt{\frac{2 \ln T}{\epsilon^2 j}} + 2m\Delta_{\max} \\
&\leq \sum_{i \in [m], \Delta_{\min}^i > 0} \int_0^{N_{T,i}} 8B_\infty \sqrt{\frac{2 \ln T}{\epsilon^2 j}} dj + 2m\Delta_{\max} \\
&\leq \sum_{i \in [m], \Delta_{\min}^i > 0} \frac{128B_\infty^2 \ln T}{\varepsilon^2 \Delta_{\min}^i} + 2m\Delta_{\max}
\end{aligned}
$$

Considering T as the dominant term, we reach the result.

$\square$

B. Proof of Theorem 5

Theorem 5. *For any m and K, and any Δ satisfying $0 < \Delta/B_\infty < 0.35$, the regret of any consistent ε-locally private algorithm π on the CSB problem with B_∞-bounded smoothness is bounded from below as*

$$
\liminf_{T \to \infty} \frac{Reg(T)}{\log T} \geq \frac{B_\infty^2 (m-1)}{64(e^\varepsilon - 1)^2 \Delta}
$$

Specifically, for $0 < \varepsilon \leq 1/2$, the regret is at least

$$
\liminf_{T \to \infty} \frac{Reg(T)}{\log T} \geq \frac{B_\infty^2 (m-1)}{128\varepsilon^2 \Delta}
$$

Proof. We slightly modify the MAB instance in Basu et al. (2019). Suppose there are m arms in a MAB problem. Each arm $i \in [m]$ is associated with an i.i.d Bernoulli random variable μ with mean $\bar{\mu}_i$. If arm i is pulled in a certain step t, instead of receiving reward $\tilde{\mu}(i)$ sampled from the distribution of μ, we receive a reward of $B_\infty \cdot \tilde{\mu}(i)$. Denote the sub-optimality gap of pulling a sub-optimal arm as Δ. Following the argument in Basu et al. (2019), we consider two "MAB" instance: ν_1 with mean weight $\bar{\mu} = \{\Delta/B_\infty, 0, ..., 0\}$ and ν_2 with $\bar{\mu} = \{\Delta/B_\infty, ..., 0, 2\Delta/B_\infty\}$. Similarly, we can show that each supoptimal arm need to be pulled at least

$$
\frac{1}{2 \min\{4, e^{2\varepsilon}\}(e^\varepsilon - 1)^2 D(f_a \| f^*)},
$$

where f_a and f^* denote the weight distribution of arm a and optimal arm. Since $D(f_a\|f^*) \le 4\Delta^2/B_\infty^2$, we have

$$
\begin{aligned}
\liminf_{T\to\infty} \frac{Reg(T)}{\ln T} &\ge (m-1)\frac{1}{2\min\{4, e^{2\varepsilon}\}(e^\varepsilon - 1)^2 D(f_a\|f^*)}\Delta \\
&\ge (m-1)\frac{B_\infty^2}{64(e^\varepsilon - 1)^2\Delta} \\
&\ge (m-1)\frac{B_\infty^2}{128\varepsilon^2\Delta}
\end{aligned}
$$

The second inequality is due to $D\left(p\|q\right) \le \frac{(p-q)^2}{q(1-q)}$ and $\Delta/(B_\infty) \le 0.35 \le \frac{\sqrt{2}}{4}$. The last inequality is for the case that $0 < \varepsilon \le 1/2$.

This special "MAB" problem can reduce to the stochastic CSB problem with B_∞-bounded smoothness. We prove the lower bound by reduction. $\qquad\square$

C. Proof of Theorem 6

Theorem 6. *(Restate) For any m and K such that m/K is an integer, and any Δ satisfying $0 < \Delta/(B_1 K) < 0.35$, the regret of any consistent ε-locally private algorithm π on the CSB problem with B_1-bounded smoothness is bounded from below as*

$$
\liminf_{T\to\infty} \frac{Reg(T)}{\log T} \ge \frac{B_1^2(m - K)K}{64(e^\varepsilon - 1)^2\Delta}
$$

Specifically, for $0 < \varepsilon \le 1/2$, the regret is at least

$$
\liminf_{T\to\infty} \frac{Reg(T)}{\log T} \ge \frac{B_1^2(m - K)K}{128\varepsilon^2\Delta}
$$

Our lower bound is derived on the K-path semi-bandit problem (Kveton et al., 2015): There are m base arms. The feasible super arms are m/K paths. That is, path i (super arm i) contains base arms $(i-1)K + 1, ..., iK$. Suppose the return of choosing super arm S is B_1 times the sum of the weight $\hat{w}_i$ for $i \in S$. The weights of different base arms in the same super arm are identical, and the weights of base arms in different paths are distributed independently. Denote the best super arm as S^*. The weight of each base arm is a Bernoulli random variable with mean:

$$
\bar{w}(i) = \begin{cases} 0.5 & i \in S^* \\ 0.5 - \Delta/(B_1 K) & \text{otherwise} \end{cases}
$$

To prove the lower bound, we adopt general canonical bandit model (Lattimore & Szepesvári, 2018). Denote the privacy-preserving algorithm as π, which maps the observation history to the probability of choosing each super arm, and the CSB instance as ν,. The interaction between the algorithm and the instance in a given horizon T can be denoted as the observation history $\mathcal{H}_T \triangleq \{(S_t, \boldsymbol{Z}_t)\}_{t=1}^T$. An observed history $\mathcal{H}_T$ is a random variable sampled from the measurable space $\left(([m]^k \times \mathbb{R}^k)^T, \mathcal{B}([m]^k \times \mathbb{R}^k)^T\right)$ and a probability measure $\mathbb{P}_{\pi\nu}$. $\mathbb{P}_{\pi\nu}$ is defined as follow:

- The probability of choosing a super arm $S_t = S$ in step t is dictated only by the algorithm $\pi(S|\mathcal{H}_{t-1})$.

- The distribution of rewards $\boldsymbol{X}_t$ in step t is $f_{S_t}^\nu$, which depends on S_t and conditionally independent on the history $\mathcal{H}_{t-1}$.

- In the case of local differential privacy, the algorithm cannot observe X_t directly, but a privated version of rewards $\boldsymbol{Z}_t$. $\boldsymbol{Z}_t$ only depends on X_t and is conditionally independent on the history $\mathcal{H}_{t-1}$. Denote the conditional distribution of $\boldsymbol{Z}$ as $M(\boldsymbol{Z}|\boldsymbol{X})$.

As a result, the distribution of the observed history $\mathcal{H}_T$ is

$$\mathbb{P}_{\pi\nu}^T (\mathcal{H}_T) = \prod_{t=1}^{T} \pi (S_t | \mathcal{H}_{t-1}) f_{S_t}^\nu (\boldsymbol{X}_t) M (\boldsymbol{Z}_t | \boldsymbol{X}_t).$$

Denote $g_{S_t}^\nu (\boldsymbol{Z}) = f_{S_t}^\nu (\boldsymbol{X}_t) M (\boldsymbol{Z}_t | \boldsymbol{X}_t)$. Before proving Theorem 6, we state following two lemmas.

Lemma 3. *Given a stochastic CSB algorithm π and two CSB environment ν_1 and ν_2, the KL divergence of two probability measure $\mathbb{P}_{\pi\nu_1}^T$ and $\mathbb{P}_{\pi\nu_2}^T$ can be decomposed as:*

$$D \left(\mathbb{P}_{\pi\nu_1}^T \| \mathbb{P}_{\pi\nu_2}^T \right) = \sum_{t=1}^{T} \mathbb{E}_{\pi\nu_1} \left[D \left(\pi (S_t | \mathcal{H}_{t-1}, \nu_1) \| \pi (S_t | \mathcal{H}_{t-1}, \nu_2) \right) \right] + \sum_{S \in \mathcal{S}} \mathbb{E}_{\pi\nu_1} \left[N_S(T) \right] D \left(g_S^{\nu_1} \| g_S^{\nu_2} \right),$$

$N_S(T)$ *denotes the number of times S is chosen in T steps.*

Proof.

$$D \left(\mathbb{P}_{\pi\nu_1}^T \| \mathbb{P}_{\pi\nu_2}^T \right) = \int_{\mathcal{H}_T} \ln \frac{d\mathbb{P}_{\pi\nu_1}^T (H)}{d\mathbb{P}_{\pi\nu_2}^T (H)} d\mathbb{P}_{\pi\nu_1}^T (H)$$

$$= \int_{\mathcal{H}_T} \sum_{t=1}^{T} \ln \frac{\pi(S_t | \mathcal{H}_{t-1}, \nu_1)}{\pi(S_t | \mathcal{H}_{t-1}, \nu_2)} d\pi(S_t | \mathcal{H}_{t-1}, \nu_1) + \int_{\mathcal{H}_T} \sum_{t=1}^{T} \ln \frac{g_{S_t}^{\nu_1}(\boldsymbol{Z})}{g_{S_t}^{\nu_2}(\boldsymbol{Z})} d \left(g_{S_t}^{\nu_1}(\boldsymbol{Z}) \right)$$

$$= \sum_{t=1}^{T} \mathbb{E}_{\pi\nu_1} \left[D \left(\pi (S_t | \mathcal{H}_{t-1}, \nu_1) \| \pi (S_t | \mathcal{H}_{t-1}, \nu_2) \right) \right] + \sum_{S \in \mathcal{S}} \left[\sum_{t=1}^{T} \mathbb{E}_{\mathbb{P}_\pi^T \nu_1} \left[\mathbb{1}_{S_t = S} \right] D \left(g_S^{\nu_1}(\boldsymbol{Z}) \| g_S^{\nu_2}(\boldsymbol{Z}) \right) \right]$$

$$= \sum_{t=1}^{T} \mathbb{E}_{\pi\nu_1} \left[D \left(\pi (S_t | \mathcal{H}_{t-1}, \nu_1) \| \pi (S_t | \mathcal{H}_{t-1}, \nu_2) \right) \right] + \sum_{S \in \mathcal{S}} \mathbb{E}_{\pi\nu_1} \left[N_S(T) \right] D \left(g_S^{\nu_1} \| g_S^{\nu_2} \right)$$

$\square$

Lemma 4. *[Theorem 1 in Duchi et al. (2016)] For any $\alpha \geq 0$, let Q be a conditional distribution that guarantees α-differential privacy. Then for any pair of distributions P_1 and P_2, the induced marginal M_1 and M_2 satisfy the bound*

$$D_{\mathrm{kl}} \left(M_1 \| M_2 \right) + D_{\mathrm{kl}} \left(M_2 \| M_1 \right) \leq \min \left\{ 4, e^{2\alpha} \right\} \left(e^\alpha - 1 \right)^2 \| P_1 - P_2 \|_{\mathrm{TV}}^2.$$

Based on these two lemmas, we are now ready to prove Theorem 6.

Proof. (Proof of Theorem 6) Suppose ν_1 denote the stochastic CSB instance with weight vector:

$$w(i) = \begin{cases} 0.5 & i \in S^* \\ 0.5 - \Delta/(B_1 K) & \text{otherwise} \end{cases}$$

For any sub-optimal super arm S^1, denote the CSB instance with the following weight vector as ν_2:

$$w(i) = \begin{cases} 0.5 & i \in S^* \\ 0.5 + \Delta/(B_1 K) & i \in S^1 \\ 0.5 - \Delta/(B_1 K) & \text{otherwise} \end{cases}$$

Denote the expected cumulative regret for a policy π on instance ν in T steps as $Reg(\pi, \nu, T)$. Then we have,

$$\text{Reg}\left(\pi, \nu_1, T\right) \geq \mathbb{P}_{\pi\nu_1}\left(N_{S^1}(T) \geq T/2\right)\frac{T\Delta}{2},$$

$$\text{Reg}\left(\pi, \nu_2, T\right) \geq \mathbb{P}_{\pi\nu_2}\left(N_{S^1}(T) \leq T/2\right)\frac{T\Delta}{2}$$

Combining these two inequality, we have

$$
\begin{aligned}
\text{Reg}\left(\pi, \nu_1, T\right) + \text{Reg}\left(\pi, \nu_2, T\right) &\geq \frac{T\Delta}{2}\left(\mathbb{P}_{\pi\nu_1}\left(N_{S^1}(T) \leq T/2\right) + \mathbb{P}_{\pi\nu_2}\left(N_{S^1}(T) \geq T/2\right)\right) \\
&\geq \frac{T\Delta}{4}\exp\left(-D\left(\mathbb{P}_{\pi\nu_1}^T \| \mathbb{P}_{\pi\nu_2}^T\right)\right)
\end{aligned}
\tag{9}
$$

The second inequality is due to probabilistic Pinsker's inequality (Lattimore & Szepesvári, 2019).

By lemma 3, we have

$$
\begin{aligned}
D\left(\mathbb{P}_{\pi\nu_1}^T \| \mathbb{P}_{\pi\nu_2}^T\right) &= \sum_{t=1}^{T} \mathbb{E}_{\pi\nu_1}\left[D\left(\pi\left(S_t|\mathcal{H}_t, \nu_1\right) \| \pi\left(S_t|\mathcal{H}_t, \nu_2\right)\right)\right] + \sum_{S\in\mathcal{S}} \mathbb{E}_{\pi\nu_1}\left[N_S(T)\right]D\left(g_S^{\nu_1} \| g_S^{\nu_2}\right) \\
&= \sum_{S\in\mathcal{S}} \mathbb{E}_{\pi\nu_1}\left[N_S(T)\right]D\left(g_S^{\nu_1} \| g_S^{\nu_2}\right) \\
&= \mathbb{E}_{\pi\nu_1}\left[N_{S^1}(T)\right]D\left(g_{S^1}^{\nu_1} \| g_{S^1}^{\nu_2}\right)
\end{aligned}
\tag{10}
$$

The second equality is because π chooses S_t based on the observed history $\mathcal{H}_t$. The third equality is because ν_1 and ν_2 only differs in S^1.

By combining Equ. 9 and Equ. 10 we get,

$$
\begin{aligned}
\mathbb{E}_{\pi\nu_1}\left[N_{S^1}(T)\right] &= D\left(\mathbb{P}_{\pi\nu_1}^T \| \mathbb{P}_{\pi\nu_2}^T\right) / D\left(g_{S^1}^{\nu_1} \| g_{S^1}^{\nu_2}\right) \\
&\geq \ln\left(\frac{T\Delta}{4\left(\text{Reg}\left(\pi, \nu_1, T\right) + \text{Reg}\left(\pi, \nu_2, T\right)\right)}\right) / D\left(g_{S^1}^{\nu_1} \| g_{S^1}^{\nu_2}\right) \\
&\geq \frac{\ln(T)/4 - \ln(8m/K)}{D\left(g_{S^1}^{\nu_1} \| g_{S^1}^{\nu_2}\right)} \\
&\geq \frac{\ln(T)/4 - \ln(8m/K)}{\min\left\{4, e^{2\varepsilon}\right\}\left(e^{\varepsilon} - 1\right)^2 \left\|f_{S^1}^{\nu_1} - f_{S^1}^{\nu_2}\right\|_{\text{TV}}^2} \\
&\geq \frac{\ln(T)/2 - 2\ln(8m/K)}{\min\left\{4, e^{2\varepsilon}\right\}\left(e^{\varepsilon} - 1\right)^2 D\left(f_{S^1}^{\nu_1} \| f_{S^1}^{\nu_2}\right)} \\
&\geq \frac{K^2 B_1^2\left(\ln(T)/16 - \ln(8m/K)/8\right)}{\min\left\{4, e^{2\varepsilon}\right\}\left(e^{\varepsilon} - 1\right)^2 \Delta^2}
\end{aligned}
$$

The first inequality is due to Equ. 9. The second inequality is due to the consistent algorithm setting, i.e. $\text{Reg}\left(\pi, \nu_1, T\right) \leq \frac{m}{k}\Delta T^p$. Here we set $p = 3/4$. The third inequality is due to Lemma 4. The forth inequality is due to Pinsker's inequality. The last inequality is due to $D\left(p\|q\right) \leq \frac{(p-q)^2}{q(1-q)}$ and $\Delta/(B_1 K) \leq 0.35 \leq \frac{\sqrt{2}}{4}$.

Now we can bound $\liminf_{T \to \infty} \frac{Reg(T)}{\log T}$:

$$
\begin{aligned}
\liminf_{T \to \infty} \frac{Reg(T)}{\ln T} &= \liminf_{T \to \infty} \frac{\sum_{S \in \mathcal{S}, S \neq S^*} \Delta \cdot \mathbb{E}_{\pi \nu_1}[N_S(T)]}{\ln T} \\
&\geq \liminf_{T \to \infty} \frac{B_1^2 \left(m/K - 1\right) \Delta K^2 \left(\ln(T)/16 - \ln(8m/K)/8\right)}{\min\{4, e^{2\varepsilon}\} \left(e^\varepsilon - 1\right)^2 \Delta^2 \ln T} \\
&= \frac{B_1^2 mK}{16 \min\{4, e^{2\varepsilon}\} \left(e^\varepsilon - 1\right)^2 \Delta} \\
&\geq \frac{B_1^2 mK}{128\varepsilon^2 \Delta}
\end{aligned}
$$

The last inequality is due to $(e^\varepsilon - 1)^2 \leq 2\varepsilon^2$ for $0 < \varepsilon \leq 1/2$. □

D. Omitted Proof of Theorem 8

Before proving Theorem 8, we consider following two events, and show that these events happen with high probability.

Lemma 5. *Let* $\text{Sum}_{t,i}$ *be the sum of previous outcome* $X_{t,i}$ *without privacy noise for base arm* i *in the first* t *steps. We denote the following event as* Λ_1*: For any step* $t \in [T]$ *and any base arm* $i \in [m]$*,*

$$
\left| \frac{\text{Sum}_{t,i}}{T_{t,i}} - \mu_i \right| \leq \sqrt{\frac{4 \ln T}{T_{t,i}}}
$$

Then $\Pr[\Lambda_1] \geq 1 - 2/T$.

Proof. The result follows directly from Hoeffding's inequality and union bounds for all steps $t \in [T]$. □

Lemma 6. *Let* $\text{Noise}_{t,i}$ *be the Laplace noise added to* $X_{t,i}$ *in step* t*. We denote the following event as* Λ_2*: For any step* $t \in [T]$ *and any base arm* $i \in [m]$*,*

$$
\left| \frac{\text{Noise}_{t,i}}{T_{t,i}} \right| \leq \frac{12K \ln^3 T}{T_{t,i}\varepsilon}
$$

Then $\Pr[\Lambda_2] \geq 1 - 1/(mT)$.

Proof. From the argument of our algorithm, $\text{Noise}_{t,i}$ is the sum of at most $\log T$ i.i.d random variables drawn from $\text{Lap}(2K \log T/\varepsilon)$. By the tail probability of Laplace distribution, we know that for any $\nu \sim \text{Lap}(2K \log T/\varepsilon)$, with prob. $1 - \delta$, $|\nu| \leq 2K \log T \ln(1/\delta)/\varepsilon$. Set $\delta = 1/(m^2 T^2 \log T)$. By union bounds over $\log T$ random variables, we have $|\text{Noise}_{t,i}| \leq 4K \log^2 T \ln(mT \log T)/\varepsilon$ with prob. $1 - 1/(m^2 T^2)$ for a fixed i and t. By union bound over all base arm i and step t, we prove that

$$
\left| \frac{\text{Noise}_{t,i}}{T_{t,i}} \right| \leq \frac{4K \log^2 T \ln(mT \log T)}{T_{t,i}\varepsilon} \leq \frac{12K \ln^3 T}{T_{t,i}\varepsilon}
$$

for any step t and base arm i for sufficiently large T with prob. $1 - 1/(mT)$. □

Proof. (Proof of Lemma 2) Suppose G_t denote the event that the oracle fails to produce an α-approximate answer with respect to the input vector in step t. Similar with the proof of Theorem 4, the cumulative regret in the steps that G_t happens is at most $R_{\text{fail}} \leq (1 - \beta)T\Delta_{\max}$.

Then we have,

$$Reg_{\mu,\alpha,\beta}(T) \leq T\alpha\beta \operatorname{opt}_\mu - \mathbb{E}\sum_{t=1}^T r_\mu(S_t)$$

$$\leq R_{\text{fail}} + T\alpha\beta \operatorname{opt}_\mu - \left(T\alpha \operatorname{opt}_\mu - \sum_{t\in[T]} \Delta_t \mathbb{1}\{\neg G_t\} \right)$$

$$\leq \sum_{t\in[T]} \Delta_t \mathbb{1}\{\neg G_t\}$$

Here Δ_t denote the sub-optimal gap in step t.

This means that we only need to consider the steps that G_t doesn't happen. Denote $\hat{R}(T)$ as the regret if event Λ_1 and Λ_2 happen.

$$Reg_{\mu,\alpha,\beta}(T) \leq \Pr\{\Lambda_1 \cap \Lambda_2\}\hat{R}(T) + \sum_{i\in[m]} \Delta_{\min}^i$$

$$+ \Pr\{\neg\Lambda_1\}T\Delta_{\max} + \Pr\{\neg\Lambda_1\}T\Delta_{\max}$$

$$\leq \hat{R}(T) + (m+2)\Delta_{\max}$$

If event Λ_1 and Λ_2 happen, we have

$$|\tilde{\mu}_t(i) - \mu_i| = \left| \frac{\operatorname{Sum}_{t,i}}{T_{t,i}} - \mu_i + \frac{\operatorname{Noise}_{t,i}}{T_{t,i}} \right|$$

$$\leq \sqrt{\frac{4\ln T}{T_{t,i}}} + \frac{12K\ln^3 T}{T_{t,i}}$$

for step $t \in [T]$, if we choose a sub-optimal super arm with sub-optimality gap $\Delta_{S_t} > 0$, then we have

$$\alpha r_\mu(S_\mu^*) - r_\mu(S_t) \leq \alpha r_{\bar{\mu}_t}(S_\mu^*) - (r_{\bar{\mu}_t}(S_t) - B_1\|\bar{\mu}_t - \mu\|_1)$$

$$\leq B_1\|\bar{\mu}_t - \mu\|_1$$

$$\leq B_1(\|\bar{\mu}_t - \tilde{\mu}_t\|_1 + \|\tilde{\mu}_t - \mu\|_1)$$

$$\leq B_1 \sum_{i\in S_t} \left(4\sqrt{\frac{\ln T}{T_{t-1,i}}} + \frac{24K\ln^3 T}{T_{t-1,i}\varepsilon} \right) \tag{11}$$

The first inequality is due to L_1 smoothness assumption. The second inequality is because the oracle returns S_t which satisfies $r_{\bar{\mu}_t}(S_t) \geq \alpha r_{\bar{\mu}_t}(S_\mu^*)$. The last inequality is due to the definition of $\bar{\mu}_t$ and the concentration bound for $\tilde{\mu}_t$.

This shows that if event Λ_1 and Λ_2 happen, and we choose a sub-optimal super arm with sub-optimality gap $\Delta_{S_t} > 0$ in step t, F_t happens.

Then we have $\hat{R}(T) \leq \sum_{t\in[T]} \Delta_{S_t}\mathbb{1}\{F_t\}$, which finishes the proof. $\square$

E. Proof of Theorem 9

Theorem 9. *For any m and K such that $m \geq 2K$,and any Δ satisfying $0 < \Delta/(B_1 K) < 0.35$, the regret for any consistent ε-DP algorithm on the CSB problem with B_1 bounded smoothness is at least $\Omega\left(\frac{B_1^2 mK\ln T}{\Delta} + \frac{B_1 mK\ln T}{\varepsilon} \right)$.*

Proof. Previous results have shown that the regret for any non-private CSB algorithm is at least $\Omega\left(\frac{mK\ln T}{\Delta} \right)$ (Kveton et al., 2015). They consider linear CSB problem, which is a special case of B_1 bounded smoothness CSB with $B_1 = 1$. We

slightly modify the hard instance in Kveton et al. (2015) and prove the regret lower bound for B_1 bounded smoothness CSB in non-private setting.

The main difference is that we assume the reward of any super arms S_t is B_1 times the sum of weights $w(i)$ for $i \in S_t$. In our hard instance, we also consider the K-path semi-bandit problem. There are m base arms. The feasible super arms are m/K paths. Path i (Super arm i) contains base arms $(i-1)K+1, (i-1)K+2, ..., iK$. The weight of base arm i is a Bernoulli random variable with mean $\bar{w}(i)$. Since Δ in our setting is B_1 times that of the instance in Kveton et al. (2015), we slightly modify the mean of $w(i)$ to make sure that the mean $\bar{w}(i) \in [0,1]$:

$$
\bar{w}(i) = \begin{cases} 0.5 & i \in S^* \\ 0.5 - \Delta/(B_1 K) & \text{otherwise} \end{cases}
$$

With the same argument in Kveton et al. (2015), we can prove that each path need to be selected at least $\frac{B_1^2 K^2 \ln T}{\Delta^2}$ times. which means that the regret is at least $\frac{B_1^2 K^2 \ln T}{\Delta^2} \Delta \cdot (L/K - 1) = \Omega\left(\frac{B_1^2 m K \ln T}{\Delta}\right)$. Since private CSB is harder than non-private CSB (There is a reduction from non-private CSB to private CSB), the regret of private CSB is at least $\Omega\left(\frac{B_1^2 m K \ln T}{\Delta}\right)$.

By the following lemma, we can show that the regret of any ε-DP consistent CSB algorithm is at least $\Omega\left(\frac{B_1 m K \ln T}{\varepsilon}\right)$. Combining both results, we can prove that the regret lower bound is $\Omega\left(\max\left\{\frac{B_1^2 m K \ln T}{\Delta}, \frac{B_1 m K \ln T}{\varepsilon}\right\}\right) = \Omega\left(\frac{B_1^2 m K \ln T}{\Delta} + \frac{B_1 m K \ln T}{\varepsilon}\right)$. $\square$

Lemma 7. *For any m and K such that $m \geq 2K$, and any Δ satisfying $0 < \Delta/(B_1 K) < 0.35$, the regret for any consistent CSB algorithm guaranteeing ε-DP is at least $\Omega\left(\frac{B_1 m K \ln T}{\varepsilon}\right)$.*

Now we only need to prove Lemma 7.

Proof. We consider the CSB instance: Suppose there are m base arms, each associated with a weight sampled from Bernoulli distribution. These m base arms are divided into three sets, $S^*, \tilde{S}, \bar{S}$. S^* contains m base arms, which build up the optimal super arm set. $\tilde{S}$ contains $K-1$ "public" base arms for sub-optimal super arms. These arms are contained in all sub-optimal super arms. $\bar{S}$ contains $m - 2K + 1$ base arms. each base arm combined with $K-1$ "public" base arms in $\tilde{S}$ builds up a sub-optimal super arm. Totally we have $m - 2K + 1$ sub-optimal super arms and one optimal super arm. The mean of the Bernoulli random variable associated to each base arm is defined as follow:

$$
w(i) = \begin{cases} 0.5 & i \in S^* \\ 0.5 - \Delta/(B_1 K) & \text{otherwise} \end{cases}
$$

The weights of base arms in $\tilde{S}$ are identical, while other weights are i.i.d sampled. The reward of pulling a super arm S is B_1 times the sum of weights of all base arm $i \in S$. As a result, the sub-optimality gap of each sub-optimal super arm is Δ. We denote this CSB instance as ν_1.

Now we fix one certain sub-optimal super arm S_1. Denote E_{S_1} as the event that super arm S_1 is pulled $\leq \frac{B_1 K \ln T}{400 \varepsilon \Delta} := t_S$ times. Our goal is to show that E_{S_1} happens with probability at most $\frac{1}{2m}$. If this is true, by union bounds over all sub-optimal super arms, all the sub-optimal super arms will be pulled at least t_S times with prob. $1 - \frac{1}{2}$. This means the regret is at least $\Omega\left(\frac{B_1 m K \ln T}{\varepsilon}\right)$.

Now we prove that $P_{\nu_1}(E_{S_1}) \leq 1/(2m)$. Our analysis is inspired by the work of Shariff & Sheffet (2018). Consider another CSB instance with all the setting same as ν_1, except that the mean weights of base arms in S_1 are increased by $2\Delta/(B_1 K)$ each. We denote this instance as ν_2. Consider the case that rewards are drawn from ν_2. Due to consistent

property, the regret of the algorithm is at most $T^{3/4}m\Delta$. For sufficiently large T, we have

$$\frac{T\Delta}{2K}\mathbb{P}_{\nu_2}[E] \leq \frac{(T-t_S)\Delta}{K}\mathbb{P}_{\nu_2}[E] \leq T^{3/4}m\Delta$$

.

The first inequality is for sufficiently large T. The second inequality is because if E happens in ν_2, the regret is at least $(T-t_s) \cdot \frac{\Delta}{K}$. This means that $\mathbb{P}_{\nu_2}[E] \leq \frac{mK}{T^{1/4}}$.

Now we consider the influence of differential privacy. The result of Karwa & Vadhan (2017) (Lemma 6.1) states that the group privacy between the case that inputs are drawn i.i.d from distribution P_1 and P_2 is proportional to $6\varepsilon n \cdot d_{\text{TV}}(P, Q)$, where n is the number of inputs data. We apply the coupling argument in Karwa & Vadhan (2017) to our setting. Suppose the algorithm turns to an oracle when she needs to sample a reward of super arm S_1. The oracle can generate at most t_S pairs of data. The left ones are i.i.d sampled from ν_1, while the right ones are i.i.d sampled from ν_2. Whether the algorithm receive a reward sampled from the left or the right depends on the true environment. The algorithm turns to another oracle if and only if the original oracle runs out of t_S samples. By Lemma 6.1 in Karwa & Vadhan (2017), the oracle runs out of t_S samples, i.e. event E_{S_1} happens with similar probability under ν_1 and ν_2. Indeed, the probability of event E_{S_1} happens under ν_1 is less than $\exp(6\varepsilon t_S \cdot d_{\text{TV}}(P, Q))$ times the probability of event E_{S_1} happens under ν_2.

That is, for sufficiently large T,

$$\mathbb{P}_{\nu_1}[E_{S_1}] \leq \exp(6\varepsilon t_S \cdot d_{\text{TV}}(\nu_1, \nu_2))\mathbb{P}_{\nu_2}[E_{S_1}]$$
$$\leq \exp\left(24\varepsilon t_S \cdot \frac{\Delta}{B_1 K}\right)\mathbb{P}_{\nu_2}[E_{S_1}]$$
$$\leq \exp(0.06 \ln T)\frac{mK}{T^{1/4}}$$
$$= mKT^{-0.19} \leq \frac{1}{2m}.$$

The second inequality is due to $d_{\text{TV}}(\nu_1, \nu_2) \leq \sqrt{\frac{D_{KL}(\nu_1\|\nu_2)}{2}} \leq 4\Delta/(B_1 K)$ by Pinsker's inequality and the setting that the public base arms are identical.

$\square$