Optimal Rates of (Locally) Differentially Private Heavy-tailed Multi-Armed Bandits

Youming Tao * Shandong University
Yulian Wu * KAUST
Peng Zhao Nanjing University
Di Wang KAUST

Abstract

In this paper we investigate the problem of stochastic multi-armed bandits (MAB) in the (local) differential privacy (DP/LDP) model. Unlike previous results that assume bounded/sub-Gaussian reward distributions, we focus on the setting where each arm's reward distribution only has $(1+v)$-th moment with some $v \in (0, 1]$. In the first part, we study the problem in the central ϵ-DP model. We first provide a near-optimal result by developing a private and robust Upper Confidence Bound (UCB) algorithm. Then, we improve the result via a private and robust version of the Successive Elimination (SE) algorithm. Finally, we establish the lower bound to show that the instance-dependent regret of our improved algorithm is optimal. In the second part, we study the problem in the ϵ-LDP model. We propose an algorithm that can be seen as locally private and robust version of SE algorithm, which provably achieves (near) optimal rates for both instance-dependent and instance-independent regret. Our results reveal differences between the problem of private MAB with bounded/sub-Gaussian rewards and heavy-tailed rewards. To achieve these (near) optimal rates, we develop several new hard instances and private robust estimators as byproducts, which might be used to other related problems. Finally, experiments also support our theoretical findings and show the effectiveness of our algorithms.

*The first two authors contributed equally. Part of the work was done when Youming Tao was a research intern at KAUST, and Yulian Wu was a student at East China Normal University.

Proceedings of the 25th International Conference on Artificial Intelligence and Statistics (AISTATS) 2022, Valencia, Spain. PMLR: Volume 151. Copyright 2022 by the author(s).

1 INTRODUCTION

As one of the most fundamental problems in statistics and machine learning, (stochastic) Multi-Armed Bandits (MAB), and its general form, bandit learning, have already been studied for more than half a century, starting from Thompson (1933) and Robbins (1952). They find numerous applications in many areas such as medicine (Gutiérrez et al., 2017), finance (Shen et al., 2015), social science (Nakayama et al., 2017), and clinical research (Press, 2009). The wide applications of bandit learning also present some new challenges to existing methods. Particularly, due to the existence of sensitive data and their distributed nature in many applications like recommendation system, biomedicine, and genomics, it is often challenging to preserve the privacy of such data, which makes the data extremely difficult to aggregate and learn from.

To preserve the privacy of these sensitive data, Differential Privacy (DP) (Dwork et al., 2006) has received a great deal of attention and now has established itself as a de facto notation of privacy for data analysis. Over the past decade, differentially private bandit learning has been extensively studied from various setups including classical stochastic MAB (Mishra and Thakurta, 2015; Tossou and Dimitrakakis, 2016; Sajed and Sheffet, 2019; Ren et al., 2020; Kalogerias et al., 2020), combinatorial semi-bandits (Chen et al., 2020), and contextual bandits (Shariff and Sheffet, 2018; Hannun et al., 2019; Malekzadeh et al., 2020; Zheng et al., 2020). Additionally, Wang et al. (2020b); Dubey and Pentland (2020a,b) recently investigated bandit learning in the federated/distributed setting.

However, these problems are still not well-understood. For example, all of the previous results and methods need to assume that the rewards are sampled from some bounded (or sub-Gaussian) distributions to guarantee the DP property. However, such assumptions may not hold when designing decision-making algorithms for complicated real-world systems. In particular, previous papers have shown that the rewards or the interactions in such systems often lead to heavy-tailed and power

Table 1: Summary of our contributions and comparison with the bounded/sub-Gaussian reward distribution. All the results are in the expected regret form. For the heavy-tailed reward distribution case, we assume the $(1+v)$-th moment of each reward distribution is bounded by 1 for some known $v \in (0,1]$. For the bounded reward distribution case, we assume the rewards are bounded by 1. For the sub-Gaussian reward distribution case, we assume the variance of each reward distributed is bounded by 1. Here K is the number of arms, T is the number of rounds, and Δ_a is the mean reward gap of arm a.

Problem	Model	Upper Bound	Lower Bound
Heavy-tailed Reward (Instance-dependent Bound)	ϵ-DP	$O\left(\frac{\log T}{\epsilon}\sum_{\Delta_a>0}\left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}}+\max_a \Delta_a\right)$	$\Omega\left(\frac{\log T}{\epsilon}\sum_{\Delta_a>0}\left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}}\right)$
	ϵ-LDP	$O\left(\frac{\log T}{\epsilon^2}\sum_{\Delta_a>0}\left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}}+\max_a \Delta_a\right)$	$\Omega\left(\frac{\log T}{\epsilon^2}\sum_{\Delta_a>0}\left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}}\right)$
Bounded/sub-Gaussian Reward (Instance-dependent Bound)	ϵ-DP	$O\left(\frac{K\log T}{\epsilon}+\sum_{\Delta_a>0}\frac{\log T}{\Delta_a}\right)$ (Sajed and Sheffet, 2019)	$\Omega\left(\frac{K\log T}{\epsilon}+\sum_{\Delta_a>0}\frac{\log T}{\Delta_a}\right)$ (Shariff and Sheffet, 2018)
	ϵ-LDP	$O\left(\frac{1}{\epsilon^2}\sum_{\Delta_a>0}\frac{\log T}{\Delta_a}+\Delta_a\right)$ (Ren et al., 2020)	$\Omega\left(\frac{1}{\epsilon^2}\sum_{\Delta_a>0}\frac{\log T}{\Delta_a}\right)$ (Ren et al., 2020)
Heavy-tailed Reward (Instance-independent Bound)	ϵ-DP	$O\left(\left(\frac{K\log T}{\epsilon}\right)^{\frac{v}{1+v}}T^{\frac{1}{1+v}}\right)$	—
	ϵ-LDP	$O\left(\left(\frac{K\log T}{\epsilon^2}\right)^{\frac{v}{1+v}}T^{\frac{1}{1+v}}\right)$	$\Omega\left(\left(\frac{K}{\epsilon^2}\right)^{\frac{v}{1+v}}T^{\frac{1}{1+v}}\right)$
Bounded/sub-Gaussian Reward (Instance-independent Bound)	ϵ-DP	$O\left(\sqrt{KT\log T}+\frac{K\log T}{\epsilon}\right)$ (Sajed and Sheffet, 2019)	$\Omega\left(\sqrt{KT}+\frac{K\log T}{\epsilon}\right)$ (Sajed and Sheffet, 2019)
	ϵ-LDP	$O\left(\frac{\sqrt{KT\log T}}{\epsilon}\right)$ (Ren et al., 2020)	$\Omega(\frac{\sqrt{KT}}{\epsilon})$ (Basu et al., 2019)

law distributions (Dubey and Pentland, 2019), such as modeling stock prices (Bradley and Taqqu, 2003), preferential attachment in social networks (Mahanti et al., 2013), and online behavior on websites (Kumar and Tomkins, 2010). Thus, it is necessary to develop new methods to deal with these heavy-tailed rewards in the private bandit learning.

To address the above issue, in this paper, we focus on the most fundamental bandit model, *i.e.,* multi-armed bandits, with heavy-tailed rewards. We conduct a comprehensive and the first study on MAB with heavy-tailed rewards in both central and local DP models, where the reward distribution of each arm only has the $(1+v)$-th moment for some $v \in (0,1]$. Our contributions are summarized as follows.

- In the first part (Section 4), we consider the problem in the central ϵ-DP model. Specifically, we first propose a method based on a robust version of the Upper Confidence Bound (UCB) algorithm, and also design a new mechanism that could be seen as an adaptive version of the Tree-based mechanism (Dwork et al., 2010). To further improve the result, we then develop a private and robust version of the Successive Elimination (SE) algorithm and show that the (expected) regret bound is improved by a factor of $\log^{1.5+\frac{1}{v}} T$, where T is the number of rounds. Moreover, we establish the lower bound and show that the instance-dependent regret bound of $O\left(\frac{\log T}{\epsilon}\sum_{\Delta_a>0}\left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}}+\max_a \Delta_a\right)$ achieved by our second algorithm is optimal (up to poly$(\log\log \frac{1}{\Delta_a})$ factors), where Δ_a is the mean gap defined in Section 3.1.

- In the second part (Section 5), we study the problem in the ϵ-LDP model. We first develop a LDP version of the SE algorithm which achieves an instance-dependent regret bound of $O\left(\frac{\log T}{\epsilon^2}\sum_{\Delta_a>0}\left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}}+\max_a \Delta_a\right)$ and an $\widetilde{O}\left(\left(\frac{K}{\epsilon^2}\right)^{\frac{v}{1+v}}T^{\frac{1}{1+v}}\right)$ instance-independent bound. Then, we show that the above instance-dependent regret bound is optimal and the instance-independent regret bound is near-optimal (up to poly$(\log T)$ factors).

- All of our results also reveal the differences between the problem of private MAB with bounded/sub-Guassian rewards and that with heavy-tailed rewards (see Table 1 for details). To achieve these (near) optimal results, we develop several new hard instances, mechanisms and private robust estimators as byproducts, which could be used to other related problems, such as private contextual bandits (Shariff and Sheffet, 2018) or private reinforcement learning (Vietri et al., 2020).

Due to space limitation, all the technical lemmas and proofs are included in the appendix. The source code is also included in Supplementary Materials.

2 RELATED WORK

As mentioned earlier, there are enormous previous works on either MAB with bounded/sub-Gaussian reward distributions in the (local) DP model (Mishra and Thakurta, 2015; Tossou and Dimitrakakis, 2016; Gajane et al., 2018; Shariff and Sheffet, 2018; Basu et al., 2019; Sajed and Sheffet, 2019; Ren et al., 2020; Vietri et al., 2020; Zheng et al., 2020) or MAB with heavy-tailed reward distributions (Bubeck et al., 2013; Lee

et al., 2020; Yu et al., 2018; Lattimore, 2017; Agrawal et al., 2021; Vakili et al., 2013; Agrawal et al., 2020). However, to the best of our knowledge, MAB with heavy-tailed reward in the (local) DP model has not been studied before. In the following we only discuss previous works that are the most close to ours.

In the previous studies of MAB with bounded/sub-Gaussian rewards, to guarantee DP property, the most direct way is to modify the classical UCB algorithm (Mishra and Thakurta, 2015; Tossou and Dimitrakakis, 2016; Ren et al., 2020). Our first algorithm is motivated by a robust version of the UCB algorithm in (Bubeck et al., 2013). However, there are several differences. First, unlike the non-private setting, in this paper we show that this approach could only achieve a suboptimal instance-dependent (expected) regret bound, due to the noises added by the Tree-based mechanism (Chan et al., 2011). Secondly, due to the added noises, parameters such as the thresholds are quite different with the non-private case. To achieve an improved regret bound, our second algorithm is based on the Successive Elimination (SE) (Even-Dar et al., 2006) algorithm, whose private version has been studied in (Sajed and Sheffet, 2019) for the bounded reward case. In this paper we extend the method to the heavy-tailed case and the LDP model. Our algorithms are provably (near) optimal. For the lower bounds, previous papers established hard instances for either private MAB with bounded rewards (Basu et al., 2019; Sajed and Sheffet, 2019; Ren et al., 2020) or heavy-tailed MAB in the non-private case (Bubeck et al., 2013). However, these instances cannot be used to our problem and this paper builds new hard instances.

Private and robust estimation has drawn much attention in recent years. Barber and Duchi (2014) provided the first study on private mean estimation for distributions with bounded moment, which is extended by Kamath et al. (2020); Brunel and Avella-Medina (2020); Liu et al. (2021) recently. However, all of them need to assume the underlying distribution has the second-order moment, while in this paper we only need to assume the reward distributions have the $(1+v)$-th moment for some $v \in (0, 1]$. Moreover, all of these works only focus on the central DP model and offline setting, and it is generally unclear whether they could be extended to the stream setting. Thus, our problem is more general. In addition to the mean estimation problem, recently Wang et al. (2020a) studied differentially private stochastic convex optimization with heavy-tailed data, while their work still requires to assume the distribution of gradient has second-order moment and cannot be used to the stream setting.

3 PRELIMINARIES

In this section, we present some preliminaries for MAB with heavy-tailed rewards and differential privacy.

3.1 MAB with Heavy-tailed Rewards

In a stochastic multi-armed bandits (MAB) problem, there is a learner interacting with the environment sequentially over T rounds. The learner is faced with a set of K independent arms $\{1, \dots, K\}$. In each round $t \in [T]$, the learner selects an arm $a_t \in [K]$ to pull and then obtains a reward x_t drawn i.i.d. from a fixed but unknown probability distribution $\mathcal{X}_{a_t}$ associated with the chosen arm. Denote by μ_a the mean of each distribution $\mathcal{X}_a$ for $a \in [K]$, and by $\mu^* = \max_{a\in[K]} \mu_a$ the maximum. Define $\Delta_a \triangleq \mu^* - \mu_a$ as the mean reward gap for arm a. The learner aims to maximize her/his expected cumulative reward over time, *i.e.,* to minimize the (expected) cumulative *regret*, defined as

$$\mathcal{R}_T \triangleq T\mu^* - \mathbb{E}\left[\sum_{t=1}^{T} x_t\right], \tag{1}$$

where the expectation is taken with respect to all the randomness. This paper considers a heavy-tailed setting where each arm's reward distribution only has finite raw moments of order $1+v$ for some $v \in (0, 1]$. Concretely, we assume that there is a constant $u > 0$ such that for each reward distribution $\mathcal{X}_a$,

$$\mathbb{E}_{X\sim\mathcal{X}_a}[|X|^{1+v}] \le u. \tag{2}$$

In this paper, we assume both v and u are known constants, *i.e.,* for any constant c we regard $c^{\frac{1}{v}}$ as a constant. Note that the assumption is commonly used in robust estimation (Catoni, 2012; Kamath et al., 2020; Wang et al., 2020a) and MAB with heavy-tailed rewards (Bubeck et al., 2013; Dubey and Pentland, 2019; Lee et al., 2020; Agrawal et al., 2021).

Instead of the assumption on the raw moment in (2), there is another assumption on the central moment instead, *i.e.,* $\mathbb{E}_{X\sim\mathcal{X}_a}[|X - \mathbb{E}(X)|^{1+v}] \le u$. We note that both of the raw moment and central moment assumptions have been studied in the previous work on private robust estimation (Wang et al., 2020a; Kamath et al., 2020) for the mean estimation of distributions with bounded second-order moment. Here, we claim that, the bounded raw moment implies that the bounded central moment, and vice versa. See Lemma 10 in Appendix for details.

3.2 Differential Privacy

We introduce the definition of differential privacy (DP) in the stream setting since rewards are released continually. According to (Dwork et al., 2010), for data

streams there are two different settings, *i.e.,* event-level setting and user-level setting. This paper will focus on the event-level setting, *i.e.,* two data streams σ and σ' are adjacent if they differ at exactly one timestep. Intuitively, an algorithm is differentially private if it cannot be used to distinguish any two adjacent streams.

Definition 1 (Differential Privacy (Dwork et al., 2010))**.** An algorithm $\mathcal{M}$ is ϵ-differentially private (DP) if for any adjacent streams σ and σ', and any measurable subset $\mathcal{O}$ of the output space of $\mathcal{M}$, we have $\mathbb{P}[\mathcal{M}(\sigma) \in \mathcal{O}] \leq e^{\epsilon} \cdot \mathbb{P}[\mathcal{M}(\sigma') \in \mathcal{O}]$.

Compared with DP, Local Differential Privacy (LDP) is a stronger notion of privacy. In LDP, each data is perturbed before collection to ensure privacy.

Definition 2 (Local Differential Privacy)**.** An algorithm $\mathcal{M}: \mathcal{X} \to \mathcal{Y}$ is said to be ϵ-locally differentially private (LDP) if for any $x, x' \in \mathcal{X}$, and any measurable subset $\mathcal{O} \subset \mathcal{Y}$, it holds that $\mathbb{P}[\mathcal{M}(x) \in \mathcal{O}] \leq e^{\epsilon} \cdot \mathbb{P}[\mathcal{M}(x') \in \mathcal{O}]$.

In this paper, we will mainly use the Laplacian and an adaptive version of the Tree-based mechanism (see Section 4 for details), and the parallel composition theorem to guarantee the DP property.

Lemma 1 (Parallel Composition)**.** Suppose there are n ϵ-differentially private mechanisms $\{\mathcal{M}_i\}_{i=1}^n$ and n disjoint datasets denoted by $\{D_i\}_{i=1}^n$. Then the algorithm, which applies each $\mathcal{M}_i$ on the corresponding D_i, preserves ϵ-DP in total.

Definition 3 (Laplacian Mechanism)**.** Given a function $f: \mathcal{X}^n \to \mathbb{R}^d$, the Laplacian Mechanism is defined as: $\mathcal{M}_L(D, f, \epsilon) = f(D) + (Y_1, Y_2, \cdots, Y_d)$, where Y_i is i.i.d. drawn from a Laplacian distribution $\text{Lap}(\frac{\Delta_1(f)}{\epsilon})$, where $\Delta_1(f)$ is the ℓ_1-sensitivity of the function f, *i.e.,* $\Delta_1(f) = \sup_{D \sim D'} ||f(D) - f(D')||_1$. Here, $D \sim D'$ denotes that D and D' are neighbouring datasets, *i.e.,* those that differ in exactly on entry. For a parameter λ, the Laplacian distribution has the density function $\text{Lap}(\lambda)(x) = \frac{1}{2\lambda} \exp(-\frac{|x|}{\lambda})$. Laplacian Mechanism preserves ϵ-DP.

4 DP HEAVY-TAILED MAB

In this section, we will study the problem of designing ϵ-DP algorithms for MAB with heavy-tailed rewards. Recall that, in the classical setting where the rewards follow some bounded distributions, the most commonly used approach is using the Tree-based mechanism to privately calculate the sum of rewards and then modify the Upper Confidence Bound (UCB) algorithm (Auer et al., 2002), such as (Mishra and Thakurta, 2015; Tossou and Dimitrakakis, 2016). However, their methods cannot be directly generalized to the heavy-tailed setting, since now the reward is unbounded. Thus, the most natural idea is to first preprocess the rewards to make them bounded and then use the Tree-based mechanism and UCB algorithm.

Algorithm 1 DP Robust Upper Confidence Bound

Input: time horizon T, parameters ϵ, v, u.

1: Create an empty tree Tree_a for each arm $a \in [K]$.
2: Initialize pull number $n_a \leftarrow 0$ for each arm $a \in [K]$.
3: Denote B_n as $(\frac{\epsilon u n}{\log^{1.5} T})^{1/(1+v)}$ for any $n \in \mathbb{N}^+$.
4: **for** $t = 1, \ldots, K$ **do**
5: Pull arm t and observe a reward x_t.
6: Update the pull number $n_t \leftarrow n_t + 1$.
7: Truncate the reward by $\widetilde{x}_t \leftarrow x_t \cdot \mathbb{I}_{|x_t| \leq B_{n_t}}$.
8: Insert $\widetilde{x}_t$ into Tree_t.
9: **end for**
10: **for** $t = K+1, \ldots, T$ **do**
11: Obtain $\widehat{S}_a(t)$ for each $a \in [K]$ via Algorithm 2.
12: Pull arm

$$a_t = \arg\max_a \frac{\widehat{S}_a(t)}{n_a} + 18u^{\frac{1}{1+v}} \left(\frac{\log(2t^4)\log^{1.5+\frac{1}{v}} T}{n_a \epsilon}\right)^{\frac{v}{1+v}}$$

and observe the reward x_t.
13: Update the pull number $n_{a_t} \leftarrow n_{a_t} + 1$.
14: Truncate the reward by $\widetilde{x}_t \leftarrow x_t \cdot \mathbb{I}_{|x_t| \leq B_{n_{a_t}}}$.
15: Insert $\widetilde{x}_t$ into Tree_{a_t}.
16: **end for**

To address MAB with heavy-tailed in the non-private case, Bubeck et al. (2013) presented a general near-optimal framework called robust-UCB by combining the UCB algorithm with several robust mean estimators. Specifically, the framework first truncates the rewards to some bounded value and then performs a robust version of UCB. Building upon the framework, we first design a method for DOP heavy-tailed MAB based on the above non-private robust-UCB algorithm, see Algorithm 1 for details.

The key idea of our algorithm is that, in the first K rounds, we establish a tree instance Tree_a for each arm $a \in [K]$ (step 4-9). After that, at round t, when the arm a_t is pulled, we truncate the newly generated reward by a certain range $B_{n_{a_t}}$ and insert the truncated reward to Tree_{a_t} (step 10-15). Here we use a robust version of UCB to select the arm, where the sum of rewards is given by the Tree-based mechanism (since we only insert the truncated rewards, we can use the mechanism). We note that in the original Tree-based mechanism in (Chan et al., 2011; Dwork et al., 2010) each element in the data steam is bounded by a uniform constant B. However, here the bound $B_{n_{a_t}}$ is adaptive and non-decreasing. Thus, we need a finer tree mechanism. To this end, we propose an adaptive Tree-based

Algorithm 2 (Adaptive) Tree-based Mechanism

Input: time horizon T, privacy budget ϵ, a stream σ.
Output: A private version $\widehat{S}(t)$ for $S(t) = \sum_{i=1}^{t} \sigma(i)$ at each $t \in [T]$

Initialize each p-sum α_i and noisy p-sum $\widehat{\alpha}_i$ to 0.
$\epsilon' \leftarrow \epsilon / \log T$.
for $t = 1, \ldots, T$ **do**
  Express t in binary form: $t = \sum_j \mathrm{Bin}_j(t) \cdot 2^j$.
  $i \leftarrow \min\{j : \mathrm{Bin}_j(t) \neq 0\}$.
  $\alpha_i \leftarrow \sum_{j<i} \alpha_j + \sigma(t)$.
  for $j = 0, \ldots, i-1$ **do**
    $\alpha_j \leftarrow 0$, $\widehat{\alpha}_j \leftarrow 0$.
  end for
  $\widehat{\alpha}_i \leftarrow \alpha_i + \mathrm{Lap}(2B_t/\epsilon')$.
  return $\hat{S}(t) \leftarrow \sum_{j:\mathrm{Bin}_j(t)=1} \widehat{\alpha}_j$.
end for

mechanism based on the earlier works (Chan et al., 2011; Dwork et al., 2009), whose procedures are presented in Algorithm 2. By the same proof as in (Chan et al., 2011; Dwork et al., 2009) we have the following guarantees as shown in Lemma 2.

Definition 4 (p-sum)**.** A p-sum is a partial sum of consecutive data items. Let $1 \le i \le j$. For a data stream σ of length T, we use $\sigma(t)$ to denote the data item at time $t \in [T]$ and $\sum[i, j] \triangleq \sum_{k=i}^{j} \sigma(k)$ to denote a partial sum involving data items i through j. We use the notation α_i^t to denote the p-sum $\sum[t - 2^i + 1, t]$.

Lemma 2 ((Adaptive) Tree-based Mechanism)**.** Given a stream σ such that $\sigma(t) \in [-B_t, B_t]$ for $\forall t \in [T]$, where B_t is non-decreasing with t, we want to privately and continually release the sum of the stream $S(t) \triangleq \sum_{i=1}^{t} \sigma(i)$ for each $t \in [T]$. Tree-based Mechanism (Algorithm 2) outputs an estimation $\widehat{S}(t)$ for $S(t)$ at each $t \in [T]$ such that $\widehat{S}(t)$ preserves ϵ-differential privacy and guarantees the following noise bound with probability at least $1 - \delta$ for any $\delta > 0$,

$$\left|\widehat{S}(t) - S(t)\right| \le \frac{2B_t}{\epsilon} \cdot \log^{1.5} T \cdot \log \frac{1}{\delta}. \tag{3}$$

When $B_t = B$, Algorithm 2 will be the same as the original one. Theorem 1 presents the privacy guarantee of overall algorithm (Algorithm 1 and Algorithm 2).

Theorem 1. For any $\epsilon > 0$, the overall algorithm (Algorithm 1 and Algorithm 2) is ϵ-differentially private.

In fact, the $\widehat{S}_a(t)/n_a$ term in step 12, which is denoted by $\widehat{\mu}_a(n_a, t)$, could be seen as a robust and private estimator of the mean μ_a after total n_a pulls of arm a till time t. Our selection strategy in step 12 is based on the following estimation error between $\widehat{\mu}_a(n_a, t)$ and μ_a, which is also a key lemma that will be used to bound the regret of Algorithm 1.

Lemma 3. In Algorithm 1, for a fixed arm a and t, we have the following estimation error with probability at least $1 - t^{-4}$,

$$\widehat{\mu}_a(n_a, t) \le \mu_a + 18u^{\frac{1}{1+v}} \left(\frac{\log(2t^4) \log^{1.5+\frac{1}{v}} T}{n_a \epsilon}\right)^{\frac{v}{1+v}}. \tag{4}$$

We have the following instance-dependent regret bound by the proof of Theorem 1 in (Auer et al., 2002).

Theorem 2. Under our assumptions, for any $0 < \epsilon \le 1$ the instance-dependent expected regret of Algorithm 1 satisfies

$$\mathcal{R}_T \le O\left(\sum_{a:\Delta_a>0} \left(\frac{\log^{2.5+\frac{1}{v}} T}{\epsilon} \left(\frac{u}{\Delta_a}\right)^{\frac{1}{v}} + \Delta_a\right)\right). \tag{5}$$

Compared with the non-private version of robust UCB (Bubeck et al., 2013), the main difference is the threshold value $B_{n_{a_t}} = (\frac{\epsilon u n_{a_t}}{\log^{1.5} T})^{\frac{1}{1+v}}$, where Bubeck et al. (2013) set it as $(\frac{u n_{a_t}}{\log(t^2)})^{\frac{1}{1+v}}$. Informally speaking, this is caused by the fact that, due to the privacy, the number of efficient samples now becomes $n\epsilon$. Specifically, due to privacy constraint, the estimation error could be decomposed into three parts: the bias, variance due to the truncation, and the noise we added. We can show that setting $B_{n_{a_t}}$ as the threshold could provide an improved bound of error. Compared with the $O(\sum_{a:\Delta_a>0} [\log T (\frac{u}{\Delta_a})^{\frac{1}{v}} + \Delta_a])$ optimal rate of the regret in the non-private version (Bubeck et al., 2013), we can see that there is an additional factor of $\frac{\log^{1.5+\frac{1}{v}} T}{\epsilon}$ in the private case. In contrast, in the problem where the reward distributions are bounded, it has been shown by Shariff and Sheffet (2018) that, there is only an additional factor of $\frac{1}{\epsilon}$ compared with the non-private case. Thus, a natural question arises here is *whether it is possible to further improve the regret.* We answer this question affirmatively by designing an optimal algorithm, see Algorithm 3 for details.

Our algorithm is based on the Successive Elimination (SE) algorithm proposed by Even-Dar et al. (2006), whose DP variant has been studied by Sajed and Sheffet (2019). Briefly speaking, we first set all the arms as viable options (step 1), then in each epoch we pull all the viable arms to get the same (private) confidence interval around their empirical rewards (step 4-18). Finally we eliminate the arms with lower empirical rewards from the viable options if they are sub-optimal compared with other viable arms (step 21-25).

Theorem 3. For any $\epsilon > 0$, Algorithm 3 is ϵ-differentially private.

Remark 1. As mentioned earlier, Sajed and Sheffet (2019) also studied a DP variant of the SE algorithm.

Algorithm 3 DP Robust Successive Elimination

Input: confidence β, parameters ϵ, v, u.
$\mathcal{S} \leftarrow \{1, \cdots, K\}$
Initialize: $t \leftarrow 0$, $\tau \leftarrow 0$.
repeat
  $\tau \leftarrow \tau + 1$.
  Set $\bar{\mu}_a = 0$ for all $a \in \mathcal{S}$.
  $r \leftarrow 0$, $D_\tau \leftarrow 2^{-\tau}$.
  $R_\tau \leftarrow \left\lceil u^{\frac{1}{v}} \left(\frac{24^{(1+v)/v} \log(4|\mathcal{S}|\tau^2/\beta)}{\epsilon D_\tau^{(1+v)/v}}\right) + 1 \right\rceil$.
  $B_\tau \leftarrow \left(\frac{uR_\tau \epsilon}{\log(4|\mathcal{S}|\tau^2/\beta)}\right)^{1/(1+v)}$.
  while $r < R_\tau$ **do**
    $r \leftarrow r + 1$.
    for $a \in \mathcal{S}$ **do**
      $t \leftarrow t + 1$.
      Sample a reward $x_{a,r}$.
      $\widetilde{x}_{a,r} \leftarrow x_{a,r} \cdot \mathbb{I}_{\{|x_{a,r}| \le B_\tau\}}$.
    end for
  end while
  For each $a \in \mathcal{S}$, compute $\bar{\mu}_a \leftarrow \left(\sum_{l=1}^{R_\tau} \widetilde{x}_{a,l}\right)/R_\tau$.
  Set $\widetilde{\mu}_a \leftarrow \bar{\mu}_a + \mathrm{Lap}\left(\frac{2B_\tau}{R_\tau \epsilon}\right)$ for all $a \in \mathcal{S}$.
  $\widetilde{\mu}_{\max} \leftarrow \max_{a \in \mathcal{S}} \widetilde{\mu}_a$.
  $err_\tau \leftarrow u^{1/(1+v)} \left(\frac{\log(4|\mathcal{S}|\tau^2/\beta)}{R_\tau \epsilon}\right)^{v/(1+v)}$.
  for all viable arm a **do**
    if $\widetilde{\mu}_{\max} - \widetilde{\mu}_a > 12 err_\tau$ **then**
      Remove arm a from $\mathcal{S}$.
    end if
  end for
until $|\mathcal{S}| = 1$
Pull the arm in $\mathcal{S}$ in all remaining $T - t$ rounds.

However, there are several differences between their result and ours. Their algorithm is only for bounded reward distributions, while here we focus on the heavy-tailed ones. Due to the irregularity of rewards, we need to preprocess and shrink these rewards. Moreover, the forms of parameters are also more complicated than the bounded distributions case. Finally, in the later section, we also extend Algorithm 3 to the local model and show its optimality.

The following lemma claims that the number of rounds to pull each arm a is at most $\widetilde{O}\left(\frac{1}{\epsilon(\Delta_a)^{\frac{1+v}{v}}}\right)$.

Lemma 4. For any instance of the K-MAB problem, denote by a^* its optimal arm. Fix the time horizon T and confidence level $\beta \in (0, 1)$. Then, with probability at least $1 - \beta$, in Algorithm 3, the total number of rounds to pull each sub-optimal arm $a \neq a^*$ is at most

$$\min\left\{T, O\left(\frac{u^{\frac{1}{1+v}}}{\epsilon(\Delta_a)^{\frac{1+v}{v}}}\left(\log\left(\frac{K}{\beta}\right) + \log\log\left(\frac{1}{\Delta_a}\right)\right)\right)\right\}. \tag{6}$$

Theorem 4 (DP Upper Bound)**.** If we set $\beta = \frac{1}{T}$ in Algorithm 3, then for sufficiently large T and any $\epsilon \in (0, 1]$, the instance-dependent expected regret of Algorithm 3 satisfies

$$\mathcal{R}_T \le O\left(\frac{u^{\frac{1}{1+v}} \log T}{\epsilon} \sum_{\Delta_a > 0} \left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}} + \max_a \Delta_a\right). \tag{7}$$

Moreover, the instance-independent expected regret of Algorithm 3 satisfies

$$\mathcal{R}_T \le O\left(u^{\frac{v}{(1+v)^2}} \left(\frac{K \log T}{\epsilon}\right)^{\frac{v}{1+v}} T^{\frac{1}{1+v}}\right), \tag{8}$$

where the $O(\cdot)$-notation omits $\log\log \frac{1}{\Delta_a}$ terms.

From Theorem 4 we can see that compared with the regret bound $O(\frac{\log^{2.5} T}{\epsilon} \sum_{\Delta_a > 0} (\frac{1}{\Delta_a})^{\frac{1}{v}})$ in Theorem 2, we achieve an improved bound of $O(\frac{\log T}{\epsilon} \sum_{\Delta_a > 0} (\frac{1}{\Delta_a})^{\frac{1}{v}})$. We think the main reason for the improvement is that the UCB-based method needs to make a reward-dependent choice in each round whereas the SE-based method only makes the reward-dependent choices in $K - 1$ special rounds when it performs arm elimination. Moreover, we also have an instance-independent regret bound. While in the bounded rewards case it has been shown that a DP variant of the SE algorithm is optimal (Sajed and Sheffet, 2019), it is still unknown whether Algorithm 3 is optimal in the heavy-tailed case. In the following we study the lower bound of regret for heavy-tailed MAB problem in the ϵ-DP model. We start from the two-armed instance-dependent regret lower bound which is specified in Theorem 10 to show that the dependency on the term of $\frac{1}{\epsilon}(\frac{1}{\Delta})^{\frac{1}{v}}$ is unavoidable in general. Due to space limitation, we put Theorem 10 and its proof in Appendix. Then we extend to the K-arm case and show the instance-dependent regret bound presented in Theorem 4 is **optimal**. The lower bound of the instance-independent regret is still unclear, and we leave it as an open problem.

Theorem 5 (DP Instance-dependent Lower Bound)**.** There exists a heavy-tailed K-armed bandit instance with $u \le 1$ in (2), $\mu_a \le \frac{1}{6}$ and $\Delta_a \in (0, \frac{1}{12})$, such that for any ϵ-DP ($0 < \epsilon \le 1$) algorithm $\mathcal{A}$ whose expected regret is at most $T^{\frac{3}{4}}$, we have

$$\mathcal{R}_T \ge \Omega\left(\frac{\log T}{\epsilon} \sum_{\Delta_a > 0} \left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}}\right). \tag{9}$$

Below we will sketch the proof of Theorem 5. Notably, previous hard instances in the bounded rewards case cannot provide tight lower bounds in our setting. Moreover, the hard instance in the non-private heavy-tailed MAB case (Bubeck et al., 2013) is only for two-armed setting and cannot be generalized to

K arms. Therefore, it is necessary to construct new hard instance to prove Theorem 5. Specifically, we design the following heavy-tailed MAB problem as a hard instance: the instance $\bar{P}$ where the distribution of each arm $a \in [K]$ is $\nu_a = \left(1 - \frac{s_a^{1+v}}{2}\right)\delta_0 + \frac{s_a^{1+v}}{2}\delta_{1/s_a}$, with some $\frac{1}{2} \geq \mu_1 \geq \cdots \geq \mu_K$ and $s_a = (2\mu_a)^{\frac{1}{v}}$. That is when $X \sim \nu_a$, $X = 0$ with probability $(1 - \frac{s_a^{1+v}}{2})$, and $X = \frac{1}{s_a}$ with probability $\frac{s_a^{1+v}}{2}$. It is easy to verify for each $a \in [K]$ that $\mathbb{E}[\nu_a] = \mu_a$ and the $(1+v)$-th moment is bounded by 1. Now, we fix an arm $a \neq 1$ and denote $\mathcal{E}_a$ as the event that the arm a is pulled at most $t_a \triangleq \frac{\log T}{100\epsilon 4^{1/v}}(\frac{1}{\Delta_a})^{\frac{1+v}{v}}$ times. We will mainly show that $\mathbb{P}_{\mathcal{A},\bar{P}}(\mathcal{E}_a) \leq \frac{1}{2K}$. To prove this, we consider another instance $\bar{Q}_a$ where the distribution of any arm $a' \neq a$ remains unchanged, and the distribution of arm a is $\nu'_a = [1 - (\frac{s_a^{1+v}}{2} + 2\Delta_a\gamma_a)]\delta_0 + (\frac{s_a^{1+v}}{2})\delta_{1/s_a} + (2\Delta_a\gamma_a)\delta_{1/\gamma_a}$, where $\gamma_a = (4\Delta_a)^{\frac{1}{v}}$. Note that when setting $\mu_a^{1+v} \leq \frac{1}{6}$ and $\Delta_a^{1+v} \leq \frac{1}{12}$, we have $\frac{s_a^{1+v}}{2} + 2\Delta_a\gamma_a = 2^{\frac{1}{v}}\mu_a^{\frac{1+v}{v}} + 2 \cdot 4^{\frac{1}{v}} \cdot \Delta_a^{\frac{1+v}{v}} \leq (\frac{1}{3})^{\frac{1}{v}} + 2 \cdot (\frac{1}{3})^{\frac{1}{v}} \leq 1$, thus the postulated ν'_a is reasonable. It is also easy to verify that $\mathbb{E}[\nu'_a] = \mu_a + 2\Delta_a = \mu_1 + \Delta_a \leq 1$ and the $(1+v)$-th moment of v'_a is bounded by 1. With the coupling lemma in Karwa and Vadhan (2018), we can prove that $\mathbb{P}_{\mathcal{A},\bar{P}}(E_a) \leq \frac{1}{2K}$. Thus, by taking the union bound we can get with probability at least $\frac{1}{2}$, the regret of $\bar{P}$ for any private algorithm $\mathcal{A}$ is $\Omega(\frac{\log T}{\epsilon}\sum_{\Delta_a>0}(\frac{1}{\Delta_a})^{\frac{1}{v}})$.

Remark 2. In the MAB with bounded rewards case, it has been shown that the optimal rate of the expected rate is $O(\frac{K\log T}{\epsilon} + \sum_{\Delta_a>0}\frac{\log T}{\Delta_a})$ (Sajed and Sheffet, 2019). Compared with the optimal rate $O(\frac{\log T}{\epsilon}\sum_{\Delta_a>0}(\frac{1}{\Delta_a})^{\frac{1}{v}})$ in the heavy-tailed case, we can see there is a huge difference. First, the dependency on $\frac{1}{\Delta_a}$ now becomes to $(\frac{1}{\Delta_a})^{\frac{1}{v}}$. Secondly, the price of privacy in the bounded rewards case is an additional term of $O(\frac{K\log T}{\epsilon})$ compared with the non-private rate, while in the heavy-tailed case, there is an additional factor of $\frac{1}{\epsilon}$ compared with the non-private one.

5 LDP HEAVY-TAILED MAB

In this section, we will investigate upper and lower bounds for heavy-tailed MAB in the *local* DP model. We start from the upper bounds. To design an ϵ-LDP algorithm, most of the previous works on MAB with bounded rewards modifies the UCB algorithm and uses the Laplacian mechanism to guarantee the LDP property, such as (Chen et al., 2020; Zhou and Tan, 2021). For sub-Gaussian rewards, Ren et al. (2020) first map the unbounded rewards with a Sigmoid function and then use the Laplacian mechanism and UCB algorithm. Moreover, it has been shown that this type of method is near-optimal (Chen et al., 2020; Ren et al., 2020). Thus, for MAB with heavy-tailed rewards, a straightforward way is to modify the UCB-based method. Specifically, each reward will be shrunken to a certain range and then added Laplacian noise, subsequently the algorithm uses the confidence bound on these perturbed rewards to pull an arm. However, such an approach may cause enormous error. The reason is that, similar to Algorithm 1, here the threshold of each reward depends on n_a. That is, the Laplacian noise we added for each reward will be proportional to n_a. As a result, unlike the Tree-based mechanism in the central model, this LDP version of the UCB algorithm will introduce a huge amount of error to estimate the mean.

To achieve a better utility, we propose an ϵ-LDP version of the SE algorithm, see Algorithm 4 for details. The basic idea is similar to Algorithm 3, where the algorithm now maintains (private) confidence interval for each arm via the perturbed rewards instead of the noisy average. However, compared with the above LDP version of the UCB algorithm, we can see that here the Laplacian noise added to each reward is *independent* on the number of rounds n_a, which could be much smaller than the noise added in the LDP version of UCB method when T is sufficiently large. The following two theorems provide the privacy and utility guarantees for Algorithm 4, respectively.

Theorem 6. For any $\epsilon > 0$, Algorithm 4 is ϵ-local differentially private.

Theorem 7 (LDP Upper Bound)**.** Set $\beta = \frac{1}{T}$ in Algorithm 4. For any $\epsilon \in (0, 1]$ and sufficiently large T, the instance-dependent expected regret of Algorithm 4 satisfies

$$\mathcal{R}_T \leq O\left(\frac{u^{\frac{2}{v}}\log T}{\epsilon^2}\sum_{\Delta_a>0}\left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}} + \max_a \Delta_a\right). \quad (10)$$

Moreover, the instance-independent expected regret of Algorithm 4 satisfies

$$\mathcal{R}_T \leq O\left(u^{\frac{2}{1+v}}\left(\frac{K\log T}{\epsilon^2}\right)^{\frac{v}{1+v}} T^{\frac{1}{1+v}}\right), \quad (11)$$

where the $O(\cdot)$-notations omit $\log\log\frac{1}{\Delta_a}$ terms.

In the following, we derive both instance-dependent and instance-independent lower bounds for heavy-tailed MAB in the ϵ-LDP model. Similar to instance-dependent lower bounds in central DP, we also first analyze the two-armed case to show that the dependency on the term of $\frac{1}{\epsilon^2}(\frac{1}{\Delta})^{\frac{1}{v}}$ is unavoidable in general (see Theorem 11 in Appendix C) and then extend to K-armed case in the following result.

Theorem 8 (LDP Instance-dependent Lower Bound)**.** There exists a heavy-tailed K-armed bandit instance with $u \leq 1$ in (2) and $\Delta_a \triangleq \mu_1 - \mu_a \in (0, \frac{1}{5})$, such

Algorithm 4 LDP Robust Successive Elimination

Input: Confidence β, parameters ϵ, v, u.
$\mathcal{S} \leftarrow \{1, \cdots, K\}$
Initialize: $t \leftarrow 0$, $\tau \leftarrow 0$.
repeat
  $\tau \leftarrow \tau + 1$.
  Set $\bar{\mu}_a = 0$ for all $a \in \mathcal{S}$.
  $r \leftarrow 0$, $D_\tau \leftarrow 4^{-\tau}$.
  $R_\tau \leftarrow \left\lceil u^{\frac{2}{v}} \left(\frac{28^{2(1+v)/v} \log(8|\mathcal{S}|\tau^2/\beta)}{\epsilon^2 D_\tau^{2(1+v)/v}}\right) + \log\left(\frac{8|\mathcal{S}|\tau^2}{\beta}\right)\right\rceil$.
  $B_\tau \leftarrow \left(\frac{u\sqrt{R_\tau}\epsilon}{\sqrt{\log(8|\mathcal{S}|\tau^2/\beta)}}\right)^{1/(1+v)}$.
  while $r < R_\tau$ **do**
    $r \leftarrow r + 1$.
    for $a \in \mathcal{S}$ **do**
      $t \leftarrow t + 1$.
      Sample a reward $x_{a,r}$ for each arm $a \in \mathcal{S}$.
      $\widetilde{x}_{a,r} \leftarrow x_{a,r} \cdot \mathbb{I}_{\{|x_{a,r}| \le B_\tau\}}$.
      $\widehat{x}_{a,r} \leftarrow \widetilde{x}_{a,r} + \mathrm{Lap}\left(\frac{2B_\tau}{\epsilon}\right)$
    end for
  end while
  For each $a \in \mathcal{S}$, compute $\bar{\mu}_a \leftarrow \left(\sum_{l=1}^{R_\tau} \widehat{x}_{a,l}\right)/R_\tau$.
  $\bar{\mu}_{\max} \leftarrow \max_{a \in \mathcal{S}} \bar{\mu}_i$.
  $err_\tau \leftarrow u^{1/(1+v)} \left(\frac{\sqrt{\log(8|\mathcal{S}|\tau^2/\beta)}}{R_\tau \epsilon}\right)^{v/(1+v)}$.
  for all viable arm a **do**
    if $\widetilde{\mu}_{\max} - \widetilde{\mu}_a > 14 err_\tau$ **then**
      Remove arm a from $\mathcal{S}$.
    end if
  end for
until $|\mathcal{S}| = 1$
Pull the arm in $\mathcal{S}$ in all remaining $T - t$ rounds.

that for any ϵ-LDP ($0 < \epsilon \le 1$) algorithm whose regret $\le o(T^\alpha)$ for any $\alpha > 0$, the regret satisfies

$$\liminf_{T\to\infty} \frac{\mathcal{R}_T}{\log T} \ge \Omega\left(\frac{1}{\epsilon^2} \sum_{\Delta_a > 0} \left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}}\right).$$

Remark 3. Theorem 11 reveals that the term of $\frac{1}{\epsilon^2 \Delta^{\frac{1}{v}}}$ is unavoidable in the regret bound. As a result, the attained bound in Theorem 7 is optimal. Compared with the optimal rate $O(\frac{1}{\Delta^{\frac{1}{v}}})$ in the non-private case, we can see the price of privacy is an additional factor of $\frac{1}{\epsilon^2}$, which is similar to other MAB with bounded/sub-Gaussian rewards problems in the LDP model (Zhou and Tan, 2021; Ren et al., 2020).

Theorem 9 (LDP Instance-independent Lower Bound)**.** There exists a heavy-tailed K-armed bandit instance with the $(1+v)$-th bounded moment of each reward distribution is bounded by 1. Moreover, if T is large enough, for any the ϵ-LDP algorithm $\mathcal{A}$ with $\epsilon \in (0, 1]$, the expected regret must satisfy

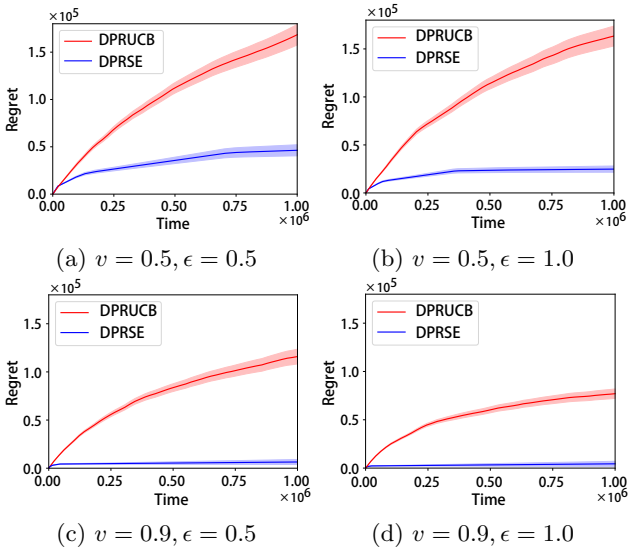
$$\mathcal{R}_T \ge \Omega\left(\left(\frac{K}{\epsilon^2}\right)^{\frac{v}{1+v}} T^{\frac{1}{1+v}}\right).$$

Remark 4. From Theorem 9, we can see the upper bound (11) of Algorithm 4 is nearly optimal. However, compared with instance-independent lower bound, there is still a poly($\log T$) factor gap. We conjecture this factor could be removed by using some more advanced robust estimator, such as the estimator in Lee et al. (2020) and we will leave it as an open problem. For MAB with bounded rewards in the LDP model, Basu et al. (2019) shows that its instance-dependent regret bound is always at least $\Omega(\frac{\sqrt{KT}}{\epsilon})$, *i.e.,* there is an additional factor of $\frac{1}{\epsilon}$ compared with the non-private case. However, for heavy-tailed MAB, compared with the lower bound of $\Omega(K^{\frac{v}{1+v}} T^{\frac{1}{1+v}})$ in the non-private case, from Theorem 9 we can observe that the difference is a factor of $(\frac{1}{\epsilon^2})^{\frac{v}{1+v}}$. Thus, combining with Remark 2, we can conclude that heavy-tailed MAB and bounded MAB are quite different in both central and local differential privacy models.

6 EXPERIMENTS

In this section, we conduct experiments on synthetic datasets to evaluate the performance of our algorithms. Since this is the first paper studying DP/LDP heavy-tailed MAB and there is no previous methods, we will only evaluate the performance of our algorithms. Our experiments consist of two parts. In the first part, we empirically compare our Algorithm 1 and Algorithm 3 in the central DP model. In the second part, we will evaluate our Algorithm 4 in the LDP model.

Datasets and Setting For the data generation, we follow similar settings as in the previous work on MAB with heavy-tailed rewards, such as Lee et al. (2020). Specifically, we set $K = 5$ and restrict the mean of each arm within $[0.1, 0.9]$ throughout the experiment. We consider three instances, denoted by S_1, S_2 and S_3. In S_1, we let the gaps of sub-optimal arms decrease linearly where the largest mean is always 0.9 and the smallest mean is always 0.1 (so the means are {0.9, 0.7, 0.5, 0.3, 0.1}). In S_2, we consider the case that a larger fraction of arms have large sub-optimal gaps, hence we set the mean of each arm a by a quadratic convex function $\mu_a = 0.05(a-5)^2 + 0.1$ (so the means are {0.9, 0.55, 0.3, 0.15, 0.1}). In S_3, we consider the case that a larger fraction of arms have small sub-optimal gaps, hence we set the mean of each arm a by a quadratic concave function $\mu_a = -0.05(a-1)^2 + 0.9$ (so the means are {0.9, 0.85, 0.7, 0.45, 0.1}). In all the settings, the reward of each arm $a \in [K]$ at each pull is drawn from a Pareto distribution with shape parameter α and scale



(a) $v = 0.5, \epsilon = 0.5$ (b) $v = 0.5, \epsilon = 1.0$

(c) $v = 0.9, \epsilon = 0.5$ (d) $v = 0.9, \epsilon = 1.0$

Figure 1: DP Setting 1 (S_1)

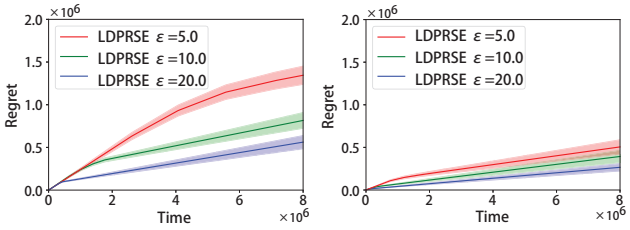
parameter λ_a. Specifically, each time after pulling arm a, the learner receives a reward x which follows the following probability density function:

$$f(x) = \begin{cases} \frac{\alpha \lambda_a^\alpha}{x^{\alpha+1}} & x \geq \lambda_a \\ 0 & x < \lambda_a. \end{cases}$$

We adopt the Pareto distribution because it is common in practice. We set $\alpha = 1.05 + v$ such that the $1 + v$-th moment of reward distribution always exists and is bounded by $\frac{\alpha \lambda_a^{1+v}}{\alpha - (1+v)}$. For a given α, we set $\lambda_a = \frac{(\alpha-1)\mu_a}{\alpha}$ since the mean of a Pareto distribution with parameters α and λ_a is $\frac{\alpha \lambda_a}{\alpha - 1}$. We take the maximum of $\frac{\alpha \lambda_a^{1+v}}{\alpha-(1+v)}$ among all arms $a \in [K]$ as u in the experiments.

For each algorithm we run 90 independent repetitions for each case. In figures, we show the average of cumulative regret (represented by the solid line) for comparing the performance of algorithms and error bars of a quarter standard deviation (represented by the shaded region) for comparing the robustness of algorithms.

Results and Discussion. In the first part, we compare the performance of our proposed Algorithm 1 (DPRUCB) and Algorithm 3 (DPRSE) for the central DP model. In each setting, we vary $\epsilon \in \{0.5, 1.0\}$ and $v \in \{0.5, 0.9\}$. The results of setting S_1 is given in Figure 1. And the results of setting S_2 and S_3 are presents in Figure 3 and Figure 4, which are included in Appendix D due to space limitation. From these results, we can see that DPRSE always outperform DPRUCB among these three settings. Moreover, we can see when ϵ decreases, the regret will increase, and when the term v becomes larger, we have smaller regret.



(a) $v = 0.5$ (b) $v = 0.9$

Figure 2: LDP Setting 3 (S_3)

In the second part, we evaluate our proposed Algorithm 4 (LDPRSE) for the local DP model in the Setting 3 (S_3), since S_3 will has larger regret than S_1 and S_2 theoretically. We vary $\epsilon \in \{5.0, 10.0, 20.0\}$ and $v \in \{0.5, 0.9\}$, The results are given in Figure 2. We can see that when ϵ or v is larger, we have smaller regret. Moreover, compared with the central DP model, we can see the regret is larger in the LDP model.

In summary, we can observe that all the above results support our previous theoretical analysis.

7 CONCLUSIONS

In this paper, we provided the first study on the problem of MAB with heavy-tailed reward distributions in the (local) Differential Privacy model. We mainly focused on the case the reward distribution of each arm only has $(1 + v)$-th moment with some $v \in (0, 1]$. In the central ϵ-DP model, we first provided a near optimal result by developing a private and robust UCB algorithm. To achieve this we provided an adaptive version of the Tree-based mechanism. Then, we improved the result via a private and robust version of the SE algorithm. Finally, we showed that the instance-dependent regret bound of our improved algorithm is optimal by showing its lower bound. In the ϵ-LDP model. We proposed an algorithm which could be seen a locally private and robust version of the SE algorithm, which provably achieve (near) optimal rates for both instance-dependent and instance-independent regret.

There are still many open problems besides the future work mentioned in the main context. First, throughout the whole paper we need to assume both u and v are known. How to address a more practical case where they are unknown? Recent work that addresses this issue for standard bandit problems (Ashutosh et al., 2021), while it remains unknown whether they can be extended to the private case. Secondly, for the setting of MAB with bounded reward, it has been shown that an UCB-based private algorithm can also attain an optimal regret guarantee. Thus, a natural question is whether it is possible to get an optimal DP variant of UCB algorithm for our problem.

Acknowledgements

Di Wang and Yulian Wu were support in part by the baseline funding BAS/1/1689-01-01 and funding from the AI Initiative REI/1/4811-10-01 of King Abdullah University of Science and Technology (KAUST). Peng Zhao was supported by the Baidu Scholarship.

References

Agrawal, S., Juneja, S., and Glynn, P. (2020). Optimal δ-correct best-arm selection for heavy-tailed distributions. In *Proceedings of the 31st International Conference on Algorithmic Learning Theory (ALT)*, pages 61–110.

Agrawal, S., Juneja, S., and Koolen, W. M. (2021). Regret minimization in heavy-tailed bandits. In *Proceedings of 34th Conference on Learning Theory (COLT)*, pages 26–62.

Ashutosh, K., Nair, J., Kagrecha, A., and Jagannathan, K. (2021). Bandit algorithms: Letting go of logarithmic regret for statistical robustness. In *Proceedings of the 24th International Conference on Artificial Intelligence and Statistics (AISTATS)*, pages 622–630.

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

Barber, R. F. and Duchi, J. C. (2014). Privacy and statistical risk: Formalisms and minimax bounds. *arXiv preprint arXiv:1412.4451*.

Basu, D., Dimitrakakis, C., and Tossou, A. (2019). Privacy in multi-armed bandits: Fundamental definitions and lower bounds. *arXiv preprint arXiv:1905.12298*.

Bradley, B. O. and Taqqu, M. S. (2003). Financial risk and heavy tails. In *Handbook of heavy tailed distributions in finance*, pages 35–103. Elsevier.

Brunel, V.-E. and Avella-Medina, M. (2020). Propose, test, release: Differentially private estimation with high probability. *arXiv preprint arXiv:2002.08774*.

Bubeck, S., Cesa-Bianchi, N., and Lugosi, G. (2013). Bandits with heavy tail. *IEEE Transactions on Information Theory*, 59(11):7711–7717.

Catoni, O. (2012). Challenging the empirical mean and empirical variance: a deviation study. In *Annales de l'IHP Probabilités et statistiques*, volume 48, pages 1148–1185.

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

Chen, X., Zheng, K., Zhou, Z., Yang, Y., Chen, W., and Wang, L. (2020). (Locally) differentially private combinatorial semi-bandits. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 1757–1767.

Dubey, A. and Pentland, A. (2019). Thompson sampling on symmetric α-stable bandits. In *Proceedings of the 28th International Joint Conference on Artificial Intelligence (IJCAI)*, pages 5715–5721.

Dubey, A. and Pentland, A. (2020a). Differentially-private federated linear bandits. In *Proceedings of the 34th Conference on Advances in Neural Information Processing Systems (NeurIPS)*.

Dubey, A. and Pentland, A. (2020b). Private and byzantine-proof cooperative decision-making. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 357–365.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Theory of Cryptography Conference (TCC)*, pages 265–284.

Dwork, C., Naor, M., Pitassi, T., and Rothblum, G. N. (2010). Differential privacy under continual observation. In *Proceedings of the 42nd Annual ACM Symposium on Theory of Computing (STOC)*, pages 715–724.

Dwork, C., Naor, M., Reingold, O., Rothblum, G. N., and Vadhan, S. P. (2009). On the complexity of differentially private data release: efficient algorithms and hardness results. In *Proceedings of the 41st Annual ACM Symposium on Theory of Computing (STOC)*, pages 381–390.

Even-Dar, E., Mannor, S., Mansour, Y., and Mahadevan, S. (2006). Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, 7(6).

Gajane, P., Urvoy, T., and Kaufmann, E. (2018). Corrupt bandits for preserving local privacy. In *Proceedings of the 29th International Conference on Algorithmic Learning Theory (ALT)*, pages 387–412.

Gutiérrez, B., Peter, L., Klein, T., and Wachinger, C. (2017). A multi-armed bandit to smartly select a training set from big medical data. In *Proceedings of the 20th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 38–45.

Hannun, A., Knott, B., Sengupta, S., and van der Maaten, L. (2019). Privacy-preserving multi-party contextual bandits. *arXiv preprint arXiv:1910.05299*.

Kalogerias, D. S., Nikolakakis, K. E., Sarwate, A. D., and Sheffet, O. (2020). Best-arm identification for quantile bandits with privacy. *arXiv preprint arXiv:2006.06792*.

Kamath, G., Singhal, V., and Ullman, J. (2020). Private mean estimation of heavy-tailed distributions. In *Proceedings of 33rd Conference on Learning Theory (COLT)*, pages 2204–2235.

Karwa, V. and Vadhan, S. P. (2018). Finite sample differentially private confidence intervals. In *Proceedings of the 9th Innovations in Theoretical Computer Science Conference (ITCS)*, pages 44:1–44:9.

Kumar, R. and Tomkins, A. (2010). A characterization of online browsing behavior. In *Proceedings of the 19th International Conference on World Wide Web (WWW)*, pages 561–570.

Lattimore, T. (2017). A scale free algorithm for stochastic bandits with bounded kurtosis. In *Proceedings of the 31st Conference on Advances in Neural Information Processing Systems (NeurIPS)*, pages 1583–1592.

Lattimore, T. and Szepesvári, C. (2020). *Bandit algorithms*. Cambridge University Press.

Lee, K., Yang, H., Lim, S., and Oh, S. (2020). Optimal algorithms for stochastic multi-armed bandits with heavy tailed rewards. In *Proceedings of the 34th Conference on Advances in Neural Information Processing Systems (NeurIPS)*.

Liu, X., Kong, W., Kakade, S., and Oh, S. (2021). Robust and differentially private mean estimation. *arXiv preprint arXiv:2102.09159*.

Mahanti, A., Carlsson, N., Mahanti, A., Arlitt, M., and Williamson, C. (2013). A tale of the tails: Power-laws in internet measurements. *IEEE Network*, 27(1):59–64.

Malekzadeh, M., Athanasakis, D., Haddadi, H., and Livshits, B. (2020). Privacy-preserving bandits. In *Proceedings of the 3rd Conference on Machine Learning and Systems (MLSys)*.

Mishra, N. and Thakurta, A. (2015). (Nearly) optimal differentially private stochastic multi-arm bandits. In *Proceedings of the 31st Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 592–601.

Nakayama, K., Hisakado, M., and Mori, S. (2017). Nash equilibrium of social-learning agents in a restless multiarmed bandit game. *Scientific reports*, 7(1):1–8.

Press, W. H. (2009). Bandit solutions provide unified ethical models for randomized clinical trials and comparative effectiveness research. *Proceedings of the National Academy of Sciences*, 106(52):22387–22392.

Ren, W., Zhou, X., Liu, J., and Shroff, N. B. (2020). Multi-armed bandits with local differential privacy. *arXiv preprint arXiv:2007.03121*.

Robbins, H. (1952). Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535.

Sajed, T. and Sheffet, O. (2019). An optimal private stochastic-mab algorithm based on optimal private stopping rule. In *Proceedings of the 36th International Conference on Machine Learning (ICML)*, pages 5579–5588.

Shariff, R. and Sheffet, O. (2018). Differentially private contextual linear bandits. *arXiv preprint arXiv:1810.00068*.

Shen, W., Wang, J., Jiang, Y.-G., and Zha, H. (2015). Portfolio choices with orthogonal bandit learning. In *Proceedings of the 24th International Joint Conference on Artificial Intelligence (IJCAI)*.

Thompson, W. R. (1933). On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294.

Tossou, A. and Dimitrakakis, C. (2016). Algorithms for differentially private multi-armed bandits. In *Proceedings of the 30th AAAI Conference on Artificial Intelligence (AAAI)*, pages 2087–2093.

Vakili, S., Liu, K., and Zhao, Q. (2013). Deterministic sequencing of exploration and exploitation for multi-armed bandit problems. *IEEE Journal of Selected Topics in Signal Processing*, 7(5):759–767.

Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.

Vietri, G., Balle, B., Krishnamurthy, A., and Wu, S. (2020). Private reinforcement learning with pac and regret guarantees. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 9754–9764.

von Bahr, B. and Esseen, C.-G. (1965). Inequalities for the rth Absolute Moment of a Sum of Random Variables, $1 \leqq r \leqq 2$. *The Annals of Mathematical Statistics*, 36(1):299 – 303.

Wang, D., Gaboardi, M., and Xu, J. (2018). Empirical risk minimization in non-interactive local differential privacy revisited. In *Proceedings of the 32nd Conference on Advances in Neural Information Processing Systems (NeurIPS)*.

Wang, D., Xiao, H., Devadas, S., and Xu, J. (2020a). On differentially private stochastic convex optimization with heavy-tailed data. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 10081–10091.

Wang, H., Zhao, Q., Wu, Q., Chopra, S., Khaitan, A., and Wang, H. (2020b). Global and local differential privacy for collaborative bandits. In *Proceedings of the 14th ACM Conference on Recommender Systems (RecSys)*, pages 150–159.

Yu, X., Shao, H., Lyu, M. R., and King, I. (2018). Pure exploration of multi-armed bandits with heavy-tailed payoffs. In *Proceedings of the 34th Conference on Uncertainty in Artificial Intelligence (UAI)*, pages 937–946.

Zheng, K., Cai, T., Huang, W., Li, Z., and Wang, L. (2020). Locally differentially private (contextual) bandits learning. In *Proceedings of the 34th Conference on Advances in Neural Information Processing Systems (NeurIPS)*, pages 12300–12310.

Zhou, X. and Tan, J. (2021). Local differential privacy for bayesian optimization. In *Proceedings of the 35th AAAI Conference on Artificial Intelligence (AAAI)*, pages 11152–11159.

Supplementary Material: Optimal Rates of (Locally) Differentially Private Heavy-tailed Multi-Armed Bandits

A TECHNICAL LEMMAS

Lemma 5 (Tail Bound of Laplacian Vairable (Dwork et al., 2006)). If $X \sim \text{Lap}(b)$, then

$$\mathbb{P}(|X| \geq t \cdot b) = \exp(-t).$$

Lemma 6 (Bernstein's Inequality (Vershynin, 2018)). Let $X_1, \cdots X_n$ be n independent zero-mean random variables. Suppose $|X_i| \leq M$ and $\mathbb{E}[X_i^2] \leq s$ for all $i \in [n]$. Then for any $t > 0$, we have

$$\mathbb{P}\left\{\frac{1}{n}\sum_{i=1}^{n} X_i \geq t\right\} \leq \exp\left(-\frac{\frac{1}{2}t^2 n}{s + \frac{1}{3}Mt}\right)$$

Lemma 7. Given a random variable X with $\mathbb{E}[|X|^{1+v}] \leq u$ for some $v \in (0, 1]$, for any $B > 0$ we have

$$\mathbb{E}\left[X \cdot \mathbb{I}_{|X|>B}\right] \leq \frac{u}{B^v}.$$

Proof [of Lemma 7] By the definition of expectation, we have

$$\begin{aligned}
u \geq \mathbb{E}[|X|^{1+v}] &= \int_0^\infty (1+v)t^{1+v-1}\mathbb{P}(|X| > t)\,\mathrm{d}t \\
&\geq \int_B^\infty t^v \mathbb{P}(|X| > t)\,\mathrm{d}t \\
&\geq B^v \int_B^\infty \mathbb{P}(|X| > t)\,\mathrm{d}t \\
&= B^v \int_0^\infty \mathbb{P}(X \cdot \mathbb{I}_{|X|>B} > t)\,\mathrm{d}t \\
&= B^v \mathbb{E}\left[X \cdot \mathbb{I}_{|X|>B}\right].
\end{aligned}$$

Rearranging the inequality finishes the proof. ■

Lemma 8 (Concentration of Laplace Variables (Wang et al., 2018)). If $X_1, \cdots X_n \sim \text{Lap}(s/\epsilon)$, then with probability at least $1 - \beta$, we have

$$\left|\frac{1}{n}\sum_{i=1}^{n} X_i\right| \leq \frac{2s}{\epsilon\sqrt{n}}\sqrt{\log\frac{2}{\beta}}.$$

Lemma 9 (Jensen's Inequality). Let X be an integrable, real-valued random variable, and ψ be a convex function. Then $\psi(\mathbb{E}[X]) \leq \mathbb{E}[\psi(X)]$.

Lemma 10 (Relation between Raw Moment and Central Moment). Let X be a random variable over $\mathbb{R}$ such that $\mathbb{E}[X] = \mu$. We have the following two results:

(1) When $\mathbb{E}[|X|^{1+v}] < \infty$ for some $v \in (0, 1]$, we have

$$\mathbb{E}[|X - \mu|^{1+v}] \leq 4\mathbb{E}[|X|^{1+v}] < \infty.$$

(2) When $\mathbb{E}[|X-\mu|^{1+v}] < \infty$ for some $v \in (0,1]$, we have

$$\mathbb{E}[|X|^{1+v}] \leq 2\mathbb{E}[|X-\mu|^{1+v}] + 2|\mu|^{1+v} < \infty.$$

Proof [of Lemma 10] When $\mathbb{E}[|X|^{1+v}] < \infty$, we have

$$\begin{aligned}
\mathbb{E}[|X-\mu|^{1+v}] &\leq \mathbb{E}[|X-\mu|^{1+v} + |X+\mu|^{1+v}] \\
&\leq \mathbb{E}[2(|X|^{1+v} + |\mu|^{1+v})] \\
&= 2\mathbb{E}[|X|^{1+v}] + 2\mathbb{E}[|\mu|^{1+v}] \\
&= 2\mathbb{E}[|X|^{1+v}] + 2|\mu|^{1+v} \\
&= 2\mathbb{E}[|X|^{1+v}] + 2|\mathbb{E}[X]|^{1+v} \\
&\leq 2\mathbb{E}[|X|^{1+v}] + 2\mathbb{E}[|X|^{1+v}] \\
&= 4\mathbb{E}[|X|^{1+v}],
\end{aligned}$$

where the second inequality is due to the inequality (6) of von Bahr and Esseen (1965), and the last inequality is due to Jensen's inequality (Lemma 9).

When $\mathbb{E}[|X-\mu|^{1+v}] < \infty$, we have

$$\begin{aligned}
\mathbb{E}[|X|^{1+v}] &\leq \mathbb{E}[|X|^{1+v}] + \mathbb{E}[|X-2\mu|^{1+v}] \\
&\leq \mathbb{E}[2(|X-\mu|^{1+v} + |\mu|^{1+v})] \\
&= 2\mathbb{E}[(|X-\mu|^{1+v})] + 2|\mu|^{1+v},
\end{aligned}$$

where the second inequality comes from the inequality (6) of von Bahr and Esseen (1965) as well. ■

B OMITTED PROOFS FOR SECTION 4 (central differential privacy)

Proof [of Lemma 2] The proof can be directly followed by the proof in the original bounded case, which is given by (Chan et al., 2011; Dwork et al., 2010). ■

Proof [of Theorem 1] Note that all rewards generated by arm a are inserted into the corresponding Tree-based Mechanism $Tree_a$ and each Tree-based Mechanism is ϵ-differentially private (by Lemma 2), the proof follows directly from the Parallel Composition Theorem of DP. ■

Proof [of Lemma 3] We will prove the following more general lemma:

Lemma 11. In Algorithm 1, for a fixed arm a and t, we have the following estimation error with probability at least $1-2\delta$ for any $\delta \geq \frac{1}{2T^4}$

$$\widehat{\mu}_a(n_a,t) \leq \mu_a + 18u^{\frac{1}{1+v}}\left(\frac{\log\frac{1}{\delta}\log^{1.5+\frac{1}{v}}T}{n_a\epsilon}\right)^{\frac{v}{1+v}}. \tag{12}$$

Thus, in Algorithm 1, we set δ to be $\frac{1}{2t^4}$. Hence with probability at least $1-\frac{1}{t^4}$, we have

$$\widehat{\mu}_a(n_a,t) \leq \mu_a + 18u^{\frac{1}{1+v}}\left(\frac{\log(2t^4)\log^{1.5+\frac{1}{v}}T}{n_a\epsilon}\right)^{\frac{v}{1+v}}.$$

In the following we will prove Lemma 11.

Denote the total noise introduced by the Tree_a by *noise* and the i-th reward obtained from arm a by $x_{a,i}$. From Bernstein's inequality for bounded random variables, noting that $\mathbb{E}(X_a^2 \cdot \mathbb{I}_{|X_a|\le B}) \le uB^{1-v}$ if $\mathbb{E}|X|^{1+v} \le u$, we have, with probability at least $1-2\delta$,

$$
\begin{aligned}
&|\widehat{\mu}_a(n_a,t) - \mu_a| \\
&= \left| \frac{1}{n_a}\sum_{i=1}^{n_a} x_{a,i}\mathbb{I}_{|x_{a,i}|\le B_i} + noise - \mu_a \right| \\
&\le \left| \frac{1}{n_a}\sum_{i=1}^{n_a} x_{a,i}\mathbb{I}_{|x_{a,i}|\le B_i} - \mu_a \right| + |noise| \\
&= \left| \frac{1}{n_a}\sum_{i=1}^{n_a} \left[x_{a,i}\mathbb{I}_{|x_{a,i}|\le B_i} - \mathbb{E}(X_a\mathbb{I}_{|X_a|\le B_i})\right] + \frac{1}{n_a}\sum_{i=1}^{n_a}\left[\mathbb{E}(X_a\mathbb{I}_{|X_a|\le B_i}) - \mathbb{E}X_a\right] \right| + |noise| \\
&\le \left| \frac{1}{n_a}\sum_{i=1}^{n_a} \left[x_{a,i}\mathbb{I}_{|x_{a,i}|\le B_i} - \mathbb{E}(X_a\mathbb{I}_{|X_a|\le B_i})\right] \right| + \left| \frac{1}{n_a}\sum_{i=1}^{n_a}\mathbb{E}(X_a\mathbb{I}_{|X_a|> B_i}) \right| + |noise| \\
&\le \sqrt{\frac{2B_{n_a}^{1-v}u\log(\frac{1}{\delta})}{n_a}} + \frac{B_{n_a}\log\frac{1}{\delta}}{3n_a} + \frac{1}{n_a}\sum_{i=1}^{n_a}\frac{u}{B_i^v} + \frac{2B_{n_a}\log\frac{1}{\delta}\log^{1.5}T}{\epsilon n_a}.
\end{aligned}
$$

Where the last inequality is due to Lemma 7 and Lemma 5, and Lemma 6 that with probability at least $1-\delta$,

$$
\left| \frac{1}{n_a}\sum_{i=1}^{n_a} \left[x_{a,i}\mathbb{I}_{|x_{a,i}|\le B_i} - \mathbb{E}(X_a\mathbb{I}_{|X_a|\le B_i})\right] \right| \le \sqrt{\frac{2B_{n_a}^{1-v}u\log(\frac{1}{\delta})}{n_a}} + \frac{B_{n_a}\log\frac{1}{\delta}}{3n_a}.
$$

Recall that $B_n = \left(\frac{\epsilon u n}{\log^{1.5}T}\right)^{\frac{1}{1+v}}$ for any $n \in \mathbb{N}^+$, with some evident calculations, we can bound each term above in the last inequality respectively.

$$
\begin{aligned}
\frac{1}{n_a}\sum_{i=1}^{n_a}\frac{u}{B_i^v} &= \frac{u}{n_a}\sum_{i=1}^{n_a}\left(\frac{\log^{1.5}T}{\epsilon u i}\right)^{\frac{v}{1+v}} \le \frac{u^{\frac{1}{1+v}}}{n_a}\frac{(\log^{1.5}T)^{\frac{v}{1+v}}}{\epsilon^{\frac{v}{1+v}}}\sum_{i=1}^{n_a} i^{-\frac{v}{1+v}} \\
&\le \frac{u^{\frac{1}{1+v}}}{n_a}\frac{(\log^{1.5}T)^{\frac{v}{1+v}}}{\epsilon^{\frac{v}{1+v}}}\cdot(1+v)\cdot n_a^{\frac{1}{1+v}} \le 2u^{\frac{1}{1+v}}\left(\frac{\log\frac{1}{\delta}\log^{1.5}T}{\epsilon n_a}\right)^{\frac{v}{1+v}}, \qquad (13) \\
\sqrt{\frac{2B_{n_a}^{1-v}u\log(\frac{1}{\delta})}{n_a}} &= \frac{\sqrt{2}u^{\frac{1}{1+v}}\epsilon^{\frac{1-v}{2(1+v)}}(\log\frac{1}{\delta})^{\frac{1}{2}}}{n_a^{\frac{v}{1+v}}(\log^{1.5}T)^{\frac{1-v}{2(1+v)}}} \le \sqrt{10}u^{\frac{1}{1+v}}\left(\frac{\log\frac{1}{\delta}\log^{1.5}T}{\epsilon n_a}\right)^{\frac{v}{1+v}}, \qquad (14) \\
\frac{B_{n_a}\log\frac{1}{\delta}}{3n_a} &\le \frac{5B_{n_a}(\log\frac{1}{\delta})^{\frac{v}{1+v}}(\log^{1.5}T)^{\frac{1}{1+v}}}{3n_a} \le \frac{5(\epsilon u n)^{\frac{1}{1+v}}(\log\frac{1}{\delta})^{\frac{v}{1+v}}(\log^{1.5}T)^{\frac{1}{1+v}}}{3(\log^{1.5}T)^{\frac{1}{1+v}}n_a} \\
&= \frac{5\epsilon^{\frac{1}{1+v}}u^{\frac{1}{1+v}}(\log\frac{1}{\delta})^{\frac{v}{1+v}}}{3n_a^{\frac{v}{1+v}}} \le 2u^{\frac{1}{1+v}}\left(\frac{\log\frac{1}{\delta}\log^{1.5}T}{\epsilon n_a}\right)^{\frac{v}{1+v}}. \qquad (15)
\end{aligned}
$$

The inequality of the second term is based on the fact that $\epsilon \le 1$ and

$$
\left(\log\frac{1}{\delta}\right)^{\frac{1}{2}-\frac{v}{1+v}} = \left(\log\frac{1}{\delta}\right)^{\frac{1-v}{2(1+v)}} \le (\log 2T^4)^{\frac{1-v}{2(1+v)}} \le (5\log T)^{\frac{1-v}{2(1+v)}} \le \sqrt{5}(\log^{1.5}T)^{\frac{1-v}{2(1+v)}}.
$$

The inequality of the third term holds due to the following fact:

$$
\frac{\log\frac{1}{\delta}}{(\log^{1.5}T)^{\frac{1}{(1+v)}}} \le \left(\frac{\log\frac{1}{\delta}}{\log^{1.5}T}\right)^{\frac{1}{(1+v)}}\left(\log\frac{1}{\delta}\right)^{\frac{v}{1+v}} \le 5\left(\log\frac{1}{\delta}\right)^{\frac{v}{1+v}}.
$$

Moreover we have with probability at least $1-\delta$

$$\begin{aligned}
&\frac{2B_{n_a}\log\frac{1}{\delta}\log^{1.5}T}{\epsilon n_a} = \frac{2u^{\frac{1}{1+v}}(\log\frac{1}{\delta}\log^{1.5}T)^{\frac{v}{1+v}}}{(\epsilon n_a)^{\frac{v}{1+v}}}(\log\frac{1}{\delta})^{\frac{1}{1+v}} \\
&\leq \frac{10u^{\frac{1}{1+v}}(\log\frac{1}{\delta}\log^{1.5}T)^{\frac{v}{1+v}}}{(\epsilon n_a)^{\frac{v}{1+v}}}\log^{\frac{v}{1+v}}T \leq \frac{10u^{\frac{1}{1+v}}(\log\frac{1}{\delta}\log^{1.5}T)^{\frac{v}{1+v}}}{(\epsilon n_a)^{\frac{v}{1+v}}}\log^{\frac{1}{1+v}}T \\
&\leq \frac{10u^{\frac{1}{1+v}}(\log\frac{1}{\delta}\log^{1.5+\frac{1}{v}}T)^{\frac{v}{1+v}}}{(\epsilon n_a)^{\frac{v}{1+v}}}. \qquad (16)
\end{aligned}$$

Based on (13), (14), (15), (16), we obtain

$$|\widehat{\mu}_a(n_a,t)-\mu_a| \leq 18u^{\frac{1}{1+v}}\left(\frac{\log\frac{1}{\delta}\log^{1.5+\frac{1}{v}}T}{\epsilon n_a}\right)^{\frac{v}{1+v}},$$

■

Proof [of Theorem 2] We denote by $N_a(t)$ the (random) number of times arm a is selected up to time t and let $I_{s,t} = 18u^{\frac{1}{1+v}}\left(\frac{\log[2(t+1)^4]\log^{1.5+\frac{1}{v}}T}{\epsilon s}\right)^{\frac{v}{1+v}}$. We first derive the upper bound on $\mathbb{E}[N_a(T)]$ for any arm a. Let ℓ be an arbitrary positive integer.

$$\begin{aligned}
N_a(T) &= 1+\sum_{t=K+1}^{T}\mathbb{I}_{a_t=a} \\
&\leq \ell+\sum_{t=K+1}^{T}\mathbb{I}_{a_t=a \text{ and } N_a(t-1)\geq\ell} \\
&\leq \ell+\sum_{t=K+1}^{T}\mathbb{I}_{\hat{\mu}_{a^*}(n_{a^*},t-1)+I_{N_{a^*}(t-1),t-1}\leq\hat{\mu}_a(n_a,t-1)+I_{N_a(t-1),t-1} \text{ and } N_a(t-1)\geq\ell} \\
&\leq \ell+\sum_{t=K+1}^{T}\mathbb{I}_{\min_{0<s<t}\hat{\mu}_{a^*}(s,t-1)+I_{s,t-1}\leq\max_{\ell<s_a<t}\hat{\mu}_a(s_a,t-1)+I_{s_a,t-1}} \\
&\leq \ell+\sum_{t=1}^{\infty}\sum_{s=1}^{t}\sum_{s_a=\ell}^{t}\mathbb{I}_{\hat{\mu}_{a^*}(s,t)+I_{s,t}\leq\hat{\mu}_a(s_a,t)+I_{s_a,t}}.
\end{aligned}$$

Note that $\hat{\mu}_{a^*}(s,t)+I_{s,t}\leq\hat{\mu}_a(s_a,t)+I_{s_a,t}$ implies that at least one of the following three inequalities is true:

$$\hat{\mu}_{a^*}(s,t) \leq \mu_{a^*}-I_{s,t} \qquad (17)$$
$$\hat{\mu}_a(s_a,t) \geq \mu_a+I_{s_a,t} \qquad (18)$$
$$\Delta_a < 2I_{s_a,t} \qquad (19)$$

Otherwise, assume that all three inequalities are false, then we have

$$\begin{aligned}
\hat{\mu}_{a^*}(s,t)+I_{s,t} &> \mu_{a^*} \\
&= \mu_a+\Delta_a \\
&> \hat{\mu}_a(s_a,t)-I_{s_a,t}+\Delta_a \\
&\geq \hat{\mu}_a(s_a,t)-I_{s_a,t}+2I_{s_a,t} \\
&= \hat{\mu}_a(s_a,t)+I_{s_a,t},
\end{aligned}$$

which contradicts the condition. By the result of Lemma 3, we know that (17) or (18) hold with probability at most $2t^{-4}$. For $s_a \geq 36u^{\frac{1}{v}}\frac{\log(2t^4)\log^{1.5+\frac{1}{v}}T}{\epsilon\Delta_a^{\frac{1+v}{v}}}$, (19) is false. Now, let $\ell = \left\lceil 36u^{\frac{1}{v}}\frac{\log(2t^4)\log^{1.5+\frac{1}{v}}T}{\epsilon\Delta_a^{\frac{1+v}{v}}}\right\rceil$, we can bound the

term $\mathbb{E}[N_a(T)]$ as follows,

$$\begin{aligned}
\mathbb{E}[N_a(T)] &\leq \left\lceil 36u^{\frac{1}{v}} \frac{\log(2t^4)\log^{1.5+\frac{1}{v}} T}{\epsilon \Delta_a^{\frac{1+v}{v}}} \right\rceil + \sum_{t=1}^{\infty}\sum_{s=1}^{t}\sum_{s_a=\ell}^{t} \mathbb{P}((17) \text{ or } (18) \text{ is true}) \\
&\leq \left\lceil 36u^{\frac{1}{v}} \frac{\log(2t^4)\log^{1.5+\frac{1}{v}} T}{\epsilon \Delta_a^{\frac{1+v}{v}}} \right\rceil + \sum_{t=1}^{\infty}\sum_{s=1}^{t}\sum_{s_a=\ell}^{t} \frac{2}{t^4} \\
&\leq \left\lceil 36u^{\frac{1}{v}} \frac{\log(2t^4)\log^{1.5+\frac{1}{v}} T}{\epsilon \Delta_a^{\frac{1+v}{v}}} \right\rceil + 2\sum_{t=1}^{\infty} \frac{1}{t^2} \\
&\leq 36u^{\frac{1}{v}} \frac{\log(2t^4)\log^{1.5+\frac{1}{v}} T}{\epsilon \Delta_a^{\frac{1+v}{v}}} + 1 + \frac{\pi^2}{3} \\
&\leq 36u^{\frac{1}{v}} \frac{\log(2t^4)\log^{1.5+\frac{1}{v}} T}{\epsilon \Delta_a^{\frac{1+v}{v}}} + 5.
\end{aligned}$$

Finally, using that $\mathcal{R}_T = \sum_{a=1}^{K}(\Delta_a \mathbb{E}N_a(T))$, we directly obtain

$$\mathcal{R}_T \leq \sum_{a=1}^{K}\left(36\frac{\log(2T^4)\log^{1.5+\frac{1}{v}} T}{\epsilon}\left(\frac{u}{\Delta_a}\right)^{\frac{1}{v}} + 5\Delta_a\right)$$

■

Proof [of Theorem 3] Consider two adjacent reward streams that differ only on one reward of arm a. In each epoch τ, the difference of the mean of arm a between the two adjacent streams is at most $\frac{2B_\tau}{R_\tau}$ since the reward of each arm is truncated by $[-B_\tau, B_\tau]$. Thus, adding noise of $\text{Lap}(\frac{2B_\tau}{\epsilon R_\tau})$ to μ_a guarantees ϵ-DP. ■

Proof [of Lemma 4] The bound of T is trivial so we focus on proving the latter bound. Denote the i-th reward obtained from arm a by $x_{a,i}$. We first bound $|\mu_a - \bar{\mu}_a|$ and $|\bar{\mu}_a - \widetilde{\mu}_a|$ for each epoch τ and each $a \in \mathcal{S}$. For $|\mu_a - \bar{\mu}_a|$, by Lemma 6, Lemma 7 and noting that $\mathbb{E}(X^2\mathbb{I}_{|X|\leq B}) \leq uB^{1-v}$, we have, with probability at least $1 - \frac{\beta}{4|\mathcal{S}|\tau^2}$,

$$\begin{aligned}
&|\mu_a - \bar{\mu}_a| \\
&= \left|\frac{1}{R_\tau}\sum_{t=1}^{R_\tau} x_{a,t}\mathbb{I}_{|x_{a,t}|\leq B_\tau} - \mu_a\right| \\
&= \left|\frac{1}{R_\tau}\sum_{t=1}^{R_\tau}\left[x_{a,t}\mathbb{I}_{|x_{a,t}|\leq B_\tau} - \mathbb{E}(x_{a,t}\mathbb{I}_{|x_{a,t}|\leq B_\tau})\right] + \frac{1}{R_\tau}\sum_{t=1}^{R_\tau}\left[\mathbb{E}(x_{a,t}\mathbb{I}_{|x_{a,t}|\leq B_\tau}) - \mathbb{E}x_{a,t}\right]\right| \\
&\leq \left|\frac{1}{R_\tau}\sum_{t=1}^{R_\tau}\left[x_{a,t}\mathbb{I}_{|x_{a,t}|\leq B_\tau} - \mathbb{E}(x_{a,t}\mathbb{I}_{|x_{a,t}|\leq B_\tau})\right]\right| + \left|\frac{1}{R_\tau}\sum_{t=1}^{R_\tau}\mathbb{E}(x_{a,t}\mathbb{I}_{|x_{a,t}|> B_\tau})\right| \\
&\leq \sqrt{\frac{2B_\tau^{1-v}u\log\left(\frac{4|\mathcal{S}|\tau^2}{\beta}\right)}{R_\tau}} + \frac{B_\tau\log\left(\frac{4|\mathcal{S}|\tau^2}{\beta}\right)}{3R_\tau} + \frac{u}{B_\tau^v} \\
&\leq 4u^{\frac{1}{1+v}}\left(\frac{\log\left(\frac{4|\mathcal{S}|\tau^2}{\beta}\right)}{R_\tau\epsilon}\right)^{\frac{v}{1+v}} = 4err_\tau.
\end{aligned}$$

For $|\bar{\mu}_a - \widetilde{\mu}_a|$, by using the concentration of the Laplace distribution Lemma 5, we have with probability at least

$1-\frac{\beta}{4|\mathcal{S}|\tau^2}$,

$$|\bar{\mu}_a-\widetilde{\mu}_a|=\left|\mathrm{Lap}\left(\frac{2B_\tau}{\epsilon R_\tau}\right)\right|\le\frac{2B_\tau}{\epsilon R_\tau}\log\frac{4|\mathcal{S}|\tau^2}{\beta}=2u^{\frac{1}{1+v}}\left(\frac{\log\left(\frac{4|\mathcal{S}|\tau^2}{\beta}\right)}{R_\tau\epsilon}\right)^{\frac{v}{1+v}}=2err_\tau.$$

Given an epoch τ, we denote by $\mathcal{E}_\tau$ the event where for all $i\in\mathcal{S}$ it holds that $|\mu_a-\bar{\mu}_a|\le 4err_\tau$ and $|\bar{\mu}_a-\widetilde{\mu}_a|\le 2err_\tau$ and denote $\mathcal{E}=\bigcup_{\tau\ge1}\mathcal{E}_\tau$. By taking the union bound, we have

$$\mathbb{P}(\mathcal{E}_\tau)\ge 1-\frac{\beta}{2\tau^2}$$

and

$$\mathbb{P}(\mathcal{E})\ge 1-\frac{\beta}{2}\left(\sum_{\tau\ge1}\tau^{-2}\right)\ge 1-\beta.$$

In the remainder of the proof, we assume that $\mathcal{E}$ holds. So for any epoch τ and any viable arm a, we have $|\widetilde{\mu}_a-\mu_a|\le 6err_\tau$. As a result, for any epoch τ and any two arms a and a', we have

$$|(\widetilde{\mu}_a-\widetilde{\mu}_{a'})-(\mu_a-\mu_{a'})|\le 12err_\tau.$$

Next, we show that under $\mathcal{E}$, the optimal arm a^* is never eliminated. For any epoch τ, let $a_\tau=\arg\max_{a\in\mathcal{S}}\widetilde{\mu}_a$. Since

$$|(\widetilde{\mu}_{a_\tau}-\widetilde{\mu}_{a^*})-(\mu_{a_\tau}-\mu_{a^*})|=|(\widetilde{\mu}_{a_\tau}-\widetilde{\mu}_{a^*})+\Delta_{a_\tau}|=(\widetilde{\mu}_{a_\tau}-\widetilde{\mu}_{a^*})+\Delta_{a_\tau}\le 12err_\tau,$$

it is easy to see that the algorithm doesn't eliminate a^*.

Next, we show that under $\mathcal{E}$, in any epoch τ the algorithm eliminate all viable arms with sub-optimality gap at least $D_\tau=2^{-\tau}$. Fix an epoch τ and a viable arm a with sup-optimality gap $\Delta_a\ge D_\tau$. Due to $R_\tau=u^{\frac{1}{v}}\left(\frac{24^{(1+v)/v}\log\left(4|\mathcal{S}|\tau^2/\beta\right)}{\epsilon D_\tau^{(1+v)/v}}\right)+1$, we know that $12err_\tau<\frac{D_\tau}{2}$. Thus,

$$\widetilde{\mu}_{a_\tau}-\widetilde{\mu}_a\ge\widetilde{\mu}_{a^*}-\widetilde{\mu}_a\ge\Delta_a-12err_\tau>D_\tau-\frac{D_\tau}{2}=\frac{D_\tau}{2}>12err_\tau,$$

which means arm a is eliminated by the algorithm.

Finally, fix a suboptimal arm a, we derive the upper bound on the total number of timesteps that arm a is pulled. Let $\tau(a)$ be the first epoch such that $\Delta_a\ge D_{\tau(a)}$, which implying $D_{\tau(a)}\le\Delta_a\le D_{\tau(a)-1}=2D_{\tau(a)}$. Due to $\Delta_a\le 2D_{\tau(a)}$ and $D_{\tau(a)}=2^{-\tau(a)}$, we have $\tau(a)\le\log_2\left(\frac{2}{\Delta_a}\right)$. Thus, the total number of pulls of arm a is

$$\begin{aligned}
\sum_{\tau\le\tau(a)}R_\tau&\le\sum_{\tau\le\tau(a)}2^{\frac{1+v}{v}[\tau-\tau(a)]}R_{\tau(a)}\le R_{\tau(a)}\sum_{i=0}^{\tau(a)-1}2^{-\frac{1+v}{v}i}\\
&\le\frac{1}{1-2^{-\frac{1+v}{v}}}R_{\tau(a)}\\
&=\frac{2^{\frac{1+v}{v}}}{2^{\frac{1+v}{v}}-1}\left(u^{\frac{1}{1+v}}\frac{24^{\frac{1+v}{v}}}{D_\tau^{\frac{1+v}{v}}}\frac{\log\left(\frac{4|\mathcal{S}|\tau(a)^2}{\beta}\right)}{\epsilon}+1\right)\\
&\le u^{\frac{1}{1+v}}\frac{48^{\frac{1+v}{v}}}{D_\tau^{\frac{1+v}{v}}}\frac{\log\left(\frac{4|\mathcal{S}|\tau(a)^2}{\beta}\right)}{\epsilon}+2\\
&\le u^{\frac{1}{1+v}}\left(\frac{96}{\Delta_a}\right)^{\frac{1+v}{v}}\frac{\log\left(\frac{4K}{\beta}\right)+\log\log\left(\frac{2}{\Delta_a}\right)}{\epsilon}+2\\
&=O\left(u^{\frac{1}{1+v}}\left(\frac{1}{\Delta_a}\right)^{\frac{1+v}{v}}\frac{\log\left(\frac{4K}{\beta}\right)+\log\log\left(\frac{2}{\Delta_a}\right)}{\epsilon}\right)
\end{aligned}$$

where the last inequality is due to the bounds $D_\tau > \Delta_a/2$, $|\mathcal{S}| \le K$, $\tau(a) \le \log_2(2/\Delta_a)$ and $K \le 2$. Combining with the trivial upper bound T, the lemma follows directly. ■

Proof [of Theorem 4] We first consider the instance-dependent expected regret. Denote the total number of rounds to pull each sub-optimal arm $a \neq a^*$ by T_a. Taking $\beta = \frac{1}{T}$, then with probability at least $1 - \frac{1}{T}$, we have

$$\begin{aligned}\mathcal{R}_T &\le \sum_{\Delta_a>0} T_a \cdot \Delta_a = \sum_{\Delta_i>0} \left[\frac{u^{\frac{1}{1+v}}}{\epsilon(\Delta_a)^{\frac{1}{v}}} \left(\log\left(\frac{K}{\beta}\right) + \log\log\left(\frac{1}{\Delta_a}\right)\right)\right] \\ &= O\left(\frac{u^{\frac{1}{1+v}} \log T}{\epsilon} \sum_{\Delta_a>0} \left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}} \right).\end{aligned}$$

With probability at most $\frac{1}{T}$, Algorithm 3 will fail to identify the optimal arm and thus incur an expected cumulative *regret* of $O(T \cdot \max_{\Delta_a>0} \Delta_a)$. Combining the two cases, we obtain

$$\begin{aligned}\mathcal{R}_T &\le (1 - \frac{1}{T}) \cdot O\left[\frac{u^{\frac{1}{1+v}} \log T}{\epsilon} \sum_{\Delta_a>0} \left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}} \right] + \frac{1}{T} \cdot O(T \cdot \max_{\Delta_a>0} \Delta_a) \\ &\le \widetilde{O}\left(\frac{u^{\frac{1}{1+v}} \log T}{\epsilon} \sum_{\Delta_a>0} \left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}} + \max_{\Delta_a>0} \Delta_a \right).\end{aligned}$$

We now consider the instance-independent expected regret, which is inspired by (Sajed and Sheffet, 2019).

Throughout the proof we assume Algorithm 3 runs with a parameter $\beta = \frac{1}{T}$. Since any arm a with $\Delta_a < \frac{1}{T}$ yields a negligible expected regret bound of at most 1, we assume $\Delta_a \ge \frac{1}{T}$. Then the bound of Lemma 4 becomes

$$\min\left\{ T, C \frac{u^{\frac{1}{1+v}} \log(KT)}{\epsilon} \left(\frac{1}{\Delta_a}\right)^{\frac{1+v}{v}} \right\}.$$

It follows that for any suboptimal arm a, the expected regret from pulling arm a is therefore at most

$$\min\left\{ \Delta_a T, C \frac{u^{\frac{1}{1+v}} \log(KT)}{\epsilon} \left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}} \right\}.$$

Denote by Δ^* the gap which equates the two possible regret bounds when all arms are pulled T/K times. That is, $\Delta^* \frac{T}{K} = C \frac{u^{\frac{1}{1+v}} \log T}{\epsilon} \left(\frac{1}{\Delta^*}\right)^{\frac{1}{v}}$. It can be easily derived that

$$\Delta^* = O\left(u^{\frac{v}{(1+v)^2}} \left(\frac{CK}{\epsilon} \frac{\log T}{T} \right)^{\frac{v}{1+v}} \right).$$

Note that in the setting where all suboptimal arms have the gap of precisely Δ^*, the expected regret bound is proportional to

$$O\left(u^{\frac{v}{(1+v)^2}} \left(\frac{CK \log T}{\epsilon} \right)^{\frac{v}{1+v}} T^{\frac{1}{1+v}} \right).$$

Next, we show that no matter how different the arm gaps are, the expected regret of Algorithm 3 is still proportional to this bound. For an arbitrary MAB instance, we rearrange arm by the increasing order of arm gaps such that arm 1 is the optimal arm. We partition the set of suboptimal arms $2, \cdots, K$ to two sets: $\{2, \cdots, k'\}$ and $\{k'+1, \cdots K\}$, where k' is the largest index of the arm with gap at most Δ^*. If we only pull the arms in the former set, the upper bound on the expected regret will be $T\Delta^*$, since the incurred expected regret at each pull

is at most Δ^*. If we only pull the arms in the latter set, the expected regret will be at most

$$C\frac{u^{\frac{1}{1+v}}\log(KT)}{\epsilon}\sum_{a=k'+1}^{K}\left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}} \le 2C\frac{u^{\frac{1}{1+v}}\log T}{\epsilon}\sum_{a=k'+1}^{K}\left(\frac{1}{\Delta^*}\right)^{\frac{1}{v}}$$
$$= 2(K-k')\cdot C\frac{u^{\frac{1}{1+v}}\log T}{\epsilon}\left(\frac{1}{\Delta^*}\right)^{\frac{1}{v}} = 2(K-k')\Delta^*\frac{T}{K} \le 2T\Delta^*.$$

Since $\{2,\cdots,k'\}$ and $\{k'+1,\cdots K\}$ is a partition of the subtoptimal arms, one of the two sets contributes at least half of the expected regret. It is simple to see that the expected regret is upper bounded by $O(T\Delta^*) = O\left(u^{\frac{v}{(1+v)^2}}\left(\frac{CK\log T}{\epsilon}\right)^{\frac{v}{1+v}}T^{\frac{1}{1+v}}\right)$. ■

Theorem 10. There exists a heavy-tailed two-armed bandit problem with arm 2 being sub-optimal, $u \le 1$ in (2), and $\Delta \triangleq \mu_1 - \mu_2 \in (0, \frac{1}{5})$. Such that for any ϵ-DP algorithm $\mathcal{A}$ with expected regret at most $T^{\frac{3}{4}}$ for sufficiently large number of rounds T,* we have

$$\mathcal{R}_T \ge \Omega\left(\frac{\log T}{\epsilon}(\frac{1}{\Delta})^{\frac{1}{v}}\right). \tag{20}$$

Proof [of Theorem 10] Let $\gamma = (5\Delta)^{\frac{1}{v}}$. Consider the instance $\bar{P}$ where the distribution of arm 1 is

$$\nu_1 = \left(1 - \frac{\gamma^{1+v}}{2}\right)\delta_0 + \frac{\gamma^{1+v}}{2}\delta_{1/\gamma}$$

and the distribution of arm 2 is

$$\nu_2 = \left[1 - \left(\frac{\gamma^{1+v}}{2} - \Delta\gamma\right)\right]\delta_0 + \left(\frac{\gamma^{1+v}}{2} - \Delta\gamma\right)\delta_{1/\gamma},$$

where δ_x is the Dirac distribution on x and the distribution $p\cdot\delta_x + (1-p)\cdot\delta_y$ takes the value x with probability p and the value y with probability $1-p$. It is easy to verify that

$$\mathbb{E}[\nu_1] = \frac{5}{2}\Delta, u(\nu_1) = \frac{1}{2} < 1$$

and

$$\mathbb{E}[\nu_2] = \frac{3}{2}\Delta, u(\nu_2) = \frac{3}{10} \le 1.$$

Denote $\mathcal{E}$ as the event that arm 2 is pulled at most $t_2 \triangleq \frac{\log T}{100\epsilon\cdot 5^{1/v}}\left(\frac{1}{\Delta}\right)^{\frac{1+v}{v}}$ times. We show in the following that $\mathbb{P}_{\mathcal{A},\bar{P}}(\mathcal{E}) \le \frac{1}{2}$. Consider another instance $\bar{Q}$ where the distribution of arm 1 remains unchanged, the distribution of arm 2 is

$$\nu_2' = \left[1 - \left(\frac{\gamma^{1+v}}{2} + \Delta\gamma\right)\right]\delta_0 + \left(\frac{\gamma^{1+v}}{2} + \Delta\gamma\right)\delta_{1/\gamma}.$$

Note that since $\Delta \in (0, \frac{1}{5})$, $(\frac{\gamma^{1+v}}{2} + \Delta\gamma) < 1$, hence the instance $\bar{Q}$ is reasonable. It is also easy to verify that

$$\mathbb{E}[\nu_2'] = \frac{7}{2}\Delta, u(\nu_2') = \frac{7}{10} \le 1.$$

Denote the regret of algorithm $\mathcal{A}$ under the instance $\bar{Q}$ by $\mathcal{R}^{\mathcal{A}}_{T,\bar{Q}}$. Then we have

$$\mathcal{R}^{\mathcal{A}}_{T,\bar{Q}} \ge \mathbb{P}_{\mathcal{A},\bar{Q}}(\mathcal{E})(T-t_2)\Delta \ge \frac{T\Delta}{2}\mathbb{P}_{\mathcal{A},\bar{Q}}(\mathcal{E}),$$

where the first inequality is since that when $\mathcal{E}$ holds, we have additional $(T-t_2)\Delta$ regret and the second inequality is since that we assume T is sufficiently large. Recall that $\mathcal{R}^{\mathcal{A}}_T$ is at most $T^{\frac{3}{4}}$, then we can obtain that

$$\mathbb{P}_{\mathcal{A},\bar{Q}}(\mathcal{E}) \le \frac{2}{\Delta T^{\frac{1}{4}}}.$$

Now we consider the influence of differential privacy. Before that, we recall the following lemma.

*Note that we can replace $\frac{3}{4}$ to other constants. The same to other results.

Lemma 12 (Lemma 6.1 in (Karwa and Vadhan, 2018)). For each pair of distribution $\mathcal{D}$ and $\mathcal{D}'$ and any ϵ-differentially private mechanism $\mathcal{M}$, let $\mathbb{P}$ and $\mathbb{P}'$ be the two marginal distributions on the output of $\mathcal{M}$ evaluated on n data sampled i.i.d. from $\mathcal{D}$ and $\mathcal{D}'$ respectively, then for any event E, we have

$$\mathbb{P}[E] \leq e^{6\epsilon n \cdot d_{TV}(\mathcal{D},\mathcal{D}')}\mathbb{P}'[E], \tag{21}$$

where $d_{TV}(\mathcal{D},\mathcal{D}')$ is the total-variation distance between $\mathcal{D}$ and $\mathcal{D}'$.

Lemma 12 suggests that the "effective" group privacy for the case that n data items of the inputs are drawn i.i.d. either from distribution $\mathcal{D}$ or from distribution $\mathcal{D}'$ is proportional to $\exp(6\epsilon n \cdot d_{TV}(\mathcal{D},\mathcal{D}'))$. We apply the coupling argument in (Karwa and Vadhan, 2018) to our setting. Note that we only consider the change under the event $\mathcal{E}$ here. Suppose there is an oracle $\mathcal{O}$ that can generate a collection of at most t_2 pairs of data, where the left ones are i.i.d. samples from ν_2 and the right ones are i.i.d. samples from ν_2'. Whenever the algorithm needs to sample a reward from arm 2, it turns to the oracle $\mathcal{O}$ and $\mathcal{O}$ provides either a fresh left-sample or a right-sample depending on the true environment (the true reward distribution of arm 2). Suppose there is a counter $\mathcal{C}$ standing between the algorithm $\mathcal{A}$ and the Oracle $\mathcal{O}$. And if $\mathcal{O}$ runs out of t_2 samples, $\mathcal{C}$ routes $\mathcal{A}$'s oracle calls to another oracle. Lemma 12 guarantees that, the oracle never runs out of t_2 samples, i.e. event E happens, with similar probabilities under ν_2 and ν_2'. Formally, using the result of Lemma 12, for sufficiently large T such that $T^{0.13} > \frac{4}{\Delta}$, we have

$$\begin{aligned}\mathbb{P}_{\mathcal{A},\bar{P}}(E) &\leq e^{6\epsilon t_2 d_{TV}\left(\nu_2,\nu_2'\right)} \cdot \mathbb{P}_{\mathcal{A},\bar{Q}}(E) \leq e^{6\epsilon t_2 \cdot 2\Delta\gamma} \cdot \frac{2}{\Delta T^{\frac{1}{4}}} \\ &\leq e^{\frac{12}{100}\log T} \cdot \frac{2}{\Delta} T^{-\frac{1}{4}} = T^{-0.13}\frac{2}{\Delta} < \frac{1}{2}.\end{aligned}$$

Thus we obtain

$$\mathcal{R}_T \geq \Omega\left(\Delta \cdot t_2\right) \geq \Omega\left(\frac{\log T}{\epsilon}\left(\frac{1}{\Delta}\right)^{\frac{1}{v}}\right).$$

■

Proof [of Theorem 5] We focus on the K arms with mean reward satisfying $\frac{1}{2} \geq \mu_1 \geq \cdots \geq \mu_K$. Consider the instance $\bar{P}$ where the distribution for each arm $a \in [K]$ is

$$\nu_a = \left(1 - \frac{s_a^{1+v}}{2}\right)\delta_0 + \frac{s_a^{1+v}}{2}\delta_{1/s_a},$$

where $s_a = (2\mu_a)^{\frac{1}{v}}$. It is easy to verify for each $a \in [K]$ that

$$\mathbb{E}[\nu_a] = \mu_a, u(\nu_a) = \frac{1}{2} < 1.$$

Now, we fix an arm $a \neq 1$ and denote $\mathcal{E}_a$ as the event that the arm a is pulled at most $t_a \triangleq \frac{\log T}{100\epsilon 4^{1/v}}\left(\frac{1}{\Delta_a}\right)^{\frac{1+v}{v}}$ times. We show in the following that $\mathbb{P}_{\mathcal{A},\bar{P}}(\mathcal{E}_a) \leq \frac{1}{2K}$. Consider another instance $\bar{Q}_a$ where the distribution of any arm $a' \neq a$ remains unchanged, and the distribution of arm a is

$$\nu_a' = \left[1 - \left(\frac{s_a^{1+v}}{2} + 2\Delta_a\gamma_a\right)\right]\delta_0 + \left(\frac{s_a^{1+v}}{2}\right)\delta_{1/s_a} + (2\Delta_a\gamma_a)\delta_{1/\gamma_a},$$

where $\gamma_a = (4\Delta_a)^{\frac{1}{v}}$. Note that since $\mu_a \leq \frac{1}{6}$ and $\Delta_a \leq \frac{1}{12}$, we have $\mu_a^{1+v} < \frac{1}{6}$ and $\Delta_a^{1+v} < \frac{1}{12}$, and then

$$\frac{s_a^{1+v}}{2} + 2\Delta_a\gamma_a = 2^{\frac{1}{v}}\mu_a^{\frac{1+v}{v}} + 2 \cdot 4^{\frac{1}{v}} \cdot \Delta_a^{\frac{1+v}{v}} < \left(\frac{1}{3}\right)^{\frac{1}{v}} + 2 \cdot \left(\frac{1}{3}\right)^{\frac{1}{v}} < 1,$$

thus the postulated ν_a' is reasonable. It is also easy to verify that

$$\mathbb{E}[\nu_a'] = \mu_a + 2\Delta_a = \mu_1 + \Delta_a, u(v_a') = 1.$$

Then for sufficiently large T we have

$$\mathcal{R}^{\mathcal{A}}_{T,\bar{Q}_a} \geq \mathbb{P}_{\mathcal{A},\bar{Q}_a}(\mathcal{E}_a) \cdot (T - t_a) \cdot \Delta_a \geq \frac{T\Delta_a}{2}\mathbb{P}_{\mathcal{A},\bar{Q}_a}(\mathcal{E}_a).$$

Combining with $\mathcal{R}^{\mathcal{A}}_T \leq T^{\frac{3}{4}}$, we have

$$\mathbb{P}_{\mathcal{A},\bar{Q}_a}(\mathcal{E}_a) \leq \frac{2}{\Delta_a \cdot T^{\frac{1}{4}}}.$$

By lemma 12, for sufficiently large T such that $T^{0.13} > \max_{\Delta_a>0} \frac{4K}{\Delta_a}$, we have

$$\begin{aligned}\mathbb{P}_{\mathcal{A},\bar{P}}(\mathcal{E}_a) &\leq e^{6\epsilon t_a d_{TV}\left(\nu_a,\nu'_a\right)} \cdot \mathbb{P}_{\mathcal{A},\bar{Q}_a}(\mathcal{E}_a) \leq e^{6\epsilon t_a \cdot 2\Delta_a\gamma_a} \cdot \frac{2}{\Delta_a T^{\frac{1}{4}}} \\ &\leq e^{\frac{12}{100}\log T} \cdot \frac{2}{\Delta_a} T^{-\frac{1}{4}} = T^{-0.13}\frac{2}{\Delta_a} < \frac{1}{2K}.\end{aligned}$$

Then, with probability at least $1 - K \cdot \frac{1}{2K} = \frac{1}{2}$, $\mathcal{A}$ will pull each $a \neq 1$ at least t_a times. Thus we obtain

$$\mathcal{R}_T \geq \Omega\left(\sum_{\Delta_a>0} \Delta_a \cdot t_a\right) \geq \Omega\left(\frac{\log T}{\epsilon} \sum_{\Delta_a>0}\left(\frac{1}{\Delta_a}\right)^{\frac{1}{v}}\right).$$

■

C OMITTED PROOFS FOR SECTION 5 (local differential privacy)

Proof [of Theorem 6] Since each $|\widetilde{x}_{a,r}|$ is bounded by B_r. Thus, adding noise of $\mathrm{Lap}(\frac{B_r}{\epsilon})$ to $\widetilde{x}_{a,r}$ guarantees ϵ-LDP. ■

Proof [of Theorem 7] Similar to Lemma 3, we prove the following lemma.

Lemma 13. For any instance of the K-armed MAB problem, denote by a^* the optimal arm and by Δ_a the gap between the mean of arm a^* and any sub-optimal arm $a \neq a^*$. Fix the time horizon T and confidence level $\beta \in (0,1)$. Then, with probability at least $1-\beta$, in Algorithm 4, the total number of rounds to pull each sub-optimal arm $a \neq a^*$, denoted by T_a, is at most

$$\min\left\{T, O\left(\frac{u^{\frac{2}{v}}}{\epsilon^2(\Delta_a)^{\frac{1+v}{v}}}\left(\log\left(\frac{K}{\beta}\right) + \log\log\left(\frac{1}{\Delta_a}\right)\right)\right)\right\}. \tag{22}$$

We first bound the error of $|\mu_a - \widetilde{\mu}_a|$ for each epoch τ and each arm a. Recall that

$$|\widetilde{\mu}_a - \mu_a| = \left|\frac{\sum_{i=1}^{R_\tau}\widetilde{x}_{a,i}}{R_\tau} - \mu_a\right| + \left|\frac{\sum_{i=1}^{R_\tau} Y_{a,i}}{R_\tau}\right| \tag{23}$$

where $Y_{a,i} \sim \mathrm{Lap}\left(2\frac{B_\tau}{\epsilon}\right)$ and $B_\tau = \left(\frac{u\sqrt{R_\tau}\epsilon}{\sqrt{\log(8|\mathcal{S}|\tau^2/\beta)}}\right)^{\frac{1}{1+v}}$. According to Heoffding bound (Lemma 8), we can get with probability $1-\delta$,

$$\left|\frac{\sum_{i=1}^{R_\tau} Y_{a,i}}{R_\tau}\right| \leq \frac{4\sqrt{\log\frac{2}{\delta}}}{\sqrt{R_\tau}\epsilon}\left(\frac{u\sqrt{R_\tau}\epsilon}{\sqrt{\log\left(8\left|\mathcal{S}\right|\tau^2/\beta\right)}}\right)^{\frac{1}{1+v}}.$$

Setting $\delta = \frac{\beta}{4|\mathcal{S}|e^2}$, we have

$$\left|\frac{\sum_{i=1}^{R_\tau} Y_{a,i}}{R_\tau}\right| \leq 4u^{\frac{1}{1+v}}\left(\frac{\sqrt{\log\left(8\left|\mathcal{S}\right|\tau^2/\beta\right)}}{\sqrt{R_\tau}\epsilon}\right)^{\frac{v}{1+v}}.$$

Now we consider the first term on the right side of Equation (23). From Lemma 6, Lemma 7, noting that $\mathbb{E}\left(X^2\mathbb{I}_{|X|\leq B}\right) \leq uB^{1-v}$, we have, with probability at least $1-\delta$

$$\begin{aligned}
&\mu_a - \frac{1}{R_\tau}\sum_{i=1}^{R_\tau}\widetilde{x}_{a,i} \\
&= \frac{1}{R_\tau}\sum_{i=1}^{R_\tau}\left(\mu_a - \mathbb{E}\left(X\cdot\mathbb{I}_{|X|\leq B_\tau}\right)\right) + \frac{1}{R_\tau}\sum_{i=1}^{R_\tau}\left(\mathbb{E}\left(X\cdot\mathbb{I}_{|X|\leq B_\tau}\right) - x_{a,i}\mathbb{I}_{|x_{a,i}|\leq B_\tau}\right) \\
&= \frac{1}{R_\tau}\sum_{i=1}^{R_\tau}\mathbb{E}\left(X\cdot\mathbb{I}_{|X|> B_\tau}\right) + \frac{1}{R_\tau}\sum_{i=1}^{R_\tau}\left(\mathbb{E}\left(X\cdot\mathbb{I}_{|X|\leq B_\tau}\right) - x_{a,i}\mathbb{I}_{|x_{a,i}|\leq B_\tau}\right) \\
&\leq \frac{u}{B_\tau^v} + \sqrt{\frac{2B_\tau^{1-v}u\log\left(\delta^{-1}\right)}{R_\tau}} + \frac{B_\tau\log\left(\delta^{-1}\right)}{3R_\tau}.
\end{aligned}$$

Taking $\delta = \frac{\beta}{8|\mathcal{S}|\tau^2}$, we have

$$\frac{u}{B_\tau^v} \leq u^{\frac{1}{1+v}}\left(\frac{\sqrt{\log\left(8\left|\mathcal{S}\right|\tau^2/\beta\right)}}{\sqrt{R_\tau}\epsilon}\right)^{\frac{v}{1+v}},$$

$$\sqrt{\frac{2B_\tau^{1-v}u\log\left(\delta^{-1}\right)}{R_\tau}} \leq \sqrt{\frac{2B_\tau^{1-v}u\log\left(\delta^{-1}\right)}{\sqrt{R_\tau}}}\sqrt{\frac{\sqrt{\log\frac{1}{\delta}}}{\sqrt{R_\tau}}} \leq \sqrt{2}u^{\frac{1}{1+v}}\left(\frac{\sqrt{\log\left(8\left|\mathcal{S}\right|\tau^2/\beta\right)}}{\sqrt{R_\tau}\epsilon}\right)^{\frac{v}{1+v}},$$

where the last inequality is due to the fact that $R_\tau \geq \log\left(8\left|\mathcal{S}\right|\tau^2/\beta\right)$. Moreover we have

$$\frac{B_\tau\log\left(\delta^{-1}\right)}{3R_\tau} \leq \frac{B_\tau\sqrt{\log\left(\delta^{-1}\right)}}{3\sqrt{R_\tau}}\frac{\sqrt{\log\left(\delta^{-1}\right)}}{\sqrt{R_\tau}} \leq \frac{1}{3}u^{\frac{1}{1+v}}\left(\frac{\sqrt{\log\left(8\left|\mathcal{S}\right|\tau^2/\beta\right)}}{\sqrt{R_\tau}\epsilon}\right)^{\frac{v}{1+v}}.$$

Thus, in total we have $\left|\frac{\sum_{i=1}^{R_\tau}\widetilde{X}_{a,i}}{R_\tau} - \mu_k\right| \leq 3u^{\frac{1}{1+v}}\left(\frac{\sqrt{\log(8|\mathcal{S}|\tau^2/\beta)}}{\sqrt{R_\tau}\epsilon}\right)^{\frac{v}{1+v}}$. Taking union bound yields that with probability at least $1-\frac{3\beta}{8|\mathcal{S}|\tau^2}$,

$$|\widetilde{\mu}_a - \mu_a| \leq 7u^{\frac{1}{1+v}}\left(\frac{\sqrt{\log\left(8\left|\mathcal{S}\right|\tau^2/\beta\right)}}{\sqrt{R_\tau}\epsilon}\right)^{\frac{v}{1+v}}. \tag{24}$$

Denote by $\mathcal{E}_\tau$ the event where for all arms $a\in\mathcal{S}$, (24) holds and denote $\mathcal{E} = \cup_{\tau\geq 1}\mathcal{E}_\tau$. Taking the union for all epochs and arms, we have $\mathcal{E}$ holds w.p. $1-\frac{3\beta}{8}\sum_{\tau\geq 1}\tau^{-2} \geq 1-\beta$. As a result, for any epoch τ and any two arms $i,j\in\mathcal{S}$ we have that

$$|(\widetilde{\mu}_i - \widetilde{\mu}_j) - (\mu_i - \mu_j)| \leq 14u^{\frac{1}{1+v}}\left(\frac{\sqrt{\log\left(8\left|\mathcal{S}\right|\tau^2/\beta\right)}}{\sqrt{R_\tau}\epsilon}\right)^{\frac{v}{1+v}}.$$

Next, we show that under $\mathcal{E}$, in each epoch, the optimal arm a^* will not be eliminated. Let $a_\tau = \mathrm{argmax}_{a\in\mathcal{S}}\widetilde{\mu}_a$, then in the epoch τ,

$$\widetilde{\mu}_{a_\tau} - \widetilde{\mu}_{a^*} + \Delta_{a_\tau} \leq 14u^{\frac{1}{1+v}}\left(\frac{\sqrt{\log\left(8\left|\mathcal{S}\right|\tau^2/\beta\right)}}{\sqrt{R_\tau}\epsilon}\right)^{\frac{v}{1+v}},$$

so the algorithm doesn't eliminate μ^*.

Next, we show that under $\mathcal{E}$, in each epoch τ, we eliminate all arms with sub-optimality gap $\geq 2^{-\tau} = D_\tau$. Fix a sub-optimal arm a such that $\Delta_a \geq D_\tau$. By the definition of R_τ, We know that $14u^{\frac{1}{1+v}}\left(\frac{\sqrt{\log(8|\mathcal{S}|\tau^2/\beta)}}{\sqrt{R_\tau}\epsilon}\right)^{\frac{v}{1+v}} \leq \frac{D_\tau}{2}$.

Therefore, since arm a^* remains viable, we have that

$$\begin{aligned}\widetilde{\mu}_{a_\tau} - \widetilde{\mu}_a \geq \widetilde{\mu}_{a^*} - \widetilde{\mu}_a &\geq \Delta_a - 14u^{\frac{1}{1+v}}\left(\frac{\sqrt{\log\left(8\left|\mathcal{S}\right|\tau^2/\beta\right)}}{\sqrt{R_\tau}\epsilon}\right)^{\frac{v}{1+v}} \\ &\geq \Delta_\tau - \frac{\Delta_\tau}{2} \geq \frac{\Delta_\tau}{2} \geq 14u^{\frac{1}{1+v}}\left(\frac{\sqrt{\log\left(8\left|\mathcal{S}\right|\tau^2/\beta\right)}}{\sqrt{R_\tau}\epsilon}\right)^{\frac{v}{1+v}},\end{aligned}$$

which ensures that arm a is removed from $\mathcal{S}$.

Lastly, for any fixed sub-optimal arm a, let $\tau(a)$ be the first epoch s.t. $\Delta_a \geq D^2_{\tau(a)}$, implying $D^2_{\tau(a)} \leq \Delta_a < D^2_{\tau(a)-1} = 4D^2_{\tau(a)}$. For any epoch τ, we have $R_{\tau+1} \geq 4^{-\frac{1+v}{v}}R_\tau$, we have that the total number of pulls of arm a is

$$\begin{aligned}\sum_{\tau\leq\tau(a)} R_\tau &\leq \sum_{\tau\leq\tau(a)}\left(4^{-\frac{1+v}{v}}\right)^{\tau-\tau(a)} R_{\tau(a)} \\ &\leq R_{\tau(a)}\sum_{i\geq 0}\left(4^{-\frac{1+v}{v}}\right)^i \\ &\leq \frac{1}{1-4^{-\frac{1+v}{v}}}\left(\frac{28^{\frac{2(1+v)}{v}}\log\left(8\left|\mathcal{S}\right|\tau^2(a)/\beta\right)}{\epsilon^2 D^{\frac{2(1+v)}{v}}_{\tau(a)}}u^{\frac{2}{v}} + \log\left(8\left|\mathcal{S}\right|\tau^2(a)/\beta\right)\right) \\ &\leq O\left(\frac{\log\left(\left|\mathcal{S}\right|\tau^2(a)/\beta\right)}{\epsilon^2\Delta_a^{\frac{1+v}{v}}}u^{\frac{2}{v}}\right)\end{aligned}$$

Note that $\tau(a) = O(\log\frac{1}{\Delta_a})$, thus the algorithm pulls sub-optimal arm a for a number of timesteps is bounded by $O\left(\frac{\left(\log T+\log\log\frac{1}{\Delta_a}\right)}{\epsilon^2\Delta_a^{\frac{1+v}{v}}}u^{\frac{2}{v}}\right)$ with probability $1-\frac{1}{T}$. Thus the regret is bounded by

$$\mathcal{R}_T \leq (1-\frac{1}{T})O\left(\frac{\left(\log T + \log\log\frac{1}{\Delta_a}\right)}{\epsilon^2\Delta_a^{\frac{1}{v}}}u^{\frac{2}{v}}\right) + \frac{1}{T}T\max_a\Delta_a.$$

The proof of the instance-independent regret is almost the same as the proof of Theorem 4, we omit it here. ■

Before proofing the lower bounds in the LDP model, we first recall the two useful lemmas provided by (Basu et al., 2019). Let $\mathcal{H}_T \triangleq \{(a_i, x_i)\}_{i=1}^T$ be the observed history produced by the interaction between the algorithm $\mathcal{A}$ and bandit problem instance $\mathcal{P}$ up to round T. Obviously, an observed history $\mathcal{H}_T$ is a random variable sampled from the measurable space $(([K]\times\mathbb{R})^T, \mathcal{B}(([K]\times\mathbb{R})^T), \mathbb{P}_{\mathcal{A}P})$, where $\mathcal{B}(([K]\times\mathbb{R})^T)$ is the Borel set on $([K]\times\mathbb{R})^T$ and $\mathbb{P}_{\mathcal{A}P}$ is the probability measure induced by the algorithm A and the instance P. The probability measure $\mathbb{P}_{\mathcal{A}P}$ satisfies the following two properties,

1. the probability of selecting arm a at time t is dictated only by the algorithm $\mathcal{A}$, and we denote the probability by $\mathcal{A}(a|\mathcal{H}_{t-1})$.
2. the reward distribution of arm a_t, denoted by $\mathcal{X}_{a_t}$, is independent of the previous observed history $\mathcal{H}_{t-1}$.

Hence, for any observed history $\mathcal{H}_T$, we have

$$\mathbb{P}^T_{\mathcal{A}P} \triangleq \mathbb{P}_{\mathcal{A}P}(\mathcal{H}_T) = \prod_{t=1}^T \mathcal{A}(a_t|\mathcal{H}_{t-1})\mathcal{X}_{a_t}(x_t). \tag{25}$$

The next lemma states the KL-divergence decomposition for canonical bandit framework. Intuitively, by the decomposition, we separate the effect of the algorithm and the reward generation.

Lemma 14. (KL-divergence Decomposition). Given a bandit algorithm $\mathcal{A}$, two distinct instances P_1, P_2 and a probability measure $\mathbb{P}_{\mathcal{A}P}$ satisfying (25). Then

$$\mathrm{KL}\left(\mathbb{P}^T_{\mathcal{A}P_1} \| \mathbb{P}^T_{\mathcal{A}P_2}\right) = \sum_{t=1}^{T} \mathbb{E}_{\mathcal{A}P_1}\left[\mathrm{KL}\left(\mathcal{A}\left(a_t \mid \mathcal{H}_t, P_1\right) \| \mathcal{A}\left(a_t \mid \mathcal{H}_t, P_2\right)\right)\right] + \sum_{a=1}^{K} \mathbb{E}_{\mathcal{A}P_1}\left[N_a(T)\right] \mathrm{KL}\left(\mathcal{X}_a^1 \| \mathcal{X}_a^2\right). \tag{26}$$

where we use $\mathcal{X}_a^1$ and $\mathcal{X}_a^2$ to represent the reward distributions of arm a in instance P_1 and P_2 respectively, and $N_a(T)$ is the times of pulling arm a among the T rounds.

For locally differentially private bandit algorithms, the first term on the LHS of (26) vanishes since given the same history $\mathcal{H}_t$, $A\left(a_t \mid \mathcal{H}_t, P_1\right)$ and $A\left(a_t \mid \mathcal{H}_t, P_2\right)$ should be the same as they depends only on the internal randomness of the algorithm $\mathcal{A}$. The following lemma is about the locally private KL-divergence decomposition.

Lemma 15. (Locally Private KL-divergence Decomposition). If the reward generation process is ϵ-local differentially private for both the instance P_1 and P_2, Then we have

$$\mathrm{KL}\left(\mathbb{P}^T_{\mathcal{A}P_1} \| \mathbb{P}^T_{\mathcal{A}P_2}\right) \leq 2\min\left\{4, e^{2\epsilon}\right\}\left(e^{\epsilon}-1\right)^2 \sum_{a=1}^{K} \mathbb{E}_{\mathcal{A}P_1}\left[N_a(T)\right] \mathrm{KL}\left(\mathcal{X}_a^1 \| \mathcal{X}_a^2\right). \tag{27}$$

Next, we prove Theorem 11, Theorem 8 and Theorem 9. We use $\mathcal{R}^{\mathcal{A}}_{T,P}$ to represent the regret of algorithm $\mathcal{A}$ under the instance P with total T rounds.

Theorem 11. (2-Armed LDP Instance-dependent Lower Bound). There exists a heavy-tailed two-armed bandit instance with $u \leq 1$ in (2) and $\Delta \triangleq \mu_1 - \mu_2 \in (0, \frac{1}{5})$, such that for any ϵ-LDP algorithm with $\epsilon \in (0, 1]$ and regret $\leq o(T^{\alpha})$ for any $\alpha > 0$, the regret satisfies

$$\liminf_{T\to\infty} \frac{\mathcal{R}_T}{\log T} \geq \Omega\left(\frac{1}{\epsilon^2 \Delta^{\frac{1}{v}}}\right).$$

Proof [of Theorem 11] Consider the following instance $\bar{P}_1$: the distribution of the first arm a_1 is

$$\nu_1 = \left(1 - \frac{\gamma^{1+v}}{2}\right)\delta_0 + \frac{\gamma^{1+v}}{2}\delta_{1/\gamma}$$

with $\gamma = (5\Delta)^{\frac{1}{v}} (\Delta \in \left(0, \frac{1}{5}\right))$, and the distribution of the second arm a_2 is

$$\nu_2 = \left(1 - \left(\frac{\gamma^{1+v}}{2} - \Delta\gamma\right)\right)\delta_0 + \left(\frac{\gamma^{1+v}}{2} - \Delta\gamma\right)\delta_{1/\gamma}.$$

Thus,

$$\mathbb{E}[\nu_1] = \frac{5}{2}\Delta, \quad u(\nu_1) = \left(\frac{1}{\gamma}\right)^{1+v} \cdot \frac{\gamma^{1+v}}{2} = \frac{1}{2} < 1,$$
$$\mathbb{E}[\nu_2] = \frac{3}{2}\Delta, \quad u(\nu_2) = \left(\frac{1}{\gamma}\right)^{1+v} \cdot \left(\frac{\gamma^{1+v}}{2} - \Delta\gamma\right) < 1.$$

Suppose we have another instance $\bar{P}_2$: the distribution of the first arm ν_1' is the same as ν_1, and the distribution of the second arm is

$$\nu_2' = \left(1 - \left(\frac{\gamma^{1+v}}{2} + \Delta\gamma\right)\right)\delta_0 + \left(\frac{\gamma^{1+v}}{2} + \Delta\gamma\right)\delta_{1/\gamma}.$$

Then

$$\mathbb{E}[\nu_2'] = \frac{7}{2}\Delta, \quad u(\nu_2') = \left(\frac{1}{\gamma}\right)^{1+v} \cdot \left(\frac{\gamma^{1+v}}{2} + \Delta\gamma\right) = \frac{1}{2} + \frac{\Delta}{\gamma^v} = \frac{7}{10}.$$

Since $\mathcal{R}_T = \sum_{a:\Delta_a>0} \Delta_a \mathbb{E}\left[N_a(T)\right]$, we have

$$\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_1}] \geq \mathbb{P}^T_{\mathcal{A}\bar{P}_1}\left(N_2(T) \geq \frac{T}{2}\right) \cdot \left(\frac{5}{2}\Delta - \frac{3}{2}\Delta\right) \cdot \frac{T}{2} = \frac{T\Delta}{2} \cdot \mathbb{P}^T_{\mathcal{A}\bar{P}_1}\left(N_2(T) \geq \frac{T}{2}\right),$$

$$\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_2}] \geq \mathbb{P}^T_{\mathcal{A}\bar{P}_2}\left(N_2(T) \leq \frac{T}{2}\right) \cdot \left(\frac{7}{2}\Delta - \frac{5}{2}\Delta\right) \cdot \frac{T}{2} = \frac{T\Delta}{2} \cdot \mathbb{P}^T_{\mathcal{A}\bar{P}_2}\left(N_2(T) \leq \frac{T}{2}\right).$$

Thus, by Bretagnolle-Huber inequality (Lattimore and Szepesvári, 2020, Theorem 14.2), we obtain

$$\begin{aligned}\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_1}] + \mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_2}] &\geq \frac{T\Delta}{2}\left[\mathbb{P}^T_{\mathcal{A}\bar{P}_1}\left(N_2(T) \geq \frac{T}{2}\right) + \mathbb{P}^T_{\mathcal{A}\bar{P}_2}\left(N_2(T) \leq \frac{T}{2}\right)\right] \\ &\geq \frac{T\Delta}{4} \cdot \exp\left(-\mathrm{KL}(\mathbb{P}^T_{\mathcal{A}\bar{P}_1} \| \mathbb{P}^T_{\mathcal{A}\bar{P}_2})\right).\end{aligned}$$

By Lemma 15, we have $\mathrm{KL}(\mathbb{P}^T_{\mathcal{A}\bar{P}_1} \| \mathbb{P}^T_{\mathcal{A}\bar{P}_2}) \leq 8\left(e^\epsilon - 1\right)^2 \cdot \mathbb{E}_{\mathcal{A}\bar{P}_1}[N_2(T)] \cdot \mathrm{KL}(\nu_2 \| \nu_2')$. Thus,

$$\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_1}] + \mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_2}] \geq \frac{T\Delta}{4} \exp\left(-8\left(e^\epsilon - 1\right)^2 \cdot \mathbb{E}_{\mathcal{A}\bar{P}_1}[N_2(T)] \cdot \mathrm{KL}(\nu_2 \| \nu_2')\right).$$

Then, we obtain

$$\mathbb{E}_{\mathcal{A}\bar{P}_1}[N_2(T)] \geq \frac{\log\frac{T\Delta}{4} - \log\left(\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_1}] + \mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_2}]\right)}{8\left(e^\epsilon - 1\right)^2 \cdot \mathrm{KL}(\nu_2 \| \nu_2')} \geq \frac{\log\frac{T\Delta}{4} - 2\alpha\log T}{8\epsilon^2 \cdot \mathrm{KL}(\nu_2 \| \nu_2')}, \tag{28}$$

where the last inequality is due to the assumption of sub-linear regret $\mathcal{R}_T \leq o(T^\alpha)$ and the fact that $e^\epsilon - 1 \approx \epsilon$ when ϵ is small.

By using $\mathrm{KL}\left(\mathrm{Ber}(p) \| \mathrm{Ber}(q)\right) \leq \frac{(p-q)^2}{q(1-q)}$, we obtain

$$\begin{aligned}\mathrm{KL}(\nu_2 \| \nu_2') &= \mathrm{KL}\left(\mathrm{Ber}\left(\frac{\gamma^{1+v}}{2} - \Delta\gamma\right) \middle\| \mathrm{Ber}\left(\frac{\gamma^{1+v}}{2} + \Delta\gamma\right)\right) \\ &\leq \frac{(2\Delta\gamma)^2}{\left(\frac{\gamma^{1+v}}{2} + \Delta\gamma\right) \cdot \left(1 - \left(\frac{\gamma^{1+v}}{2} + \Delta\gamma\right)\right)}.\end{aligned}$$

Note that $\gamma = (5\Delta)^{\frac{1}{v}}$, we get

$$\begin{aligned}\mathrm{KL}(\nu_2 \| \nu_2') &\leq \frac{\left(2 \cdot 5^{\frac{1}{v}} \cdot \Delta^{\frac{1+v}{v}}\right)^2}{\left(\frac{5^{\frac{1+v}{v}} \cdot \Delta^{\frac{1+v}{v}}}{2} + 5^{\frac{1}{v}} \cdot \Delta^{\frac{1+v}{v}}\right)\left(1 - \frac{7}{2} \cdot 5^{\frac{1}{v}}\Delta^{\frac{1+v}{v}}\right)} \\ &\leq \frac{4 \cdot 5^{\frac{1}{v}} \cdot \Delta^{\frac{1+v}{v}}}{\frac{7}{2}\left(1 - \frac{7}{2} \cdot 5^{\frac{1}{v}}\Delta^{\frac{1+v}{v}}\right)} \leq C \cdot 5^{\frac{1}{v}} \cdot \Delta^{\frac{1+v}{v}},\end{aligned}$$

where C is some constant and the last inequality holds when $\Delta \in (0, \frac{1}{5})$ is sufficiently small.

Thus, according to (28),

$$\liminf_{T\to\infty} \frac{\mathbb{E}_{\mathcal{A}\bar{P}_1}[N_2(T)]}{\log T} \geq \Omega\left(\frac{1}{5^{\frac{1}{v}}\Delta^{\frac{1+v}{v}}\epsilon^2}\right),$$

then,

$$\liminf_{T\to\infty} \frac{\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_1}}{\log T} \geq \liminf_{T\to\infty} \frac{\mathbb{E}_{\mathcal{A}\bar{P}_1}[N_2(T)]}{\log T} \cdot \Delta \geq \Omega\left(\frac{1}{5^{\frac{1}{v}}\Delta^{\frac{1}{v}}\epsilon^2}\right).$$

■

Proof [of Theorem 8] We focus on the K arms with mean reward satisfying $\frac{1}{2} \geq \mu_1 \geq \cdots \geq \mu_K$ and $\frac{1}{5}\mu_a \leq \Delta_a \leq \frac{1}{2}\mu_a$. Consider the instance $\bar{P}$ where the distribution for each arm $a \in [K]$ is

$$\nu_a = \left(1 - \frac{s_a^{1+v}}{2}\right)\delta_0 + \frac{s_a^{1+v}}{2}\delta_{1/s_a},$$

where $s_a = (2\mu_a)^{\frac{1}{v}}$. It is easy to verify for each $a \in [K]$ that

$$\mathbb{E}[\nu_a] = \mu_a, u(\nu_a) = \frac{1}{2} < 1.$$

Then consider another instance $\bar{Q}_a$, where the reward distribution of any arm $a' \neq a$ remains unchanged and the reward distribution of a becomes

$$\nu'_a = [1 - (\frac{s_a^{1+v}}{2} + 2\Delta_a s_a)]\delta_0 + (\frac{s_a^{1+v}}{2} + 2\Delta_a s_a)\delta_{1/s_a}.$$

Note that, $\frac{s_a^{1+v}}{2} + 2\Delta_a s_a = 2^{\frac{1}{v}}\mu_a^{\frac{1+v}{v}} + 2^{\frac{1+v}{v}}\Delta_a\mu_a^{\frac{1}{v}} \leq \mu_a + 2\Delta_a = \mu_1 + \Delta_a \leq 1$, where the first inequality is due to $\mu_a \leq \frac{1}{2}$, hence the postulated ν'_a is reasonable.

For ν'_a, we have $\mathbb{E}[\nu'_a] = \mu_a + 2\Delta_a = \mu_1 + \Delta_a$ and $u(\nu'_a) = \frac{1}{2} + \frac{\Delta_a}{\mu_a} \leq 1$, where the inequality is due to $\Delta_a \leq \frac{1}{2}\mu_a$.

Since $\mathcal{R}_T = \sum_{a:\Delta_a>0} \Delta_a \mathbb{E}[N_a(T)]$, we have

$$\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}}] \geq \mathbb{P}^T_{\mathcal{A}\bar{P}}\left(N_a(T) \geq \frac{T}{2}\right) \cdot \Delta_a \cdot \frac{T}{2} = \frac{T}{2}\Delta_a \mathbb{P}^T_{\mathcal{A}\bar{P}}\left(N_a(T) > \frac{T}{2}\right),$$

$$\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{Q}_a}] \geq \mathbb{P}^T_{\mathcal{A}\bar{Q}_a}\left(N_a(T) \leq \frac{T}{2}\right) \cdot \Delta_a \cdot \frac{T}{2} = \frac{T}{2}\Delta_a \mathbb{P}^T_{\mathcal{A}\bar{Q}_a}\left(N_a(T) \leq \frac{T}{2}\right).$$

By Bretagnolle-Huber inequality (Lattimore and Szepesvári, 2020, Theorem 14.2), we obtain

$$\begin{aligned}\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}}] + \mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{Q}_a}] &\geq \frac{T\Delta_a}{2}\left[\mathbb{P}^T_{\mathcal{A}\bar{P}}\left(N_a(T) \geq \frac{T}{2}\right) + \mathbb{P}^T_{\mathcal{A}\bar{Q}_a}\left(N_a(T) \leq \frac{T}{2}\right)\right] \\ &\geq \frac{T\Delta_a}{4} \cdot \exp\left(-\mathrm{KL}(\mathbb{P}^T_{\mathcal{A}\bar{P}} \| \mathbb{P}^T_{\mathcal{A}\bar{Q}_a})\right).\end{aligned}$$

Due to Lemma 15, we have $\mathrm{KL}(\mathbb{P}^T_{\mathcal{A}\bar{P}} \| \mathbb{P}^T_{\mathcal{A}\bar{Q}_a}) \leq 8\left(e^\epsilon - 1\right)^2 \cdot \mathbb{E}_{\mathcal{A}\bar{P}}[N_a(T)] \cdot \mathrm{KL}(\nu_a \| \nu'_a)$. Thus we obtain that

$$\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}}] + \mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{Q}_a}] \geq \frac{T\Delta_a}{4} \cdot \exp\left(-8\left(e^\epsilon - 1\right)^2 \cdot \mathbb{E}_{\mathcal{A}\bar{P}}[N_a(T)] \cdot \mathrm{KL}(\nu_a \| \nu'_a)\right),$$

which gives that

$$\mathbb{E}_{\mathcal{A}\bar{P}}[N_a(T)] \geq \frac{\log\frac{T\Delta_a}{4} - \log(\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}}] + \mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{Q}_a}])}{8(e^\epsilon - 1)^2 \cdot \mathrm{KL}(\nu_a \| \nu'_a)} \geq \frac{\log\frac{T\Delta_a}{4} - 2\alpha\log T}{8\epsilon^2 \cdot \mathrm{KL}(\nu_a \| \nu'_a)}, \tag{29}$$

where the last inequality is due to the assumption of sub-linear regret $\mathcal{R}_T \leq o(T^\alpha)$ and the fact that $e^\epsilon - 1 \approx \epsilon$ when ϵ is small.

By using the fact that $\mathrm{KL}\left(\mathrm{Ber}(p) \| \mathrm{Ber}(q)\right) \leq \frac{(p-q)^2}{q(1-q)}$, we can obtain that

$$\begin{aligned}\mathrm{KL}(\nu_a \| \nu'_a) &= \mathrm{KL}\left(\mathrm{Ber}\left(\frac{s_a^{1+v}}{2}\right) \middle\| \mathrm{Ber}\left(\frac{s_a^{1+v}}{2} + 2\Delta_a s_a\right)\right) \\ &\leq \frac{(2\Delta\gamma)^2}{\left(\frac{\gamma^{1+v}}{2} + \Delta\gamma\right) \cdot \left(1 - \left(\frac{\gamma^{1+v}}{2} + \Delta\gamma\right)\right)} \\ &\leq \frac{25^{\frac{1}{v}}\Delta_a^{\frac{1+v}{v}}}{1 - 7 \cdot 10^{\frac{1}{v}}\Delta_a^{\frac{1+v}{v}}} \leq C \cdot (25)^{\frac{1}{v}}\Delta_a^{\frac{1+v}{v}},\end{aligned}$$

where C is some constant and the last inequality holds since $\frac{1}{5}\mu_a \leq \Delta_a \leq \frac{1}{2}\mu_a$ and $\mu_a \leq \frac{1}{2}$. Thus, according to (29), we have

$$\liminf_{T\to\infty} \frac{\mathbb{E}_{\mathcal{A}\bar{P}}[N_a(T)]}{\log T} \geq \Omega\left(\frac{1}{\epsilon^2\Delta_a^{\frac{1+v}{v}}}\right),$$

then,

$$\liminf_{T\to\infty}\frac{\mathcal{R}^{\mathcal{A}}_{T,\bar{P}}}{\log T}\geq\liminf_{T\to\infty}\sum_{a:\Delta_a>0}\frac{\mathbb{E}_{\mathcal{A}\bar{P}}[N_a(T)]\cdot\Delta_a}{\log T}\geq\Omega\left(\frac{1}{\epsilon^2}\sum_{\Delta_a>0}(\frac{1}{\Delta_a})^{\frac{1}{v}}\right).$$

■

Proof [of Theorem 9] We first define the instance $\bar{P}_1$. In $\bar{P}_1$, the optimal arm (denoted by a_1) follows the reward distribution

$$\nu_1=\left(1-\frac{\gamma^{1+v}}{2}\right)\delta_0+\frac{\gamma^{1+v}}{2}\delta_{1/\gamma},$$

where $\gamma=(5\Delta)^{\frac{1}{v}}(\Delta\in\left(0,\frac{1}{5}\right))$. Note that $\mathbb{E}[\nu_1]=\frac{5}{2}\Delta, u(\nu_1)=\frac{1}{2}$.

Any other sub-optimal arm $a\neq a_1$ in $\bar{P}_1$ follows the same reward distribution

$$\nu_a=\left(1-\frac{\gamma^{1+v}}{2}+\Delta\gamma\right)\delta_0+\left(\frac{\gamma^{1+v}}{2}-\Delta\gamma\right)\delta_{1/\gamma}.$$

Note that for all $a\neq a_1$ $\mathbb{E}[\nu_a]=\frac{3}{2}\Delta$, $u(\nu_a)=\frac{1}{2}-\frac{1}{5}=\frac{3}{10}<1$. We denote the corresponding locally private reward distribution for each arm $a\in[K]$ as $\bar{\nu}_a$.

For algorithm $\mathcal{A}$ and instance $\bar{P}_1$, we denote $i=\arg\min_{a\in\{2,\cdots,K\}}\mathbb{E}_{\mathcal{A}\bar{P}_1}[N_a(T)]$. Thus, $\mathbb{E}_{\mathcal{A}\bar{P}_1}[N_i(T)]\leq\frac{T}{K-1}$.

Now, consider another instance $\bar{P}_i$ where $\nu_1,\cdots,\nu_K$ are the same as those in $\bar{P}_1$ except the i-th arm such that

$$\nu_i'=\left(1-\frac{\gamma^{1+v}}{2}-\Delta\gamma\right)\delta_0+\left(\frac{\gamma^{1+v}}{2}+\Delta\gamma\right)\delta_{1/\gamma}.$$

Note that now $\mathbb{E}[\nu_i']=\frac{7}{2}\Delta$, $u(\nu_i')=\frac{7}{10}<1$. Similarly, we denote the corresponding locally private reward distribution for arm i as $\bar{\nu}_i'$.

Thus,

$$\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_1}]\geq\mathbb{P}^T_{\mathcal{A}\bar{P}_1}\left[N_i(T)\geq\frac{T}{2}\right]\frac{T}{2}\Delta,$$
$$\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_i}]\geq\mathbb{P}^T_{\mathcal{A}\bar{P}_i}\left[N_i(T)\leq\frac{T}{2}\right]\frac{T}{2}\Delta.$$

Thus by Bretagnolle-Huber inequality (Lattimore and Szepesvári, 2020, Theorem 14.2) and Lemma 15 we have,

$$\begin{aligned}\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_1}]+\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_i}]&\geq\frac{T\Delta}{4}\exp\left(-\mathrm{KL}(\mathbb{P}^T_{\mathcal{A}\bar{P}_1}\|\mathbb{P}^T_{\mathcal{A}\bar{P}_i})\right)\\&\geq\frac{T\Delta}{4}\exp\left(-\mathbb{E}_{\mathcal{A}\bar{P}_1}[N_i(T)]\cdot\mathrm{KL}(\bar{\nu}_\mathrm{i}\|\bar{\nu}_\mathrm{i}')\right)\\&\geq\frac{T\Delta}{4}\exp\left(-8(\mathrm{e}^\epsilon-1)^2\cdot\mathbb{E}_{\mathcal{A}\bar{\mathrm{P}}_1}[\mathrm{N_i(T)}]\cdot\mathrm{KL}(\nu_\mathrm{i}\|\nu_\mathrm{i}')\right).\end{aligned}$$

Since

$$\mathrm{KL}(\nu_\mathrm{i}\|\nu_\mathrm{i}')\leq\frac{(2\Delta\gamma)^2}{\left(\frac{\gamma^{1+v}}{2}+\Delta\gamma\right)\left(1-\left(\frac{\gamma^{1+v}}{2}+\Delta\gamma\right)\right)}\leq\mathrm{C}\cdot5^{\frac{1}{v}}\cdot\Delta^{\frac{1+v}{v}},$$

for some constant $C>0$ and Δ is sufficiently small.

We obtain

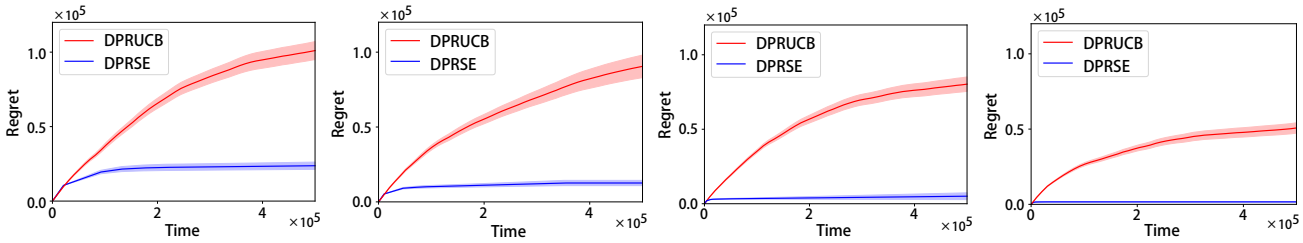
$$\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_1}]\geq\frac{T\Delta}{8}\exp\left(-8\epsilon^2\cdot\frac{T}{K-1}\cdot C\cdot5^{\frac{1}{v}}\cdot\Delta^{\frac{1+v}{v}}\right).$$

Taking $\Delta=\left(\frac{K}{T\epsilon^2}\right)^{\frac{v}{1+v}}$, we get the result

$$\mathbb{E}[\mathcal{R}^{\mathcal{A}}_{T,\bar{P}_1}]\geq\Omega\left(T^{\frac{1}{1+v}}\left(\frac{K}{\epsilon^2}\right)^{\frac{v}{1+v}}\right).$$

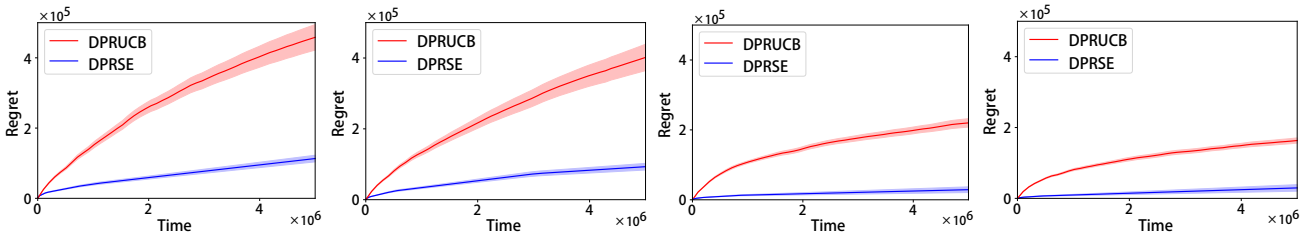
■

D OMITTED EXPERIMENTAL RESULTS FOR SECTION 6



(a) $v = 0.5, \epsilon = 0.5$ (b) $v = 0.5, \epsilon = 1.0$ (c) $v = 0.9, \epsilon = 0.5$ (d) $v = 0.9, \epsilon = 1.0$

Figure 3: DP Setting 2 (S_2)



(a) $v = 0.5, \epsilon = 0.5$ (b) $v = 0.5, \epsilon = 1.0$ (c) $v = 0.9, \epsilon = 0.5$ (d) $v = 0.9, \epsilon = 1.0$

Figure 4: DP Setting 3 (S_3)