Locally Private and Robust Multi-Armed Bandits

Xingyu Zhou[*] Wei Zhang[†]

Abstract

We study the interplay of local differential privacy (LDP) and robustness with respect to Huber corruption and possibly heavy-tailed rewards in the context of multi-armed bandits (MABs). We consider two different practical settings: LDP-then-Corruption (LTC) where each user's locally private response might be further corrupted during the data collection process, and Corruption-then-LDP (CTL) where each user may be adversary or corrupted such that each LDP mechanism will be applied to the corrupted data. To start with, we present the first tight characterization of high-probability mean estimation error under both LTC and CTL settings. Leveraging this result, we then present an almost tight (up to log factor) characterization of the minimax regret in online MABs and sub-optimality in offline MABs under both LTC and CTL settings. One key message behind all three problems is that LTC is a more difficult setting and leads to a worse performance guarantee compared to the CTL setting (in the minimax sense). This interesting interplay between privacy and corruption highlights that one needs to carefully design and analyze bandit algorithms when considering both privacy and corruption rather than treating them as linear combinations. Along the way, several results are of independent interest. Our proposed minimax optimal mean estimators can find application in many other scenarios. As a by-product, we also improved the state-of-the-art regret lower bound for locally private and heavy-tailed online MABs, i.e., without Huber corruption.

[*]Wayne State University, Detroit, USA. Email: xingyu.zhou@wayne.edu

[†]Texas A&M University, College Station, USA. Email: komo@tamu.edu Work done during research intern at Wayne State University

Contents

1 Introduction 3

2 Related Work 4

3 Problem Setup 5

4 Preview of Main Results 8

5 Mean Estimation 10

6 Online MABs 12

7 Offline MABs 14

8 Simulations 15
- 8.1 Mean estimation . . . 16
- 8.2 Online MABs . . . 18
- 8.3 Offline MABs . . . 19

9 Proof 20
- 9.1 Proof of Proposition 1 . . . 20

10 Conclusion 23

A Proof of Proposition 2 24

B Proof of Proposition 3 28

C Proof of Proposition 4 29

D Proof of Proposition 5 31

E Proof of Proposition 6 32

F Online MABs under Strong Huber Corruption 32

1 Introduction

The Multi-Armed Bandit (MAB) problem [BF85] offers a fundamental approach for sequential decision-making under uncertainty based on only bandit feedback. Take online advertising as an illustrative example, where the advertising platform (i.e., the central learner) sequentially and adaptively displays ads (i.e., arm) based on users' reward feedback (e.g., engagement score) so as to maximize the cumulative rewards. In practice, several important factors have to be considered when designing real-world MAB algorithms, as illustrated below using our online advertising example.

- **Privacy.** The raw engagement score (which is calculated based on clicks, purchases, and time spent viewing the ad, etc.) from a user's device may lead to privacy leakage. For instance, when the ad is about a medicine on some rare or uncommon disease, a high engagement score might imply interest or association with the uncommon disease. Such privacy leakage may lead to unintended personal and social consequences as well as trust issues on the platform. One principled way to mitigate it is via local differential privacy (LDP) [KLNRS11; DJW18], i.e., each user's device locally adds a suitable amount of noise (depending on the privacy mechanism and budget) to obfuscate the raw feedback before sending it out.

- **Robustness.** Another important factor in real-world scenarios is the robustness of MAB algorithms under both possibly heavy-tailed feedback and adversary corruption.
  - *Heavy-tailed feedback.* For example, the engagement score in our example could often be heavy-tailed, i.e., non-negligible probabilities of observing extremely high values. This might happen due to some special events and seasons (e.g., Black Friday) or influencer interaction (e.g., a celebrity or influencer shares an ad).
  - *Adversary corruption.* For example, during the collection/transmission of users' feedback, there could be malicious attacks/corruption on the engagement scores, e.g., with some probability, each score could be replaced by any *arbitrary* value, i.e., Huber corruption [Hub64]. On the other hand, (Huber) corruption can also happen on each user's side before transmission. For instance, a user might manipulate or spoof interactions to artificially inflate or deflate the engagement score. Moreover, malicious software or hacking could also alter the engagement score recorded on a user's device before transmission.

To tackle the above privacy and robustness issues in MABs, there has been a large related literature, which, however, mainly investigates the two issues in an isolated way (see Section 2 for details). Motivated by this, in this work, we are particularly interested in the following question:

Is there any interesting interplay between privacy and robustness in MABs?

Our contributions. We give an affirmative answer to the above question by unveiling a fundamental interplay between privacy protection (in particular, local differential privacy (LDP)) and robustness under Huber corruption and heavy-tailedness. Our main message is a separation result between two MAB settings that differ in the order of privacy protection and corruption, i.e., LDP-then-corruption (LTC) vs. Corruption-then-LDP (CTL). That is, under LTC, corruption happens after LDP mechanism while under CTL, corruption happens before the LDP mechanism. To obtain our separation result for the two settings, we take the following principled approach:

(i) We first study the mean estimation problem – a cornerstone step in the analysis of stochastic MABs – under both LTC and CTL settings. We give the first tight characterization of the high-probability concentration result with respect to privacy budget, corruption level, and heavy-tailedness under the two settings. Specifically, we first establish a lower bound on the minimax high-probability mean error rate and then propose minimax optimal estimators that achieve matching worst-case upper bounds. The key observation here is that mean estimation error under LTC is larger than that under CTL and moreover the gap becomes larger as the privacy requirement becomes stronger.

(ii) Leveraging the above tight mean estimation results, we then study both online MABs and offline MABs under both LTC and CTL. We present an almost tight characterization (up to log factor) of the corresponding minimax performances (i.e., regret in online MABs and sub-optimality in offline MABs) by deriving lower bounds and proposing almost optimal algorithms. As in mean estimation, there is a seperation between LTC and CTL, i.e., LTC is a more difficult setting that leads to a worse performance in the minimax sense, highlighting the interesting interplay between privacy and robustness in MABs.

(iii) Along the way, several results could be of independent interest. First, our optimal locally private and robust mean estimators can be applied to many other applications beyond MABs. Moreover, as a by-product, we also improved the state-of-the-art regret lower bound in [TWZW22] for locally private and heavy-tailed MABs, i.e., without corruption. Finally, our separation result also offer useful insights for practitioners when designing private and robust algorithms.

2 Related Work

Private MABs. To offer mathematically rigorous privacy protection, LDP is first introduced to MABs in [RZLS20] where the authors establish private lower bounds on both problem-dependent and problem-independent (minimax) regrets as well as several LDP mechanisms and learning algorithms that achieve nearly-optimal performance. Later, it is generalized to the heavy-tailed setting in [TWZW22]. LDP has also been considered in various other bandit settings [CZZYCW20; ZCHLW20; ZT21]. In addition to LDP, other strictly weaker privacy models have also been considered in MABs to achieve a better regret, such as central DP where users need to trust the central learner [MT15; TD16; SS19] and distributed DP where users need to trust the intermediate third-party [TKMS21; CZ22a]. In addition to the above online MABs, recent work [QW22] also considers offline RL (hence MABs) under central DP with bounded rewards.

Robust MABs. Robust MABs under Huber corruption have been recently studied in [KPK19; MTCD21; BMM22; AMBM23]. Several other corruption models have also been considered in MABs, such as budgeted-corruption model where the cumulative difference between observed reward and true reward is bounded by some constant budget [LMP18; GKT19] and strong contamination model [NT20; ABM19]. Robust regret minimization in MABs under heavy-tailed rewards have also been studied, e.g., [BCL13; AJK21].

Private and Robust MABs. As mentioned above, the existing literature largely investigate privacy and robustness in MABs separately. To the best of our knowledge, there are only two very recent works that consider privacy and robustness in MABs simultaneously. In [WZTW23], the authors consider the central DP model where the raw non-private feedback received by the central learner can be first corrupted under Huber model. This is in sharp contrast to our local DP model, which is not only stronger but allows us to study the order of corruption and privacy. In [CEM23], the authors study linear bandits (which includes MAB as a special case) under LDP and then Huber corruption (i.e., LTC setting). As will be discussed in Section 6, their regret bound is sub-optimal and worse than ours when reduced to the MAB case. Note that we also study the CTL setting, which in turn highlights the interplay between privacy and corruption.

Private and Robust Mean Estimation. Our work is inspired by recent advances in (locally) private and robust mean estimation. In particular, for the CTL setting, the authors of [LBY22] give the tight characterization in terms of mean-square-error (MSE). In contrast, we derive the high probability concentration. For LTC, both [CSU21; CS23] give constant-probability concentration when the inlier distribution is bounded. Instead, we present the high-probability version even for heavy-tailed inlier distribution, which requires new analysis and design of the estimators. We would also like to point out some other related private and/or robust mean estimation results. For instance, under central DP, [KSU20] gives the first high probability mean concentration for heavy-tailed distributions. For standard non-private mean estimation under heavy tails, we refer readers to the nice survey [LM19], and for non-private mean estimation under corruption in general high-dimension space, we refer readers to the nice book [DK23]. We finally remark that there are recent exciting advances in understanding the connection between robustness and privacy in mean estimation (e.g., robustness induces privacy [HKMN23; AUZ23] and vice versa [GH22]).

3 Problem Setup

In this section, we will formally introduce the three problems considered in this paper: mean estimation, online and offline MABs, under the constraints of both local differential privacy (LDP) and robustness (with respect to heavy tails as well as Huber corruption). To start with, we introduce the privacy and corruption models. Specifically, as for privacy, we consider (non-interactive) local differential privacy (LDP) as defined below.

Definition 1 (ε-LDP, [DJW18]). For a privacy parameter $\varepsilon \in [0, 1]$, the random variable $\widetilde{X}$ is an ε-*locally differentially private* view of X via privacy channel/mechanism Q if

$$\sup_{S\in\sigma(\widetilde{\mathcal{X}}),x,x'\in\mathcal{X}} \frac{Q(\widetilde{X} \in S \mid X = x)}{Q(\widetilde{X} \in S \mid X = x')} \leq e^{\varepsilon},$$

where $\sigma(\widetilde{\mathcal{X}})$ denotes an appropriate σ-field on $\widetilde{\mathcal{X}}$. In this case, we also say that the conditional distribution (privacy channel) Q is an ε-LDP privacy mechanism. We write $\mathcal{Q}_{\varepsilon}$ as the set of all ε-LDP mechanisms (channels).

As for corruption (contamination), we consider the following standard Huber model.

Definition 2 (α-Huber corruption, [Hub64]). Given a parameter $\alpha \in [0, 1/2)$ and a distribution D on inliers, the output distribution under α-Huber model is $O = (1 - \alpha)D + \alpha E$. That is, a sample from O returns a sample from D with probability $1 - \alpha$ and otherwise returns a sample from some (unconstrained and unknown) corruption distribution E. We write $\mathcal{C}_{\alpha}(D)$ as the set of all possible α-Huber corruptions (channels) of inlier distribution D.

With the two definitions in hand, we are able to introduce the two settings in this paper: (i) LDP-then-Corruption (LTC) vs. (ii) Corruption-then-LDP (CTL), as also illustrated in Fig. 1.

Definition 3 (LTC vs. CTL). We consider the following two settings which include both ε-LDP and α-Huber corruption.

(i) **LDP-then-Corruption (LTC):** Each user $i \in [n]$ first generates an ε-LDP view of raw data X_i. Then, the private data Y_i from each device is independently corrupted by an α-Huber channel that outputs Z_i to the central analyzer/agent.

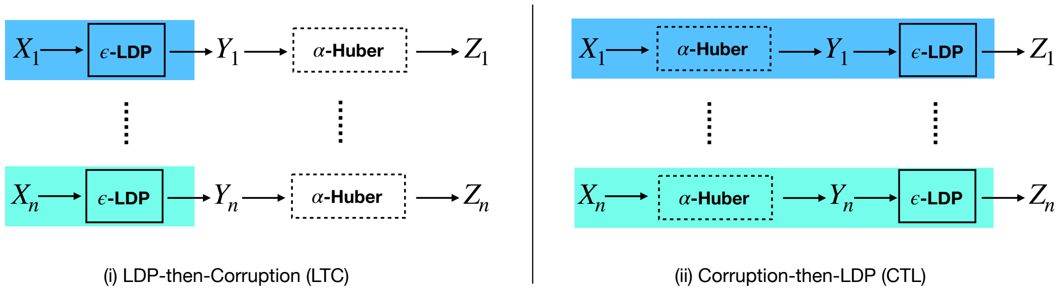


Figure 1: Illustration of the two settings considered in the paper (LTC vs. CTL). The colored region represents the boundary of each user's device.

(ii) **Corruption-then-LDP (CTL):** Each user's raw data X_i is first independently corrupted by an α-Huber model. Then, the corrupted data Y_i passes through an ε-LDP mechanism at each device that outputs Z_i to the central analyzer/agent.

Under both settings, the ultimate goal is to design the ε-LDP mechanism at each user's device and the analyzer at the central side so as to guarantee that each user's output is locally private while the computation at the central side is robust to arbitrary α-Huber corruption and possible heavy-tailed distributions of users' raw data. One particular motivation of this paper is to study how the different orders of LDP and corruption in the two settings impact the performance of different problems, i.e., mean estimation, online and offline MABs.

Mean estimation. As in [DJW18], given a real number $k > 1$, we consider the following class of distributions

$$\mathcal{P}_k := \{\text{distributions } P \text{ such that } \mathbb{E}_{X\sim P}[X] \in [-1,1] \text{ and } \mathbb{E}_{X\sim P}\left[|X|^k\right] \le 1\}. \tag{1}$$

That is, k controls the tail behavior of the distribution with smaller k meaning heavier of the tails. Given any distribution $P \in \mathcal{P}_k$, our goal is to estimate its mean $\mu(P)$ as accurately as possible. In contrast to the standard case where the analyzer has access to $i.i.d$ samples $\{X_i\}_{i=1}^n$ from P, the analyzer in this paper now only observes samples $\{Z_i\}_{i=1}^n$ that are private and corrupted view of $\{X_i\}_{i=1}^n$. Specifically, we are interested in the high probability error rate under our two different settings (LTC vs. CTL), as formally defined below.

Definition 4 (Minimax mean estimation error rate)**.** Given $\delta > 0$ and sample size $n > 0$, the minimax mean estimation error rate of the class $\mathcal{P}_k$ under ε-LDP and α-Huber corruption is defined as follows

$$\phi_\delta^*(k,\varepsilon,\alpha,n) := \inf\left\{\phi > 0 \mid \inf_{Q\in\mathcal{Q}_\varepsilon} \inf_{\widehat{\mu}_n} \sup_{P\in\mathcal{P}_k} \sup_{C\in\mathcal{C}_\alpha(P)} \mathbb{P}\left[|\widehat{\mu}_n - \mu(P)| > \phi\right] \le \delta\right\}, \tag{2}$$

where $\widehat{\mu}_n$ is a measurable function of $\{Z_i\}_{i=1}^n$, i.e., private and corrupted view of n $i.i.d$ samples $\{X_i\}_{i=1}^n$ from $P \in \mathcal{P}_k$ that pass through ε-LDP channel Q and α-Huber corruption channel C. We write $\phi^*_{\delta,\text{LTC}}(k,\varepsilon,\alpha,n)$ and $\phi^*_{\delta,\text{CTL}}(k,\varepsilon,\alpha,n)$ for the settings of LTC and CTL, respectively.

Intuitively speaking ϕ_δ^* represents the minimal error rate that any ε-LDP estimator can achieve with high probability $1-\delta$ for all distributions $P \in \mathcal{P}_k$ and all α-Huber corruption models, i.e., worst-case. Thus, the goal in our mean estimation problem is to design the optimal ε-LDP mechanism $Q^\star$ at each user's side and the optimal analyzer $\widehat{\mu}_n^\star$ at the central analyzer so as to attain the minimax mean estimate error rate in (2).

Online MABs. At each round $t \in [T]$, the central learner/analyzer chooses an action/arm $a_t \in [K]$ according to a policy π and receives a reward sample X_t that is drawn from some distribution P_{a_t} with unknown mean $r(a_t) := \mu(P_{a_t})$. Here, the policy is $\pi = \{\pi_t\}_{t=1}^T$ and π_{t+1} is a measurable function of the data received by the end of round t, i.e., for each $t \in [T]$, $\mathcal{D}_t = \{(a, X^{(a)}(t))\}_{a\in[K]}$ where $X^{(a)}(t) := \{X_1^{(a)}, \ldots, X_{N_a(t)}^{(a)}\}$ and $\sum_{a\in K} N_a(t) = t$. That is, for each round t, $X^{(a)}(t)$ groups together all $N_a(t)$ rewards from each arm $a \in [K]$ where $N_a(t)$ is the total number of times that arm a has been pulled by time t.

The goal in online MABs is to find the best policy that maximizes the total expected rewards or equivalently, to minimize the expected cumulative regret, which is the total expected loss of not choosing the best arm $a^\star$ with the largest mean. In contrast to the standard case, the learner in this paper only has access to a private and corrupted view of the true rewards and the goal is to characterize the minimax *clean* regret under our LTC and CTL settings defined below.

Definition 5 (Minimax clean regret)**.** Let $\mathrm{MAB}(k) := \{\{P_a\}_{a\in K} \mid P_a \in \mathcal{P}_k\}$ be the class of K-armed MAB instances with inlier distributions for each arm in $\mathcal{P}_k$. Then, the minimax clean regret is defined as

$$R^*(k, \varepsilon, \alpha, T) := \inf_{Q\in\mathcal{Q}_\varepsilon} \inf_{\pi} \sup_{I\in\mathrm{MAB}(k)} \sup_{C\in\mathcal{C}_\alpha(I)} \mathbb{E}\left[T \cdot r(a^\star) - \sum_{t=1}^{T} r(a_t)\right], \tag{3}$$

where a_{t+1} is a measurable function (via policy π) of private and corrupted dataset $\{(a, Z^{(a)}(t))\}_{a\in[K]}$. Here, for any arm $a \in [K]$ and $t \in [T]$, $Z^{(a)}(t) := \{Z_1^{(a)}, \ldots, Z_{N_a(t)}^{(a)}\}$ is the private and corrupted view of $N_a(t)$ samples of P_a that pass through ε-LDP channel Q and α-Huber corruption channel C. Recall that $r(a) := \mu(P_a)$, i.e., the true mean of the inlier distribution. We write $R^*_{\mathrm{LTC}}(k, \varepsilon, \alpha, T)$ and $R^*_{\mathrm{CTL}}(k, \varepsilon, \alpha, T)$ for the settings of LTC and CTL, respectively.

The goal here is to design the optimal ε-LDP mechanism $Q^\star$ (providing privacy for each user's reward feedback) and optimal learning policy $\pi^\star$ so as to attain the minimax clean regret in (3).

Remark 1. As standard in the literature [WZTW23; CKMY22; NT20], $r(\cdot)$ in (3) is the mean of inlier distributions while the randomness in the expectation is generated by both privacy and corruption.

Offline MABs. In the offline case, the analyzer cannot interact with users and instead, it is given a batch pre-collected dataset $\mathcal{D} = \{(a_i, X_i)\}_{i=1}^N$ sampled from some joint distribution of a behavior policy π and reward distributions $\{P_a\}_{a\in[K]}$. That is, each $a_i \sim \pi$ and given that $a_i = a \in [K]$, $X_i \sim P_a$ and $r(a) := \mu(P_a)$. As above, one can also rewrite $\mathcal{D}$ as $\mathcal{D} = \{(a, X^{(a)})\}_{a\in[K]}$ where $X^{(a)} := \{X_1^{(a)}, \ldots, X_{N_a}^{(a)}\}$ and $\sum_{a\in K} N_a = N$. That is, we group together all the reward feedbacks for each arm $a \in [K]$. As in [RZMJR21], we assume a finite concentrability coefficient $\beta^\star$ such that $1/\pi(a^\star) \le \beta^\star$, where $a^\star$ is the optimal arm that has the largest mean and $\beta^\star$ captures deviation between the behavior distribution π and the distribution induced by the optimal policy (i.e., always choosing the best arm $a^\star$, which has the largest mean).

The goal of offline MABs is to select an arm $\widehat{a}$ that minimizes the expected sub-optimality: $\mathrm{SubOpt} = \mathbb{E}_{\mathcal{D}}\left[r(a^\star) - r(\widehat{a})\right]$. In contrast to the standard case, the analyzer in our paper now only has access to private and corrupted view of X_i and the goal is to characterize the minimax sub-optimality under our LTC and CTL settings defined below.

Definition 6 (Minimax sub-optimality)**.** Let $\mathrm{MAB}(\beta^\star, k) := \{(\pi, \{P_a\}_{a\in K}) \mid P_a \in \mathcal{P}_k \text{ and } 1/\pi(a^\star) \le \beta^\star\}$ be the class of K-armed MAB instances with distributions in $\mathcal{P}_k$ and concentrability coefficient $\beta^\star$. Then, the minimax sub-optimality is defined as

$$\mathrm{SubOpt}^*(\beta^\star, k, \varepsilon, \alpha, N) := \inf_{Q\in\mathcal{Q}_\varepsilon} \inf_{\widehat{a}} \sup_{I\in\mathrm{MAB}(\beta^\star,k)} \sup_{C\in\mathcal{C}_\alpha(I)} \mathbb{E}\left[|r(a^\star) - r(\widehat{a})|\right], \tag{4}$$

where $\widehat{a}$ is a measurable function of private and corrupted dataset $\{(a, Z^{(a)})\}_{a\in[K]}$ and $Z^{(a)} := \{Z_1^{(a)}, \ldots, Z_{N_a}^{(a)}\}$ is the private and corrupted view of N_a samples of P_a that pass through ε-LDP channel Q and α-Huber corruption channel C. As before, we will write $\mathrm{SubOpt}^*_{\mathrm{LTC}}(\beta^\star, k, \varepsilon, \alpha, N)$ and $\mathrm{SubOpt}^*_{\mathrm{CTL}}(\beta^\star, k, \varepsilon, \alpha, N)$ for the settings of LTC and CTL, respectively.

We remark that we assume the batch data is collected by an ε-LDP mechanism that can be specified by the learner. Note that as in the standard case, we do not control the behavior policy π other than a finite $\beta^\star$. The goal here is to design the optimal ε-LDP mechanism $Q^\star$ (which protects local privacy for any users offering batch data) and optimal offline learning algorithm $\widehat{a}^\star$ so as to attain the minimax sub-optimality in (4).

4 Preview of Main Results

In this section, we first present a preview of our main results. We give tight characterizations of all the above three problems – mean estimation, online and offline MABs – in both LTC and CTL settings. One common key message behind all of them is that LTC is a more difficult setting and leads to a worse performance guarantee compared to CTL setting, in the minimax sense.

Let us start with the following theorem on the minimax error rate (cf. Def. 4) for mean estimation.

Theorem 1 (Mean Estimation)**.** *Consider the mean estimation problem under both LTC and CTL settings. Given any fixed* $\delta \in (0, 1/2)$[1]*,* $\varepsilon \in [0, 1]$*,* $\alpha \in [0, 1/2)$ *and* $k > 1$*, we have that for all large enough* n*,*

$$\phi^*_{\delta,\mathrm{LTC}}(k, \varepsilon, \alpha, n) = \Theta\left(\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} + \left(\frac{1}{\varepsilon}\sqrt{\frac{\log(1/\delta)}{n}}\right)^{1-1/k}\right),$$

$$\phi^*_{\delta,\mathrm{CTL}}(k, \varepsilon, \alpha, n) = \Theta\left(\alpha^{1-1/k} + \left(\frac{1}{\varepsilon}\sqrt{\frac{\log(1/\delta)}{n}}\right)^{1-1/k}\right).$$

Remark 2. To the best of our knowledge, this is the first high-probability concentration bound for mean estimation under both LTC and CTL, which tightly captures the dependence on the corruption level α, privacy budget ε and heavy-tail parameter k, simultaneously. It can be seen that for LTC setting, there is an additional $(1/\varepsilon)^{1-1/k}$ factor, which implies that introducing LDP guarantee first would make it more vulnerable to corruption/data manipulation attacks. Interestingly, for a fixed ε, this additional vulnerability due to LDP decreases as the tail becomes heavier, which offers additional insight into the interplay of privacy, heavy-tailedness, and robustness. Our LTC result also complements the result in [CSU21], which established consdiers the bounded case (i.e., $k = \infty$) under constant probability rather than high probability. On the other hand, for CTL, we note that the impact of corruption and privacy is separable. Our high probability bound for CTL complements the error bound in terms of mean-square error (MSE) in [LBY22].

[1] Here, we assume that δ does not depends on n; otherwise, $\delta \in (\delta_{\min}, 1/2)$ where $\delta_{\min} = e^{-cn}$ for some constant $c > 0$.

As two interesting applications of Theorem 1, we then study both online and offline MABs under LTC and CTL settings. As mentioned before, LTC captures the practical bandit scenarios where the locally private reports of each user may be further corrupted during data collection process while CTL captures the case where each user may be corrupted or malicious so that her data is first contaminated before passing through the LDP mechanism on each device.

For online MABs, the following theorem gives an almost tight characterization (up to log factor) of its minimax clean regret (cf. Def. 5) for both LTC and CTL settings.

Theorem 2 (Online MABs)**.** *Consider the online MAB problem under both LTC and CTL settings. Given any $\varepsilon \in [0,1]$, $\alpha \in [0,1/2)$ and $k > 1$, we have for all T large enough*

$$R^*_{\delta,\mathrm{LTC}}(k,\varepsilon,\alpha,T) = \widetilde{\Theta}\left(T \cdot \left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} + T^{\frac{k+1}{2k}}\left(\frac{K}{\varepsilon^2}\right)^{\frac{k-1}{2k}}\right),$$

$$R^*_{\delta,\mathrm{CTL}}(k,\varepsilon,\alpha,T) = \widetilde{\Theta}\left(T \cdot \alpha^{1-1/k} + T^{\frac{k+1}{2k}}\left(\frac{K}{\varepsilon^2}\right)^{\frac{k-1}{2k}}\right),$$

where $\widetilde{\Theta}(\cdot)$ hides log factor in T.

Remark 3 (Linear term). For both settings, due to corruption, the minimax clean regret (i.e., problem-independent regret) has a linear dependence on T, as in previous works under Huber corruption, e.g., [WZTW23] for bandits under central DP and Huber corruption and [CKMY22] for non-private contextual bandits under Huber model. The key here is to capture the tight factor in front of T, where the additional $1/\varepsilon$ factor in LTC again demonstrates the sharp difference between the two settings as in the mean estimation problem.

Remark 4 (Tighter lower bound for non-corrupted case). If one only considers privacy and heavy-tailed rewards for online MABs (hence $\alpha = 0$), our lower bound is tighter (with respect to T) than the state-of-the-art in [TWZW22], which is $\Omega\left(\left(\frac{K}{\varepsilon^2}\right)^{1-1/k} T^{1/k}\right)$ for $k \in (1,2]$ when adapted to our setting[2].

Finally, for offline MABs, the following theorem gives an almost tight characterization (up to log factor) of its minimax sub-optimality (cf. Def. 6) for both LTC and CTL settings.

Theorem 3 (Offline MABs)**.** *Consider the offline MAB problem under both LTC and CTL settings. Given any $\varepsilon \in [0,1]$, $\alpha \in [0,1/2)$ and $k > 1$, we have for all N large enough and $\beta^\star \geq 2$*

$$\mathrm{SubOpt}_{\mathrm{LTC}}(\beta^\star,k,\varepsilon,\alpha,N) = \widetilde{\Theta}\left(\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} + \left(\frac{1}{\varepsilon}\sqrt{\frac{\beta^\star}{N}}\right)^{1-1/k}\right),$$

$$\mathrm{SubOpt}_{\mathrm{CTL}}(\beta^\star,k,\varepsilon,\alpha,N) = \widetilde{\Theta}\left(\alpha^{1-1/k} + \left(\frac{1}{\varepsilon}\sqrt{\frac{\beta^\star}{N}}\right)^{1-1/k}\right),$$

where $\widetilde{\Theta}(\cdot)$ hides log factor in N.

Remark 5. Similar to the previous two problems, LTC has an additional $1/\varepsilon$ factor compared to the CTL setting. We also note that the condition on the coverage $\beta^\star \geq 2$ is assumed even in case of non-private, non-corrupted, and bounded reward case as in [RZMJR21].

[2] In fact, our lower bound also implies that the claimed upper bound of the order $O(T^{1/k})$ in [TWZW22] is ungrounded.

The rest of the paper is organized as follows. In Section 5 for mean estimation, we will state the lower bounds along with intuition, and give the optimal estimators that enjoy the matching upper bounds. Then, in Section 6 for online MABs and Section 7 for offline MABs, we will present nearly optimal algorithms that are able to achieve the corresponding minimax performance. Finally, in Section 8, we provide more insights into our separation results via numerical simulations.

5 Mean Estimation

In this section, we first formally state the lower bound on the minimax error rate and provide some key intuition on the difference between LTC and CTL settings. Then, we present the optimal mean estimators for the two settings along with their performance upper bounds.

Let us start with the lower bound, the detailed proof of which is given in Section 9.1.

Proposition 1 (Lower Bounds)**.** *Consider the mean estimation problem under LTC and CTL settings. Given any fixed $\delta \in (0, 1/2)$, $\varepsilon \in [0, 1]$, $\alpha \in [0, 1/2)$, $k > 1$ and n large enough, for all ε-LDP mechanism Q and all estimator $\widehat{\mu}_n$, there exists a distribution $P \in \mathcal{P}_k$ and α-Huber corruption channel $C \in \mathcal{C}_\alpha(P)$ such that with probability at least δ*

(i) For LTC: $|\widehat{\mu}_n - \mu(P)| \geq \Omega\left(\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} + \left(\frac{1}{\varepsilon}\sqrt{\frac{\log(1/\delta)}{n}}\right)^{1-1/k}\right)$,

(ii) For CTL: $|\widehat{\mu}_n - \mu(P)| \geq \Omega\left(\alpha^{1-1/k} + \left(\frac{1}{\varepsilon}\sqrt{\frac{\log(1/\delta)}{n}}\right)^{1-1/k}\right)$,

where recall that $\widehat{\mu}_n$ is a measurable function of $\{Z_i\}_{i=1}^n$, i.e., private and corrupted view of n $i.i.d$ samples $\{X_i\}_{i=1}^n$ from $P \in \mathcal{P}_k$ that pass through ε-LDP channel Q and α-Huber corruption channel C.

Proof ideas and intuition. We provide a summary of the key ideas and intuition behind the proof. Essentially, we divide the proof into two parts. First, we consider the case without corruption and aim to establish the second term in the bound. To this end, we will leverage some tools from information theory, e.g., maximal coupling, strong data processing inequality of LDP, and Bretagnolle–Huber inequality between TV and KL distance. Then, we turn to give the first term related to corruption. To this end, we will leverage a folklore but important fact about Huber model. Roughly speaking, this fact says that given two inlier distributions D_1 and D_2 that satisfy $\mathrm{TV}(D_1, D_2) \leq O(\alpha)$, then after α-Huber channel, one cannot distinguish between D_1 and D_2. Another important fact is that ε-LDP channel is a "contraction" channel in terms of TV distance, i.e., $\mathrm{TV}(M_1, M_2) \leq O(\varepsilon)\mathrm{TV}(P_1, P_2)$ where M_1, M_2 are induced marginals of P_1, P_2 after any ε-LDP channel. Building upon the above facts, one can immediately see that under the LTC setting, due to the "contraction" of LDP, one can choose two distributions that have a larger mean difference by a factor of $1/\varepsilon$, explaining the key difference of $1/\varepsilon$ between LTC and CTL. □

Now, we turn to our proposed optimal mean estimators for LTC and CTL settings, see Algorithms 1 and 2, respectively. Given their similarity, we will mainly focus on the LTC setting. Algorithm 1 consists of two components: an ε-LDP mechanism/randomizer Q at each user's side and an analyzer $\mathcal{A}$ at the aggregator/learner. The task of Q is to guarantee that its output is an ε-LDP view of its input. To this end, for each input X_i, it first truncates it into $\bar{X}_i$ using a properly chosen threshold M. Then, it converts the real number to binary data via random rounding. Next, it applies random response technique to generate the final output Y_i, i.e., with probability $\frac{e^\varepsilon}{e^\varepsilon+1}$, outputs a number of the same sign (with additional scaling

Algorithm 1 Mean estimation under LTC

Procedure: ε-LDP Mechanism Q

//Input: X_i, Parameters $\Phi := \{M, \varepsilon\}$

//Output: Y_i

Truncate: $\bar{X}_i = X_i \mathbb{1}(|X_i| \leq M)$

Random rounding:

$$X_i' = \begin{cases} M & w.p. \ \frac{1+\bar{X}_i/M}{2} \\ -M & w.p. \ \frac{1-\bar{X}_i/M}{2} \end{cases}$$

Random response:

$$Y_i = \begin{cases} \frac{e^\varepsilon+1}{e^\varepsilon-1} X_i' & w.p. \ \frac{e^\varepsilon}{e^\varepsilon+1} \\ -\frac{e^\varepsilon+1}{e^\varepsilon-1} X_i' & w.p. \ \frac{1}{e^\varepsilon+1} \end{cases}$$

Return Y_i

Procedure: Analyzer $\mathcal{A}$

//Input : $\{Z_i\}_{i=1}^n$, Parameters Φ

//Output : estimator $\widehat{\mu}_n$

Return $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n Z_i \mathbb{1}(|Z_i| \leq M \cdot \frac{e^\varepsilon+1}{e^\varepsilon-1})$

Algorithm 2 Mean estimation under CTL

Procedure: ε-LDP Mechanism Q

//Input: Y_i, Parameters $\Phi := \{M, \varepsilon\}$

//Output: Z_i

Truncate: $\bar{Y}_i = Y_i \mathbb{1}(|X_i| \leq M)$

Random rounding:

$$Y_i' = \begin{cases} M & w.p. \ \frac{1+\bar{Y}_i/M}{2} \\ -M & w.p. \ \frac{1-\bar{Y}_i/M}{2} \end{cases}$$

Random response:

$$Z_i = \begin{cases} \frac{e^\varepsilon+1}{e^\varepsilon-1} Y_i' & w.p. \ \frac{e^\varepsilon}{e^\varepsilon+1} \\ -\frac{e^\varepsilon+1}{e^\varepsilon-1} Y_i' & w.p. \ \frac{1}{e^\varepsilon+1} \end{cases}$$

Return Z_i

Procedure: Analyzer $\mathcal{A}$

//Input : $\{Z_i\}_{i=1}^n$

//Output : estimator $\widehat{\mu}_n$

Return $\widehat{\mu}_n = \frac{1}{n}\sum_{i=1}^n Z_i$

for unbiasedness); otherwise flips the sign. Upon receiving Z_i, i.e., corrupted version of Y_i, the analyzer $\mathcal{A}$ first simply filters out the data if it is out of the bounded range and then returns the sample mean. The entire procedure of Algorithm 1 can be illustrated as follows.

$$X_i \xrightarrow{\text{Trunc.}(M)} \bar{X}_i \xrightarrow{\text{Random Rounding}} X_i' \xrightarrow{\text{Random Response}} Y_i \xrightarrow{\text{Corruption}} Z_i \xrightarrow{\text{Trunc.}(M\frac{e^\varepsilon+1}{e^\varepsilon-1})} \bar{Z}_i \xrightarrow{\text{Sample Mean}} \widehat{\mu}_n$$

Algorithm 2 for the CTL setting basically has the same local randomizer Q as in Algorithm 1 and now the analyzer $\mathcal{A}$ directly returns the sample mean. The performance guarantees of both algorithms are given in the following theorem. See Section A for the proof.

Proposition 2 (Upper Bounds)**.** *Consider the mean estimation problem under both LTC and CTL settings. Given any fixed $\delta \in (0,1)$, $\varepsilon \in [0,1]$, $\alpha \in (0,1/2)$ and $k > 1$, for any distribution $P \in \mathcal{P}_k$ and any α-Huber channel $C \in \mathcal{C}_\alpha$, Algorithms 1 and 2 satisfy that the mechanism Q is ε-LDP and each returned estimate $\widehat{\mu}_n$ guarantees that with probability at least $1-\delta$*

(i) For LTC: $|\widehat{\mu}_n - \mu(P)| \leq O\left(\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} + \left(\frac{1}{\varepsilon}\sqrt{\frac{\log(1/\delta)}{n}}\right)^{1-1/k}\right)$,

(ii) For CTL: $|\widehat{\mu}_n - \mu(P)| \leq O\left(\alpha^{1-1/k} + \left(\frac{1}{\varepsilon}\sqrt{\frac{\log(1/\delta)}{n}}\right)^{1-1/k}\right)$,

where (i) holds for $M = \min\left\{\left(\frac{\varepsilon}{\alpha}\right)^{1/k}, \left(\frac{\varepsilon\sqrt{n}}{\sqrt{\log(1/\delta)}}\right)^{1/k}\right\}$ and for all $n \geq 3\log(1/\delta)/\alpha$ when $\alpha > 0$, otherwise for all n and $M = \left(\frac{\varepsilon\sqrt{n}}{\sqrt{\log(1/\delta)}}\right)^{1/k}$; (ii) holds for $M = \min\left\{\left(\frac{n}{\log(1/\delta)}\right)^{1/k}, \left(\frac{1}{\alpha}\right)^{1/k}, \left(\frac{\varepsilon\sqrt{n}}{\sqrt{\log(1/\delta)}}\right)^{1/k}\right\}$ and $n \geq 3\log(1/\delta)/\alpha$ when $\alpha > 0$, otherwise for $n \geq \log(1/\delta)$ and $M = \min\left\{\left(\frac{n}{\log(1/\delta)}\right)^{1/k}, \left(\frac{\varepsilon\sqrt{n}}{\sqrt{\log(1/\delta)}}\right)^{1/k}\right\}$.

Remark 6 (Burn-in period). For both estimators, when $\alpha > 0$, the concentration kicks in when the sample size n is larger than a threshold. This type of burn-in period also exists in previous works on concentration under the Huber model, though in different contexts (e.g., non-private case [CKMY22] or central model of DP [WZTW23]) or with different estimators (e.g., trimmed mean [MTCD21]).

Remark 7 (Random response vs. Laplace mechanism). One may wonder if the standard Laplace mechanism can be applied in replace of the random response for ε-LDP in Q. The answer depends on the setting and the analyzer $\mathcal{A}$. For CTL, one can still derive a similar optimal concentration bound as in Proposition 4 by the concentration of Laplace noise. On the other hand, for LTC, simply replacing random response with Laplace mechanism in Q will lead to an additional $\log(1/\alpha)$ factor. This aligns with the fact that truncation-based estimators even cannot achieve optimal mean estimation for Gaussians under the Huber model [DK23]. The above discussion indicates another difference between LTC and CTL, i.e., the choice of ε-LDP mechanisms.

As two interesting applications of our mean estimation results, we will study both online MABs and offline MABs in the next two sections, highlighting again the sharp differences between LTC and CTL settings, in terms of both regret and sub-optimality performance.

6 Online MABs

As the first application of our mean estimation result, we study online MABs with the constraints of LDP as well as robustness with respect to Huber corruption and possibly heavy-tailed rewards.

Let us start with the following lower bound on the minimax clean regret (cf. Def. 5) for online MABs under both LTC and CTL settings. See Section B for the proof.

Proposition 3 (Regret Lower bounds)**.** *Consider the online MABs under LTC and CTL settings. Let $\varepsilon \in [0,1]$, $\alpha \in [0,1/2)$, $k > 1$ and T be large enough. Then, the minimax clean regrets satisfy the following results.*

(i) For LTC: $R^*_{\mathrm{LTC}}(k,\varepsilon,\alpha,T) \geq \Omega\left(T \cdot \left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} + T^{\frac{k+1}{2k}} \left(\frac{K}{\varepsilon^2}\right)^{\frac{k-1}{2k}}\right)$;

(ii) For CTL: $R^*_{\mathrm{CTL}}(k,\varepsilon,\alpha,T) \geq \Omega\left(T \cdot \alpha^{1-1/k} + T^{\frac{k+1}{2k}} \left(\frac{K}{\varepsilon^2}\right)^{\frac{k-1}{2k}}\right)$.

Remark 8 (Comparisons to related work). We first remark that [TWZW22] studied a similar case but without corruption and established a lower bound on the order of $\Omega\left(\left(\frac{K}{\varepsilon^2}\right)^{1-1/k} T^{1/k}\right)$ (for $k \in (1,2]$ when adapted to our setting), which is weaker with respect to T compared to our lower bound. In [TWZW22], the authors also claimed to achieve their lower bound via some arm-elimination algorithm, which becomes ungrounded given our lower bound, since for a large enough T, our lower bound is even larger than their upper bound for fixed ε, k and K. Another recent work [WZTW23] also studies online MABs with both privacy and Huber corruption but under the (weaker) central model of DP. In particular, the true reward from each user may be first corrupted before being observed by the central learner, who is then responsible for taking care of privacy guarantees. That is, the central learner has access to users' raw (corrupted) data rather than only a private view of data as in our LDP case. Under this strictly weaker privacy model, the authors in [WZTW23] establish the following lower bound on the minimax clean regret: $\Omega\left(\sqrt{KT} + (K/\varepsilon)^{1-\frac{1}{k}} T^{\frac{1}{k}} + T\alpha^{1-\frac{1}{k}}\right)$. Compared to our CTL setting, one can see that our stronger LDP privacy incurs a larger privacy cost in the regret.

Algorithm 3 UCB for online MABs under LTC and CTL

1: **Input:** private and robust mean estimator $\widehat{\mu}_n(k, \varepsilon, \alpha, \delta)$, positive constants c

2: **Initialize:** For each arm $a \in [K]$, let $\widehat{\mu}_{a,s}(t)$ be the estimate $\widehat{\mu}_s(k, \varepsilon, \alpha, t^{-4})$ based on the first s observed values of $Z_{a,1}, \ldots, Z_{a,s}$ of the rewards for arm a; $N_a(t)$ be the number of pulls of arm i by the beginning of time t.

3: **for** $t \in [T]$ **do**

4:   **if** $\exists a \in [K], N_a(t) \le 6 \log(t)/\alpha$ **then**

5:     $a_t = a$

6:   **else**

7:     Let $\beta_a(t) = \begin{cases} c\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} + c\left(\frac{1}{\varepsilon}\sqrt{\frac{\log(t^4)}{N_a(t)}}\right)^{1-1/k} & \text{for LTC} \\ c\alpha^{1-1/k} + c\left(\frac{1}{\varepsilon}\sqrt{\frac{\log(t^4)}{N_a(t)}}\right)^{1-1/k} & \text{for CTL} \end{cases}$

8:     Let $\mathrm{UCB}_a(t) = \widehat{\mu}_{a,N_a(t)}(t) + \beta_a(t)$

9:     $a_t = \mathrm{argmax}_{a \in [K]} \mathrm{UCB}_a(t)$

10:   **end if**

11: **end for**

Now, let us turn to our proposed algorithm (i.e., Algorithm 3) for achieving matching regret upper bounds (up to log factor). Algorithm 3 is a variant of upper confidence bound (UCB)-based algorithm (cf. [ACF02]), which computes the UCB index for each arm at each round $t \in [T]$ and then selects the one with the highest UCB, i.e., optimism in the face of uncertainty. To construct a valid UCB, we resort to our mean estimation results in the last section. In particular, we will need the two estimators (Algorithms 1 and 2) to compute the private and robust sample mean $\widehat{\mu}_{a,N_a(t)}(t)$ for each arm $a \in [K]$ at each round t. Then, the bonus term (i.e., radius of the confidence bound) $\beta_a(t)$ comes from the high probability mean estimation error established in Proposition 2. Note that due to burn-in period of the concentration results, Algorithm 3 has an additional exploration period to guarantee that the number of arm pulls is larger than a threshold (line 4).

The following proposition formally states the regret guarantees of Algorithm 3 with the proof given in Section C.

Proposition 4 (Regret Upper Bounds)**.** *Consider the online MABs under LTC and CTL settings. Let $\varepsilon \in [0, 1]$, $\alpha \in (0, 1/2)$, $k > 1$ and T be large enough. Then, for any $1/2 > \bar{\alpha} \ge \alpha$, the expected clean regret of Algorithm 3 satisfies the following guarantees.*

(i) For LTC: $R_{\mathrm{LTC}}(k, \varepsilon, \alpha, T) \le O\left(T\left(\frac{\bar{\alpha}}{\varepsilon}\right)^{1-1/k} + \left(\frac{K \log T}{\varepsilon^2}\right)^{\frac{k-1}{2k}} T^{\frac{k+1}{2k}} + \frac{K \log T}{\bar{\alpha}}\right)$;

(ii) For CTL: $R_{\mathrm{CTL}}(k, \varepsilon, \alpha, T) \le O\left(T\left(\bar{\alpha}\right)^{1-1/k} + \left(\frac{K \log T}{\varepsilon^2}\right)^{\frac{k-1}{2k}} T^{\frac{k+1}{2k}} + \frac{K \log T}{\bar{\alpha}}\right)$.

Remark 9 ($k = \infty$). It is worth comparing our result to the very recent similar result in [CEM23], where the authors present regret for linear bandits under LTC setting when $k = \infty$[3]. Their result is worse than ours when reduced to MAB, as the scaling with respect to α is $\sqrt{\alpha}$ in the first liner term rather than our α. Another minor difference is that our algorithm is anytime while theirs is not.

[3]Note that even though [CEM23] states their result for any sub-Gaussian reward. However, their current privacy guarantee (cf. their Eq. 15) only holds for the bounded case.

Remark 10 (No-corruption case). We remark that when $\alpha = 0$, with a direct modification of the burn-in period, the regret bound only has the second term $O\left(\left(\frac{K \log T}{\varepsilon^2}\right)^{\frac{k-1}{2k}} T^{\frac{k+1}{2k}}\right)$.

Remark 11 (Small α). When the α is very small but not equal to zero, one can tune the choice of $\bar{\alpha}$ (hence truncation threshold M) to balance the first and third term in the bound. Similar comments have been made in related work as in [CKMY22; WZTW23].

Remark 12 (Problem-dependent bounds). Although we mainly focus on minimax regret (i.e., problem-independent bound) in this paper, under some conditions of corruption level and the minimum mean gap, Algorithm 3 is also able to offer some problem-dependent bounds (Section C).

7 Offline MABs

In this section, we turn to the problem of offline MABs as another application of our high probability mean estimation results developed in Section 5. Let us start with the following result on the lower bound on the sub-optimality with its proof given in Section D.

Proposition 5 (Sub-optimality Lower Bounds)**.** *Consider the offline MABs under LTC and CTL settings. Let $\varepsilon \in [0, 1]$, $\alpha \in [0, 1/2)$, $k > 1$ and N be large enough. Then, for $\beta^\star \geq 2$, the minimax expected sub-optimality satisfies the following results.*

(i) For LTC: $\mathrm{SubOpt}^*_{\mathrm{LTC}}(\beta^\star, k, \varepsilon, \alpha, N) \geq \Omega\left(\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} + \left(\frac{1}{\varepsilon}\sqrt{\frac{\beta^\star}{N}}\right)^{1-1/k}\right)$;

(ii) For CTL: $\mathrm{SubOpt}^*_{\mathrm{CTL}}(\beta^\star, k, \varepsilon, \alpha, N) \geq \Omega\left(\alpha^{1-1/k} + \left(\frac{1}{\varepsilon}\sqrt{\frac{\beta^\star}{N}}\right)^{1-1/k}\right)$.

Remark 13. To the best of our knowledge, this is the first result on offline MABs with heavy-tailed rewards, even without privacy and corruption. As $k \to \infty$, our lower bound also offers new results for offline MABs with bounded data under privacy and corruption. It can be seen that LTC setting is again more difficult compared to the CTL setting, i.e., α/ε vs. α, due to the interplay of privacy and Huber corruption.

Now, let us turn to our proposed algorithm (Algorithm 4), which is able to achieve a matching expected sub-optimality (up to log factor) for both LTC and CTL settings. Algorithm 4 is a simple variant of the classic Lower Confidence Bound (LCB)-based algorithm as in [RZMJR21], i.e., pessimism in the offline setting. The key difference compared to [RZMJR21] is our new private and robust estimator (line 8) and penalty term (line 9), which come from our high probability mean estimation error. Another modification is due to our burn-in period of concentration result (line 4). Putting all of these together, Algorithm 4 is able to achieve the following guarantees on the expected sub-optimality, with the proof given in Section E.

Proposition 6 (Sub-optimality Upper Bounds)**.** *Consider the offline MABs under LTC and CTL settings. Let $\varepsilon \in [0, 1]$, $\alpha \in (0, 1/2)$, $k > 1$ and $\delta = 1/N$. Then, for all finite $\beta^\star \geq 1$ and N large enough, the expected sub-optimality of Algorithm 4 satisfies the following guarantees.*

(i) For LTC: $\mathrm{SubOpt}_{\mathrm{LTC}}(\beta^\star, k, \varepsilon, \alpha, N) \leq O\left(\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} + \left(\frac{1}{\varepsilon}\sqrt{\frac{\beta^\star \log(KN)}{N}}\right)^{1-1/k}\right)$;

Algorithm 4 LCB for offline MABs under LTC and CTL

Input: Offline private and corrupted data $\mathcal{D} = \{(a, Z^{(a)})\}_{a\in[K]}$, $\delta \in (0,1)$, $k > 1$, $\varepsilon \in [0,1]$ and $\alpha \in (0, 1/2)$, private and robust mean estimator $\widehat{\mu}_n(k, \varepsilon, \alpha, \delta)$, positive constant c

Initialize: $N_a = |Z^{(a)}|$ for all $a \in [K]$, i.e., number of pulls for arm a in $\mathcal{D}$

for $a \in [K]$ **do**

  if $N_a < 3\log(1/\delta)/\alpha$ **then**

    Set the empirical mean reward $\widehat{r}(a) = 0$

    Set the penalty $b(a) = 1$

  else

    Compute the empirical mean reward $\widehat{r}(a) = \widehat{\mu}_{N_a}(k, \varepsilon, \alpha, \delta)$

    Compute the penalty $b(a) = \begin{cases} c\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} + c\left(\frac{1}{\varepsilon}\sqrt{\frac{\log(2K/\delta)}{N_a}}\right)^{1-1/k} & \text{for LTC} \\ c\alpha^{1-1/k} + c\left(\frac{1}{\varepsilon}\sqrt{\frac{\log(2K/\delta)}{N_a}}\right)^{1-1/k} & \text{for CTL} \end{cases}$

  end if

end for

Return $\widehat{a} = \mathrm{argmax}_{a\in[K]}\, \widehat{r}(a) - b(a)$

(ii) For CTL: $\mathrm{SubOpt}_{\mathrm{CTL}}(\beta^\star, k, \varepsilon, \alpha, N) \le O\left(\alpha^{1-1/k} + \left(\frac{1}{\varepsilon}\sqrt{\frac{\beta^\star \log(KN)}{N}}\right)^{1-1/k}\right).$

Remark 14 (No-corruption case). For the case of $\alpha = 0$, as before one can simply choose to use the mean estimate result for $\alpha = 0$ as shown in Proposition 2 and adjust the burn-in period accordingly. This will lead to a bound that only has the second term in the above upper bounds.

Remark 15 (Almost optimal for $\beta^\star \ge 2$). For $\beta^\star \ge 2$, one can observe that the upper bound of Algorithm 4 almost matches the lower bounds in Proposition 5 for both LTC and CTL settings. However, when $\beta^\star \in [1, 2)$ (i.e., good coverage case), it is known that the performance of LCB is worse than imitation learning, i.e., simply returning the most frequently selected arm in the offline dataset (when there is no privacy and corruption) [RZMJR21]. We leave it to future work to give a tight characterization of the sub-optimality when $\beta^\star \in [1, 2)$.

Remark 16 (High-probability sub-optimality bound). In fact, the proof of Proposition 6 also naturally gives us high-probability bounds without specifying $\delta = 1/N$ in the end.

8 Simulations

In this section, we conduct numerical simulations to assess the performance of our algorithms in three problems (i.e., mean estimation, online MABs and offline MABs), under both LTC and CTL settings. Recall that our performance metrics for all three problems are minimax ones, which capture the worst-case performance. As a result, we are particularly interested in the following two questions in our simulations:

(i) *Can we simulate the worst-case scenario and test the performance of our proposed algorithms?*

(ii) *How about their performance in non-worst-case scenarios?*

Note that (i) essentially sheds further light on how to design the most powerful adversary Huber corruption model, which in turn could explain the separation result between LTC and CTL from the perspective of attacking. On the other hand, (ii) would help to illustrate our algorithms' performance in some mild/real-world non-adversary Huber corruption. For example, although the minimax regret for online MABs has a linear term in the worst case, the actual performance under the non-adversary corruption model can be sub-linear as we will show later.

8.1 Mean estimation

We start with the worst-case scenario for the mean estimation under a large sample size regime where the minimax error rate is dominated by the corruption part, i.e., the separation result $(\alpha/\varepsilon)^{1-1/k}$ under LTC vs. $\alpha^{1-1/k}$ under CTL. To this end, we need to design the most powerful adversary corruption for both LTC and CTL. Here, we allow the (white-box) adversary to choose inlier distribution over X and can adaptively choose Huber corruption distribution based on inlier distribution and the knowledge of our algorithm, e.g., LDP mechanism Q in the LTC setting.

In particular, the adversary chooses the following inlier distribution:

$$P(X = 1/\gamma) = \frac{1}{2}\gamma^k, \quad P(X = -1/\gamma) = \frac{1}{2}\gamma^k, \quad P(X = 0) = 1 - \gamma^k \tag{5}$$

where $\gamma = (\alpha/\varepsilon)^{1/k}$ under LTC and $\gamma = (\alpha)^{1/k}$ under CTL. One can clearly see that $\mathbb{E}\left[|X|^k\right] \le 1$ for all $k > 1$, hence $P \in \mathcal{P}_k$ for any $k > 1$ and $\alpha \le \varepsilon$. Moreover, we have $\mathbb{E}\left[X\right] = 0$.

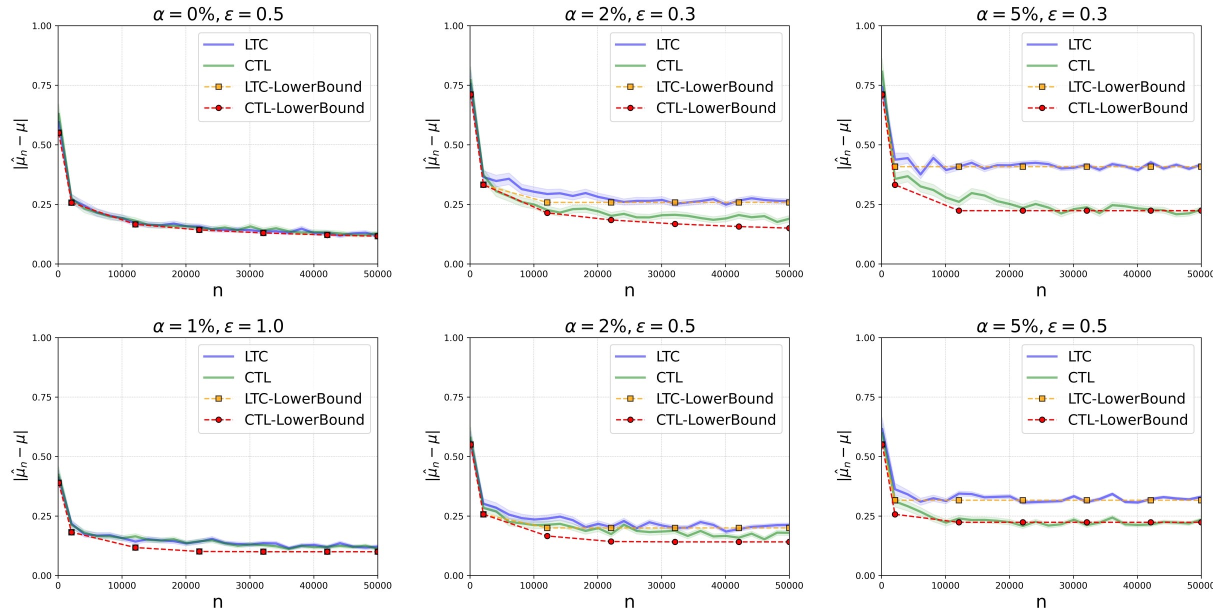
Now, we first consider the following strong Huber corruption model.

Definition 7 (Strong Huber corruption for mean estimation)**.** Let the inlier distribution over X be given by (5). Under LTC: for each input Y_i, with probability α, replace it with $M \cdot \frac{e^\varepsilon+1}{e^\varepsilon-1}$; Under CTL: for each input X_i, with probability α, replace it with M;

Note that, the white-box adversary knows our algorithm and hence M. We are going to show that no matter how large the sample size is, the mean error has to be large for both LTC and CTL under the above strong Huber corruption.

Let us start with CTL and consider the sample size n to be large. Then, according to Algorithm 2, $M = (1/\alpha)^{1/k} = 1/\gamma$, which leads to the fact that the mean of Y is now $\alpha M = \alpha^{1-1/k}$ (note $\mathbb{E}\left[X\right] = 0$). Then, our estimator will essentially at best return the mean of Y, hence leading to the error of $\Omega(\alpha^{1-1/k})$. For LTC, with the choice of γ and M, we also have $M = 1/\gamma$. By our design of LDP mechanism Q in Algorithm 1, the mean of Y is still zero and hence after the corruption, the mean of Z becomes $\alpha \cdot M\frac{e^\varepsilon+1}{e^\varepsilon-1}$, which is the best outcome of our estimator, hence the error of $\Omega((\alpha/\varepsilon)^{1-1/k})$. Note that in both cases, the choice of corruption distribution needs care (i.e., adaptation to our algorithm), since otherwise, our estimator may still have an accurate estimate, as some other outlier values can be simply filtered out by our algorithm. More importantly, an alternative explanation of our separation result becomes evident: under LTC, the error is larger because the adversary has the capability to select a corruption value that is magnified by a factor of $1/\varepsilon$.

In our experiments, we choose $k = 2$ and consider various corruption level $\alpha \in \{0, 0.02, 0.05\}$ and privacy budget $\varepsilon \in \{0.3, 0.5, 1\}$. For each set of parameters, we conduct 300 runs and plot the average of the estimation error and corresponding confidence region. Fig. 2 illustrates our simulation results under strong Huber corruption in Definition 7. A common pattern behind all the plots in Fig. 2 is that due to strong corruption, the estimation error will only converge to a plateau and almost match the lower bounds.



(a) $\alpha = 0$ or $\varepsilon = 1$ (b) $\alpha = 2\%$, vary ε (c) $\alpha = 5\%$, vary ε

Figure 2: Mean estimation error with strong Huber corruption in Definition 7 under LTC and CTL settings.

Specifically, from the two plots in column (a), we see that when $\alpha = 0$ or $\varepsilon = 1$, the performance under LTC and CTL is close (i.e., no-separation), which aligns with our theoretical results. In the two plots of column (b), we see that LTC has a larger error than CTL and as ε decreases (i.e., stronger privacy), the difference becomes larger, which matches our theoretical separation results. Finally, comparing the plots in column (c) with those in (b), we see that as the corruption level increases, the performance becomes worse.

We also consider the following weak corruption model, which simply flips the sign of the data.

Definition 8 (Weak Huber corruption for mean estimation)**.** Let the inlier distribution over X be given by (5). Under LTC: for each input Y_i, with probability α, replace it with $-Y_i$; Under CTL: for each input X_i, with probability α, replace it with $-X_i$;

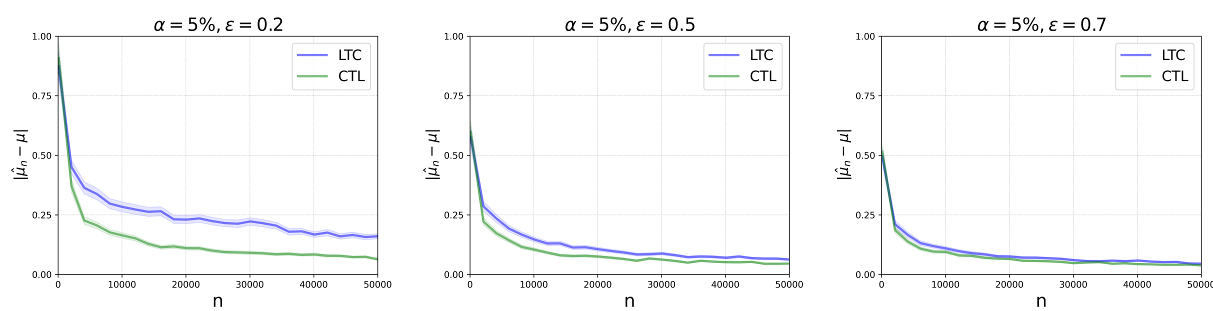
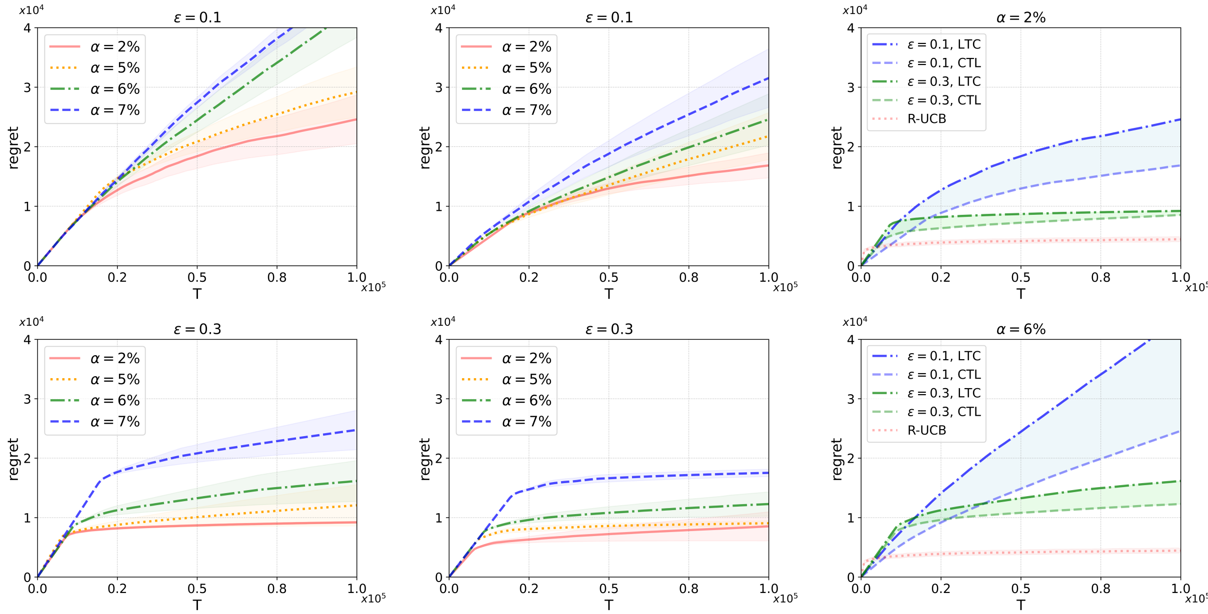


Figure 3: Mean estimation error with weak Huber corruption in Definition 8 under LTC and CTL settings

In Fig.3, we can see that under weak Huber corruption, the estimation error under our estimators can indeed decrease as the sample size increases. This demonstrates that in some real-world mild corruption



(a) LTC (b) CTL (c) Comparison of regret

Figure 4: Regret performance with Huber corruption in Definition 9 under LTC and CTL settings.

scenarios, our estimators can yield promising performance.

8.2 Online MABs

In this section, we first consider some classic heavy-tailed distributions under non-adversary corruption. The main purpose is to show that our proposed algorithm (i.e., Algorithm 3) can indeed achieve sublinear regret under certain scenarios. Moreover, our simulations will also provide some insights into our proof.

Settings. As in previous works [TWZW22; WZTW23], we consider Pareto distribution, whose probability distribution is given by

$$f(x; x_m, s) = \begin{cases} \frac{s x_m^s}{x^{s+1}}, & \text{if } x \geq x_m \\ 0, & \text{otherwise} \end{cases}$$

where $s > 0$ is the shape parameter and $x_m > 0$ is the scale parameter. In our experiments, we consider there are $K = 10$ arms, and for each arm $i \in [K]$, the distribution is Pareto with $x_m = i$ and $s = 11$. To ensure that each arm's reward distribution is in $\mathcal{P}_k$ (i.e., $\mathbb{E}_{X \sim P}[|X|^k] \leq 1$), we normalize the reward by the k-th moment, which is $\frac{s x_m^k}{s-k}$. Consequently, the mean of each arm is $\frac{s-k}{x_m^{k-1}(s-1)}$. We consider $k = 2$, which along with our choices of s and x_m, yields that arm 1 is the best arm with a mean of 0.9 while arm 10 is the worst arm with a mean of 0.09. For the corruption, we consider the following Huber model.

Definition 9 (Huber corruption for online/offline MABs)**.** Let each arm's inlier distribution be Pareto with the parameters described above. Under LTC, for each private view of reward from each $a \in [K]$, with probability α, replace it with $M \cdot \frac{e^{\varepsilon}+1}{e^{\varepsilon}-1}$. Under CTL, for each raw reward from each arm $a \in [K]$, with probability α, replace it with M.

Remark 17. It is worth noting that even though the above corruption values are the same as in Definition 7, it is not necessarily the worst-case as the inliers are now Pareto. That is, even after corruption, the agent can possibly still distinguish between different arms. We also consider strong corruption cases where after corruption, the agent cannot distinguish the distributions of two arms, hence a linear regret, see Fig. 5 and Appendix F for details.

Fig. 4 illustrates the regret performance of our proposed algorithm (i.e., Algorithm 3) for online MABs under LTC and CTL settings, with the specific corruption given by Definition 9. The two plots in column (a) capture the LTC setting while the two plots in column (b) denote the CTL setting. In both settings, we can see that for small corruption level α, our algorithm can achieve sublinear regret, even though in the *worst-case* our minimax bounds are linear. In column (c), we also directly compare the regret performance under LTC and CTL with different sets of parameters of α and ε. As expected, the regret performance under LTC is worse than that under CTL, and as α increases or ε decreases, the gap becomes larger. This demonstrates separation results in terms of actual performance rather than only in terms of theoretical upper bounds. As a baseline, we also compare with one classic robust MAB algorithm under heavy-tailed rewards proposed in [BCL13].

Note that our purpose in this section is not to demonstrate the superior performance of our proposed algorithm over all existing robust or/and private algorithms (given a large number of different existing ones). Rather, one of the goals is to use simulations to highlight the separation between LTC and CTL. Another important goal is to provide more insights into our proof of the regret upper bounds. Specifically, in our proof of the LTC setting (similar in CTL setting), we will divide the set of all sub-optimal arms $\mathcal{G}$ into two groups $\mathcal{G}_1$ and $\mathcal{G}_2$ where $\mathcal{G}_2 = \{a \in [K] \setminus a^* : c \left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} \geq \frac{1}{2}\Delta_a\}$ for some constant c. Then, we argue that if $\mathcal{G}_2$ is empty, then one can still derive the standard logarithmic problem-dependent regret bound. This can also be somehow validated partially by our simulation results. In particular, under our problem instances described above, when $\alpha = 0.02$, $\varepsilon = 0.1$, and $c = 0.5$, we have $|\mathcal{G}_2| = 0$ under LTC (i.e., no sub-optimal arms in $\mathcal{G}_2$). In this case, as illustrated in the top plot of column (a) in Fig. 4, we can observe logarithmic order regret. This naturally extends to the larger ε case, as illustrated in the bottom plot in column (a).

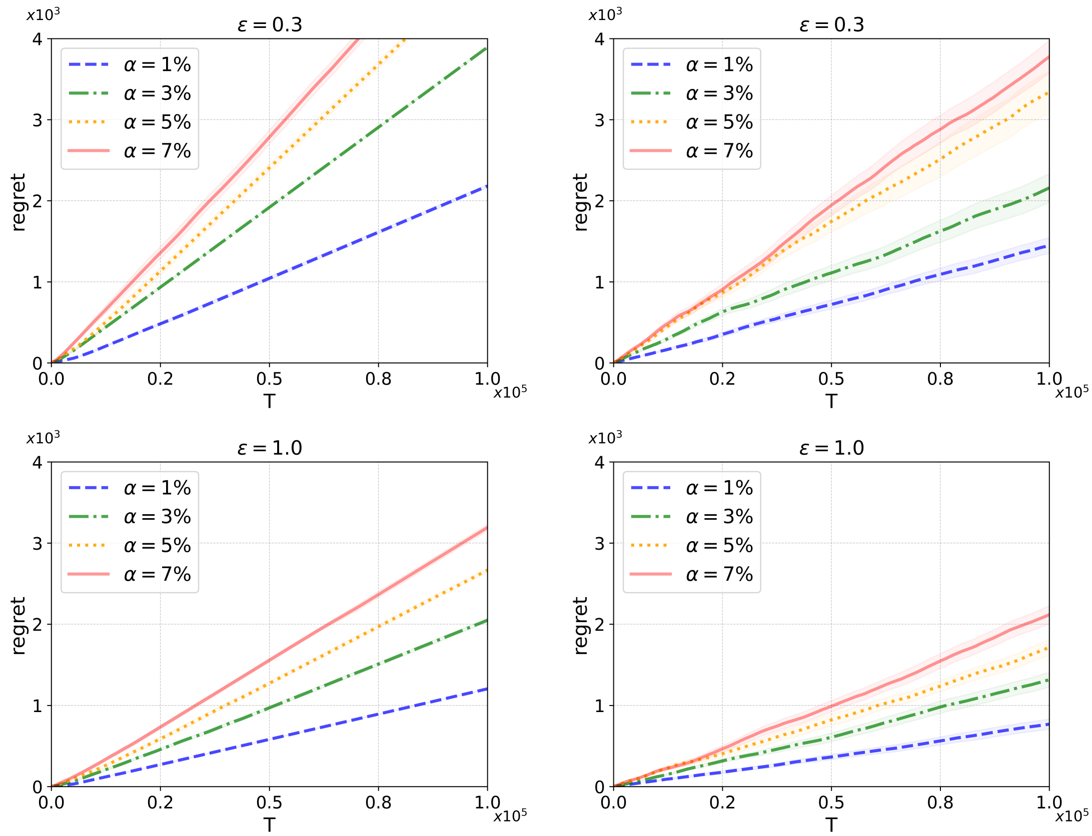
As mentioned above, we also create a strong Huber corruption for online MABs, see Appendix F for details. In this case, the regret becomes linear which matches our minimax lower bound,

8.3 Offline MABs

In the offline case, the analyzer/agent is given a batch of pre-collected data with private and corrupted view. In our experiments, we again consider the case that there are $K = 10$ arms and each arm's raw reward distribution is Pareto with the same parameters as in the online case. For corruption, we again consider the one given by Definition 9.

One difference here is that we need to specify the behavior policy π that is used to collect the data. To this end, we consider the following policy π in our simulation results: for each sample size N, we pulled the best arm (i.e., arm 1) $\frac{N}{3}$ times and each other arm $i \neq 1$ uinformly, i.e., $\frac{2N}{3(K-1)}$ times. That is, roughly speaking, we approximately have $1/\pi(a^\star) = 3$, which aligns with our theoretical assumption (i.e., the finite concentrability coefficient $\beta^\star \geq 2$ when our upper bounds are tight in minimax sense).

Fig. 6 illustrates the suboptimality of our algorithm (i.e., Algorithm 4) under both LTC and CTL settings. We can see that in both settings, the sub-optimality could approach zero under several values of privacy parameters. This again highlights that under mild/non-adversary corruption, the algorithm could yield reasonably good performance, rather than the pessimistic worst-case one. Also, we observe that even in this non-adversary corruption case, suboptimality under LTC in general is still worse than that under CTL. Finally,



(a) LTC, vary α (b) CTL, vary α

Figure 5: Regret performance with strong Huber corruption given by Appendix F under LTC and CTL settings.

it is not surprising that for both LTC and CTL, as α increases or ε decreases, sub-optimality will increase.

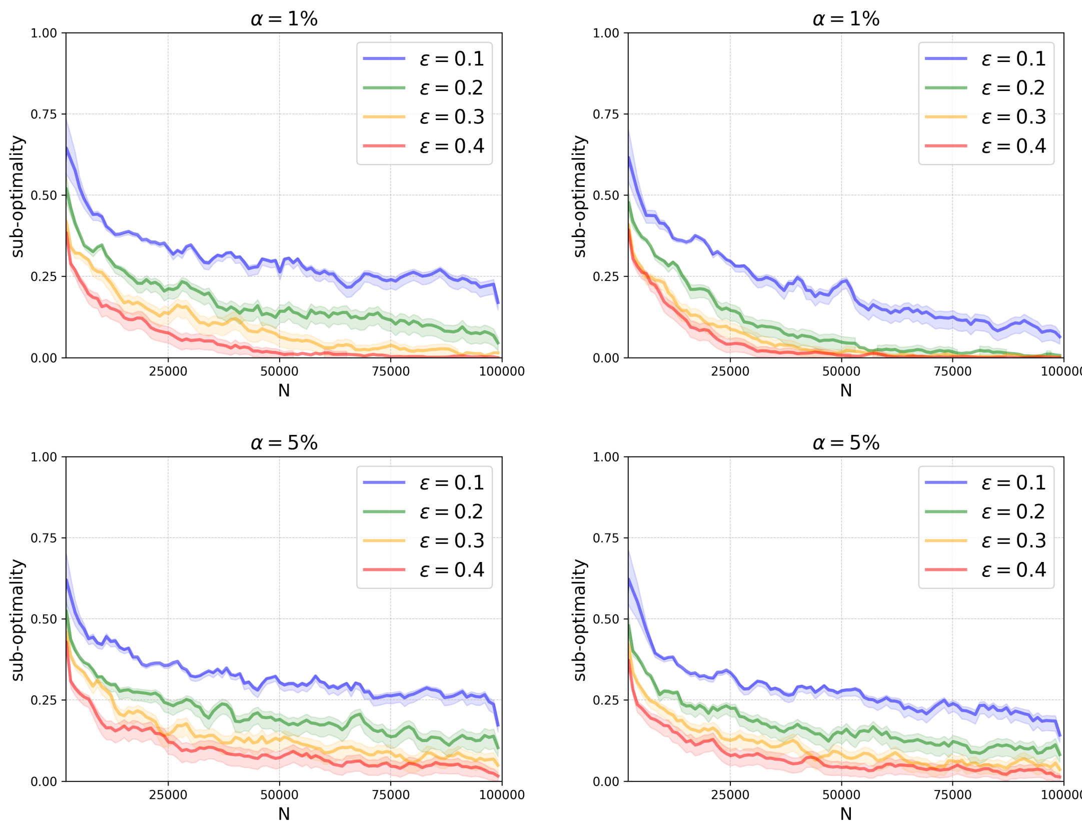
9 Proof

In the main paper, we will provide proof for the lower bound of mean estimation, i.e., Proposition 1, which will help to gain more intuition about the separation between LTC and CTL. Once armed with this intuition, it will be easier to follow the rest proofs of other results, which are all relegated to the Appendix due to space limitations.

9.1 Proof of Proposition 1

Proof. We first focus on the LTC setting and divide the proof into two steps.

Step 1: Without corruption. By definition, it suffices to establish a lower bound on the concentration even without corruption. That is, under LTC, $Z_i = Y_i$ for all $i \in [n]$. This will give us the second term in the bound.



(a) LTC, vary ε

(b) CTL, vary ε

Figure 6: Suboptimality performance with Huber corruption in Definition 9 under LTC and CTL settings.

Consider the following two distributions P and P'. Let $\gamma > 0$, specified later and

$$P(X = 1/\gamma) = \gamma^k, \quad P(X = 0) = 1 - \gamma^k$$
$$P'(X = 1/\gamma) = 1/2 \cdot \gamma^k, \quad P'(X = 0) = 1 - 1/2 \cdot \gamma^k. \tag{6}$$

It is easy to see that for both P, P', $\mathbb{E}\left[|X|^k\right] \le 1$ for all $k > 1$, hence $P, P' \in \mathcal{P}_k$ for any $k > 1$. Moreover, we have $|\mu(P) - \mu(P')| = 1/2 \cdot \gamma^{k-1}$ and $\mathrm{TV}\left(P, P'\right) = 1/2 \cdot \gamma^k$. For any ε-LDP channel Q, let M and M' be the induced marginal distribution from P and P', respectively. That is, for $i \in [n]$, $Y_i \sim M$ and $Y_i' \sim M'$. Let $Y_{[n]} = \{Y_i\}_{i=1}^n$ and $Y'_{[n]} = \{Y'_i\}_{i=1}^n$, i.e., $Y_{[n]} \sim M^{\otimes n}$ and $Y'_{[n]} \sim M'^{\otimes n}$.

The high-level idea behind our proof is as follows: Given any sample size n, if there exists at least probability 2δ such that $Y_{[n]} = Y'_{[n]}$, then one has to incur $\Omega(\gamma^{k-1})$ estimation error with probability δ. This naturally reminds us to think about maximal coupling, since it maximizes the probability that $Y_{[n]} = Y'_{[n]}$ and is also closely related to TV distance. In particular, we have the following textbook facts.

Lemma 1. *Let P_1 and P_2 be two distributions on $\mathcal{X}$ that share the same σ-algebra. There exists a coupling $\omega^*(P_1, P_2)$, which is a distribution over $\mathcal{X}^2$ such that*

$$\mathbb{P}_{(X_1,X_2)\sim\omega^*(P_1,P_2))}(X_1 \neq X_2) = \mathrm{TV}(P_1, P_2)$$
$$\forall S \text{ measurable}, \mathbb{P}_{(X_1,X_2)\sim\omega^*(P_1,P_2))}(X_1 \in S) = P_1(X_1 \in S)$$
$$\forall S \text{ measurable}, \mathbb{P}_{(X_1,X_2)\sim\omega^*(P_1,P_2))}(X_2 \in S) = P_2(X_2 \in S).$$

This coupling is called maximal coupling.

Based on this fact, fix some n, if $(Y_{[n]}, Y'_{[n]})$ is sampled from the maximal coupling $\omega^*(M^{\otimes n}, M'^{\otimes n})$, then we know that there exists a probability $p = 1 - \mathrm{TV}(M^{\otimes n}, M'^{\otimes n})$ such that $Y_{[n]} = Y'_{[n]}$. To lower bound p, we need to upper bound the TV distance. To this end, we will leverage Bretagnolle–Huber inequality and strong data processing inequality (i.e., Corollary 3 in [DJW18]). In particular, we have

$$\begin{aligned}
\mathrm{TV}(M^{\otimes n}, M'^{\otimes n}) &\overset{(a)}{\le} 1 - \frac{1}{2}\exp\left(-\mathrm{KL}\left(M^{\otimes n} \middle\| M'^{\otimes n}\right)\right) \\
&\overset{(b)}{=} 1 - \frac{1}{2}\exp\left(-4(e^\varepsilon - 1)^2 \cdot n \cdot (\mathrm{TV}\left(P, P'\right))^2\right) \\
&= 1 - \frac{1}{2}\exp\left(-4(e^\varepsilon - 1)^2 \cdot n \cdot \gamma^{2k}\right) \\
&\overset{(c)}{\le} 1 - \frac{1}{2}\exp\left(-16\varepsilon^2 \cdot n \cdot \gamma^{2k}\right),
\end{aligned}$$

where (a) holds by Bretagnolle–Huber inequality; (b) holds by Corollary 3 in [DJW18]; (c) is true since $e^\varepsilon - 1 \le 2\varepsilon$ for $\varepsilon \in [0, 1]$. Thus, let $\gamma = c_1 \left(\frac{\sqrt{\log(1/\delta)}}{\varepsilon\sqrt{n}}\right)^{1/k}$ for some constant c. Then, for large enough n, $\gamma^{k-1} < 1$ and $\mathrm{TV}(M^{\otimes n}, M'^{\otimes n}) \le 1 - 2\delta$, which implies that with probability at least δ, the error is $\Omega(\gamma^{k-1}) = \Omega\left(\left(\frac{\sqrt{\log(1/\delta)}}{\varepsilon\sqrt{n}}\right)^{1-1/k}\right)$.

Step 2: Corruption part. Recall that under α-Huber, for each private view Y_i, it is independently corrupted with probability α, and when it happens, Z_i is sampled from an arbitrary noise distribution N; otherwise, $Z_i = Y_i$. To proceed, we will utilize the following useful fact.

Lemma 2 (Theorem 5.1 in [CGR18]). *Let R_1 and R_2 be two distributions on $\mathcal{X}$; If for some $\alpha \in [0, 1/2)$, we have that* $\mathrm{TV}(R_1, R_2) \leq \frac{\alpha}{1-\alpha}$*, then there exits two distributions N_1 and N_2 on the same probability space such that*

$$(1-\alpha)R_1 + \alpha N_1 = (1-\alpha)R_2 + \alpha N_2.$$

This result says that the Huber model with parameter α can corrupt two distributions that are close in TV distance so that the outputs are essentially sampled from the same distribution, hence indistinguishable.

Another fact we will leverage is that LDP mechanism is a "contraction" in that it will make the TV distance closer.

Lemma 3 (Corollar 2.9 in [KOV14]). *For any $\varepsilon > 0$, let Q be any ε-LDP mechanism. Then, for any pair of distributions P_1 and P_2, the induced marginals M_1 and M_2 satisfy*

$$\mathrm{TV}(M_1, M_2) \leq \frac{e^\varepsilon - 1}{e^\varepsilon + 1} \mathrm{TV}(P_1, P_2).$$

The above fact indicates that for $\varepsilon \in [0, 1]$, $\mathrm{TV}(M_1, M_2) \leq O(\varepsilon)\mathrm{TV}(P_1, P_2)$. With the above two facts, it suffices for us to find two distributions P and P' for X_i with a "large" mean difference, such that the induced marginal distributions for Y_i is $O(\alpha)$. To this end, we again consider the two distributions in (6) with a different choice of γ. Since $\mathrm{TV}(P, P') = 1/2 \cdot \gamma^k$, by Lemma 3, choosing $\gamma = c' \cdot (\alpha/\varepsilon)^{1/k}$ for some small constant $c' > 0$ yields that $\mathrm{TV}(M, M') \leq \alpha \leq \alpha/(1-\alpha)$. Hence, by Lemma 2, there exists Huber contamination such that it is impossible to distinguish the final outputs. Hence, with a probability of at least $1/2$, the error is $\Omega\left(\gamma^{k-1}\right) = \Omega\left((\alpha/\varepsilon)^{1-1/k}\right)$. We finally conclude that for any $\delta \in (0, 1/2)$, with probability at least δ, for all large enough n, estimation error is $\Omega\left(\gamma^{k-1}\right) = \Omega\left((\alpha/\varepsilon)^{1-1/k}\right)$. This finishes the proof for the LTC setting.

As for the CTL setting, the second term in the lower bound follows the same proof as in Step 1. The key difference lies in Step 2, i.e., the first term in the bound. In particular, since the contamination is before LDP, one can now only choose $\gamma = c'\alpha^{1/k}$, i.e., no "contraction" from LDP anymore. As a result, the estimation error is $\Omega\left(\gamma^{k-1}\right) = \Omega\left(\alpha^{1-1/k}\right)$. □

10 Conclusion

In this paper, we have demonstrated an interesting interplay between privacy and robustness in three problems: mean estimation, online and offline MABs. The punchline is that corruption after any LDP mechanism becomes easier, hence the same amount of corruption leading to worse performance compared to corruption before LDP mechanisms. Some interesting future directions include (i) designing estimators without knowledge of corruption level or heavy-tail parameters [JOR22]; (ii) improving the sub-optimal result for linear bandit in [CEM23] and generalizing it to other privacy models such as shuffle DP [CZ22b]; (iii) studying the case where the heavy-tailedness is characterized by the central moment rather than the raw moment currently considered in our paper.

A Proof of Proposition 2

Proof. Let us start with the LTC setting. As for privacy, it builds on the privacy guarantee of random response.

Privacy. By definition, we need to show that for any two inputs $x, x' \in \mathcal{X}$ and $y \in \left\{M\frac{e^\varepsilon+1}{e^\varepsilon-1}, -M\frac{e^\varepsilon+1}{e^\varepsilon-1}\right\}$

$$\frac{\mathbb{P}\left[Y = y|X = x\right]}{\mathbb{P}\left[Y = y|X = x'\right]} \leq e^\varepsilon.$$

Consider the case $y = M\frac{e^\varepsilon+1}{e^\varepsilon-1}$ and similar analysis applies to the other case. Let $P_{x\to M^+}$ be the probability that x is translated to M in our mechanism Q and $P_{x\to M^-}$ be the probability that x is translated to $-M$ in our mechanism Q. Similarly defines $P_{x'\to M^+}$ and $P_{x'\to M^-}$.

Thus, according to our Q in Algorithm 1 and let $P_\varepsilon := \frac{e^\varepsilon}{e^\varepsilon+1}$, we have

$$\mathbb{P}\left[Y = y|X = x\right] = P_{x\to M^+}P_\varepsilon + P_{x\to M^-}(1 - P_\varepsilon)$$
$$\mathbb{P}\left[Y = y|X = x'\right] = P_{x'\to M^+}P_\varepsilon + P_{x'\to M^-}(1 - P_\varepsilon)$$

As a result,

$$\frac{\mathbb{P}\left[Y = y|X = x\right]}{\mathbb{P}\left[Y = y|X = x'\right]} = \frac{P_{x\to M^+}P_\varepsilon + P_{x\to M^-}(1 - P_\varepsilon)}{P_{x'\to M^+}P_\varepsilon + P_{x'\to M^-}(1 - P_\varepsilon)} \leq \frac{P_\varepsilon}{1 - P_\varepsilon} \leq e^\varepsilon.$$

Utility. For the utility part, we will divide the proof into four steps.

We draw the following informal diagram for an illustration of Algorithm 1.

$$X_i \xrightarrow{\text{Trunc.}(M)} \bar{X}_i \xrightarrow{\text{Random Rounding}} X_i' \xrightarrow{\text{Random Response}} Y_i \xrightarrow{\text{Corruption}} Z_i \xrightarrow{\text{Trunc.}(M\frac{e^\varepsilon+1}{e^\varepsilon-1})} \bar{Z}_i \xrightarrow{\text{Sample Mean}} \widehat{\mu}_n$$

Step 1: Bound the number of corrupted points.

By Chernoff bound for the binomial distribution, we have that for $n \geq 3\log(1/\delta)/\alpha$

$$|\mathcal{B}| \leq 2\alpha n, \quad w.p. \quad 1 - \delta,$$

where $|\mathcal{B}|$ denotes the total number of corrupted ("bad") points. Let this event be $\mathcal{E}$, and in the following steps, we will condition on this event.

Step 2: Bound the distance $|\mathbb{E}\left[X_i'\right] - \mathbb{E}\left[X_i\right]|$.

$$\begin{aligned}
|\mathbb{E}\left[X_i\right] - \mathbb{E}\left[X_i'\right]| &\leq |\mathbb{E}\left[X_i\right] - \mathbb{E}\left[\bar{X}_i\right]| + |\mathbb{E}\left[\bar{X}_i\right] - \mathbb{E}\left[X_i'\right]| \\
&\overset{(a)}{=} |\mathbb{E}\left[X_i\right] - \mathbb{E}\left[\bar{X}_i\right]| + 0 \\
&\leq \mathbb{E}\left[|X_i|\mathbb{1}(|X_i| \geq M)\right] \\
&\overset{(b)}{\leq} \frac{1}{M^{k-1}}
\end{aligned}$$

where (a) holds by the property of random rounding. Recall that, for any $\bar{X}_i \in [-M, M]$, $X_i' = M$ w.p. $\frac{1+\bar{X}_i/M}{2}$ and $X_i' = -M$ w.p. $\frac{1-\bar{X}_i/M}{2}$. Thus, one can see $\mathbb{E}\left[X_i'|\bar{X}_i\right] = \bar{X}_i$, hence $\mathbb{E}\left[\bar{X}_i\right] = \mathbb{E}\left[X_i'\right]$; (b) holds by Hölder's inequality and the fact k-th moment of X_i is upper bounded by one.

Step 3: Bound the distance $|\mathbb{E}\left[X_i'\right] - \widehat{\mu}_n|$.

$$
\begin{aligned}
|\widehat{\mu}_n - \mathbb{E}\left[X_i'\right]| &= |\frac{1}{n}\sum_i \bar{Z}_i - \mathbb{E}\left[X_i'\right]| \\
&= |\frac{1}{n}\sum_i \bar{Z}_i - \frac{1}{n}\sum_i Y_i + \frac{1}{n}\sum_i Y_i - \mathbb{E}\left[X_i'\right]| \\
&\stackrel{(a)}{\leq} 2\alpha \cdot M \cdot \frac{e^\varepsilon + 1}{e^\varepsilon - 1} + |\frac{1}{n}\sum_i Y_i - \mathbb{E}\left[X_i'\right]| \\
&\stackrel{(b)}{\leq} 2\alpha \cdot M \cdot \frac{e^\varepsilon + 1}{e^\varepsilon - 1} + O\left(M \cdot \frac{e^\varepsilon + 1}{e^\varepsilon - 1} \cdot \sqrt{\frac{\log(1/\delta)}{n}}\right) \quad w.p. \quad 1 - \delta
\end{aligned}
$$

where (a) holds by triangle inequality, the even $\mathcal{E}$ in step 1, and the fact that $\bar{Z}_i$, Y_i are both bounded; (b) holds by Hoeffding inequality. Note that $Y_i = \frac{e^\varepsilon+1}{e^\varepsilon-1}X_i'$ w.p. $\frac{e^\varepsilon}{e^\varepsilon+1}$ and $Y_i = -\frac{e^\varepsilon+1}{e^\varepsilon-1}X_i'$ w.p. $\frac{1}{e^\varepsilon+1}$. That is, $\mathbb{E}\left[Y_i\right] = \mathbb{E}\left[X_i'\right]$ and $Y_i = \{M \cdot \frac{e^\varepsilon+1}{e^\varepsilon-1}, -M \cdot \frac{e^\varepsilon+1}{e^\varepsilon-1}\}$.

Step 4: Put the above two parts together.

For any $\varepsilon \in [0,1]$, any $\delta \in (0,1)$ and any $P \in \mathcal{P}_k$, we have with probability at least $1-\delta$,

$$
|\widehat{\mu}_n - \mu(P)| \leq O\left(\frac{1}{M^{k-1}} + \frac{\alpha M}{\varepsilon} + \frac{M}{\varepsilon}\sqrt{\frac{\log(1/\delta)}{n}}\right).
$$

Thus, choosing $M = \min\left\{\left(\frac{\varepsilon}{\alpha}\right)^{1/k}, \left(\frac{\sqrt{n}\varepsilon}{\sqrt{\log(1/\delta)}}\right)^{1/k}\right\}$, yields that

$$
|\widehat{\mu}_n - \mu(P)| \leq O\left(\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} + \left(\frac{1}{\varepsilon}\sqrt{\frac{\log(1/\delta)}{n}}\right)^{1-1/k}\right),
$$

which finishes the proof for the LTC setting.

Now, let us move to the CTL setting. For privacy, it follows from the same idea as in the LTC setting.

For utility, we will divide the proof into five steps and leverage the following informal diagram for an illustration of Algorithm 2.

$$
X_i \xrightarrow{\text{Corruption}} Y_i \xrightarrow{\text{Trunc.}(M)} \bar{Y}_i \xrightarrow{\text{Random Rounding}} Y_i' \xrightarrow{\text{Random Response}} Z_i \xrightarrow{\text{Sample Mean}} \widehat{\mu}_n
$$

Step 1: Bound the number of corrupted points.

By Chernoff bound for the binomial distribution, we have that for $n \geq 3\log(1/\delta)/\alpha$

$$
|\mathcal{B}| \leq 2\alpha n, \quad w.p. \quad 1-\delta,
$$

where $|\mathcal{B}|$ denotes the total number of corrupted ("bad") points. Let this event be $\mathcal{E}$, and in the following steps, we will condition on this event.

Step 2: Bound the distance $|\widehat{\mu}_n - \frac{1}{n}\sum_i \mathbb{E}\left[\bar{Y}_i\right]|$.

$$
\begin{aligned}
|\widehat{\mu}_n - \frac{1}{n}\sum_i \mathbb{E}\left[\bar{Y}_i\right]| &\stackrel{(a)}{=} |\frac{1}{n}\sum_i Z_i - \frac{1}{n}\sum_i \mathbb{E}\left[Y_i'\right]| \\
&\stackrel{(b)}{\leq} O\left(M \cdot \frac{e^\varepsilon + 1}{e^\varepsilon - 1} \cdot \sqrt{\frac{\log(1/\delta)}{n}}\right) \quad w.p. \quad 1-\delta,
\end{aligned}
$$

where (a) holds by property of random rounding, i.e., $\mathbb{E}\left[\bar{Y}_i\right] = \mathbb{E}\left[Y_i'\right]$; (b) holds by property of random response, i.e., $\mathbb{E}\left[Z_i\right] = \mathbb{E}\left[Y_i'\right]$ and Hoeffding inequality.

Step 2: Bound the distance $|\frac{1}{n}\sum_i \bar{Y}_i - \frac{1}{n}\sum_i \mathbb{E}\left[\bar{Y}_i\right]|$.

$$|\frac{1}{n}\sum_i \bar{Y}_i - \frac{1}{n}\sum_i \mathbb{E}\left[\bar{Y}_i\right]| \le O\left(M \cdot \sqrt{\frac{\log(1/\delta)}{n}}\right), \quad w.p. \quad 1-\delta$$

where it simply follows from Hoeffding's inequality.

Step 3: Bound the distance $|\frac{1}{n}\sum_i \bar{Y}_i - \mathbb{E}\left[X_i\right]|$.

$$\begin{aligned}
|\frac{1}{n}\sum_{i\in[n]} \bar{Y}_i - \mathbb{E}\left[X_i\right]| &\overset{(a)}{=} |\frac{1}{n}\sum_{i\in\mathcal{G}} \bar{Y}_i - \mathbb{E}\left[X_i\right] + \frac{1}{n}\sum_{i\in\mathcal{B}} \bar{Y}_i| \\
&\overset{(b)}{\le} |\frac{1}{n}\sum_{i\in\mathcal{G}} \bar{Y}_i - \mathbb{E}\left[X_i\right]| + 2\alpha M \\
&= |\frac{1}{n}\sum_{i\in[n]} X_i \mathbb{1}(|X_i| \le M) - \mathbb{E}\left[X_i\right] - \frac{1}{n}\sum_{i\in[\mathcal{B}]} X_i \mathbb{1}(|X_i| \le M)| + 2\alpha M \\
&\le |\frac{1}{n}\sum_{i\in[n]} X_i \mathbb{1}(|X_i| \le M) - \mathbb{E}\left[X_i\right]| + 4\alpha M \\
&\le \underbrace{|\frac{1}{n}\sum_{i\in[n]} X_i \mathbb{1}(|X_i| \le M) - \mathbb{E}\left[X_i \mathbb{1}(|X_i| \le M)\right]|}_{\mathcal{T}_1} + \underbrace{|\mathbb{E}\left[X_i \mathbb{1}(|X_i| \le M)\right] - \mathbb{E}\left[X_i\right]|}_{\mathcal{T}_2} + 4\alpha M
\end{aligned}$$

where in (a), $\mathcal{G}$ represents all "good" indexes that are not corrupted and $\mathcal{B}$ represents all "bad" indexes that are corrupted; (b) follows from the boundedness of $\bar{Y}_i$ and the event $\mathcal{E}$ in step 1.

For $\mathcal{T}_2$, by Hölder's inequality and the fact k-th moment of X_i is upper bounded by one, we have

$$\mathcal{T}_2 \le O\left(\frac{1}{M^{k-1}}\right).$$

For $\mathcal{T}_1$, we consider two cases: (i) $k \in (1,2)$ and (ii) $k \ge 2$ when applying Bernstein's inequality.

For case (i), we note that $\mathbb{E}\left[X_i^2 \mathbb{1}(|X_i| \le M)\right] = \mathbb{E}\left[|X_i|^k |X_i|^{2-k} \mathbb{1}(|X_i| \le M)\right] \overset{(a)}{\le} \mathbb{E}\left[|X_i|^k M^{2-k}\right] \le M^{2-k}$, where (a) follows from $k < 2$. Thus, by Bernstein's inequality, we have

$$\mathcal{T}_1 \le O\left(\sqrt{\frac{M^{2-k}\log(1/\delta)}{n}} + \frac{M\log(1/\delta)}{n}\right).$$

For case (ii), we note that $\mathbb{E}\left[X_i^2 \mathbb{1}(|X_i| \le M)\right] \le \mathbb{E}\left[X_i^2\right] \le 1$. Thus, by Bernstein's inequality, we have

$$\mathcal{T}_1 \le O\left(\sqrt{\frac{\log(1/\delta)}{n}} + \frac{M\log(1/\delta)}{n}\right).$$

Step 4: Put everything together.

Case (i): for any $k \in (1,2)$, $\varepsilon \in [0,1]$, any $\delta \in (0,1)$ and any $P \in \mathcal{P}_k$, we have with probability at least $1-\delta$,

$$|\widehat{\mu}_n - \mu(P)| \le O\left(\sqrt{\frac{M^{2-k}\log(1/\delta)}{n}}\right) + O\left(\frac{M\log(1/\delta)}{n}\right) + O\left(\frac{1}{M^{k-1}}\right) + O(\alpha M) + O\left(\frac{M}{\varepsilon}\cdot\sqrt{\frac{\log(1/\delta)}{n}}\right).$$

Thus, choosing $M = \min\left\{\left(\frac{n}{\log(1/\delta)}\right)^{1/k}, \left(\frac{1}{\alpha}\right)^{1/k}, \left(\frac{\varepsilon\sqrt{n}}{\sqrt{\log(1/\delta)}}\right)^{1/k}\right\}$, yields that

$$|\widehat{\mu}_n - \mu(P)| \le O\left(\left(\frac{\log(1/\delta)}{n}\right)^{1-1/k} + \alpha^{1-1/k} + \left(\frac{\sqrt{\log(1/\delta)}}{\varepsilon\sqrt{n}}\right)^{1-1/k}\right).$$

Hence, when $n \ge \log(1/\delta)$, we have

$$|\widehat{\mu}_n - \mu(P)| \le O\left(\left(\frac{\sqrt{\log(1/\delta)}}{\varepsilon\sqrt{n}}\right)^{1-1/k} + \alpha^{1-1/k}\right).$$

Case (ii): for any $k \ge 2$, $\varepsilon \in [0,1]$, any $\delta \in (0,1)$ and any $P \in \mathcal{P}_k$, we have with probability at least $1-\delta$,

$$|\widehat{\mu}_n - \mu(P)| \le O\left(\sqrt{\frac{\log(1/\delta)}{n}}\right) + O\left(\frac{M\log(1/\delta)}{n}\right) + O\left(\frac{1}{M^{k-1}}\right) + O(\alpha M) + O\left(\frac{M}{\varepsilon}\cdot\sqrt{\frac{\log(1/\delta)}{n}}\right).$$

Thus, choosing $M = \min\left\{\left(\frac{n}{\log(1/\delta)}\right)^{1/k}, \left(\frac{1}{\alpha}\right)^{1/k}, \left(\frac{\varepsilon\sqrt{n}}{\sqrt{\log(1/\delta)}}\right)^{1/k}\right\}$, yields that

$$|\widehat{\mu}_n - \mu(P)| \le O\left(\sqrt{\frac{\log(1/\delta)}{n}} + \left(\frac{\log(1/\delta)}{n}\right)^{1-1/k} + \alpha^{1-1/k} + \left(\frac{\sqrt{\log(1/\delta)}}{\varepsilon\sqrt{n}}\right)^{1-1/k}\right).$$

Hence, when $n \ge \log(1/\delta)$, we have

$$|\widehat{\mu}_n - \mu(P)| \le O\left(\left(\frac{\sqrt{\log(1/\delta)}}{\varepsilon\sqrt{n}}\right)^{1-1/k} + \alpha^{1-1/k}\right).$$

Finally, combining the above two cases, we see that when $n \ge \log(1/\delta)$, for any $k > 1$, it suffices to choose $M = \min\left\{\left(\frac{n}{\log(1/\delta)}\right)^{1/k}, \left(\frac{1}{\alpha}\right)^{1/k}, \left(\frac{\varepsilon\sqrt{n}}{\sqrt{\log(1/\delta)}}\right)^{1/k}\right\}$ and obtain that

$$|\widehat{\mu}_n - \mu(P)| \le O\left(\left(\frac{\sqrt{\log(1/\delta)}}{\varepsilon\sqrt{n}}\right)^{1-1/k} + \alpha^{1-1/k}\right).$$

This completes the proof of Proposition 2. $\square$

B Proof of Proposition 3

Proof. As in the section for mean estimation, we first focus on the LTC setting and divide the lower bound proof into two steps.

Step 1: Without corruption. In this case, we aim to establish the second term in the lower bound. We consider the first MAB instance I as follows. Let $\gamma > 0$ be determined later and

$$\begin{aligned} &P_1(X = 1/\gamma) = 1/2 \cdot \gamma^k, \quad P_1(X = 0) = 1 - 1/2 \cdot \gamma^k \\ &P_a(X = 1/\gamma) = 1/4 \cdot \gamma^k, \quad P_a(X = 0) = 1 - 1/4 \cdot \gamma^k. \quad \forall a \neq 1. \end{aligned} \tag{7}$$

Thus, one can see that $I \in \mathrm{MAB}(k)$ for a proper choice of γ and arm 1 is the optimal arm for instance I. We let M_a be the induced marginal distribution of P_a via any ε-LDP channel and $\mathbb{E}_I[\cdot]$ denote the expectation over $\mathbb{P}_I$, which is over the randomness in the marginal distributions $\{M_a\}_{a\in[K]}$ and policy π.

Then, we construct a "coupled" instance I' of I as follows. Let $i = \mathrm{argmin}_{j>1} \mathbb{E}_I[N_j(T)]$, i.e., the arm between a_2 and a_K that has the minimum number of pulls under instance I. Define the second instance I' that only differs in the distribution for arm i compared to instance I

$$P_i(X = 1/\gamma) = 3/4 \cdot \gamma^k, \quad P_i(X = 0) = 1 - 3/4 \cdot \gamma^k. \tag{8}$$

Thus, $I' \in \mathrm{MAB}(k)$ and arm i is the optimal arm for instance I'. By definition, we also have $\mathbb{E}_I[N_i(T)] \leq T/(K-1)$.

For any instance I and policy π, we let $\mathcal{R}_T(\pi, I)$ be its corresponding expected regret. Then, by standard argument and noting that the mean gap is $\Delta := 1/4 \cdot \gamma^{k-1}$, we have

$$\begin{aligned} \mathcal{R}_T(\pi, I) + \mathcal{R}_T(\pi, I') &\geq \frac{T}{2} \cdot \Delta \cdot (\mathbb{P}_I[N_1(T) \leq T/2] + \mathbb{P}_{I'}[N_1(T) \geq T/2]) \\ &\overset{(a)}{\geq} \frac{T\Delta}{4} \exp(-\mathrm{KL}(\mathbb{P}_I \| \mathbb{P}_{I'})) \\ &\overset{(b)}{=} \frac{T\Delta}{4} \exp(-\mathbb{E}_I[N_i(T)] \cdot \mathrm{KL}(M_i \| M_i')) \\ &\overset{(c)}{\geq} \frac{T\Delta}{4} \exp(-\mathbb{E}_I[N_i(T)] \cdot 4(e^\varepsilon - 1)^2 \cdot (\mathrm{TV}(P_i, P_i'))^2) \\ &\overset{(d)}{\geq} \frac{T\Delta}{4} \exp\left(-\frac{T}{K-1} \cdot 4(e^\varepsilon - 1)^2 \cdot (\mathrm{TV}(P_i, P_i'))^2\right) \\ &\overset{(e)}{=} \frac{T\Delta}{4} \exp\left(-\frac{T}{K-1} \cdot 4(e^\varepsilon - 1)^2 \cdot \frac{\gamma^{2k}}{4}\right) \end{aligned}$$

where (a) holds by Bretagnolle–Huber inequality; (b) follows from chain rule of KL divergence; (c) holds by Theorem 1 in [DJW18]; (d) is true since $\mathbb{E}_I[N_i(T)] \leq T/(K-1)$; (e) holds by definition of TV distance.

Thus, putting everything together and noting that for $\varepsilon \in [0, 1]$, $e^\varepsilon - 1 \leq 2\varepsilon$, yields that

$$\mathcal{R}_T(\pi, I) + \mathcal{R}_T(\pi, I') \geq \frac{T\Delta}{4} \exp\left(-4\frac{\varepsilon^2 T \gamma^{2k}}{K-1}\right).$$

Thus, suppose $T \geq K/\varepsilon^2$ and choosing $\gamma = (K/(\varepsilon^2 T))^{1/2k}$, one can check that all the required conditions on γ are satisfied and we finally have that $\max\{\mathcal{R}_T(\pi, I), \mathcal{R}_T(\pi, I')\} \geq \Omega\left(T\gamma^{k-1}\right) = \Omega\left(T^{\frac{k+1}{2k}} \left(\frac{K}{\varepsilon^2}\right)^{\frac{k-1}{2k}}\right)$.

Step 2: Corruption part. In this case, we aim to establish the first term in the lower bound.

This part basically shares the same argument as before for mean estimation. Note that the only difference between I and I' is the distribution for arm i. Then, we apply the same argument as in the proof of Proposition 1 to P_i and P'_i. Hence, we have that there exists Huber corruptions so that one cannot distinguish between P_i and P'_i, and hence I and I'. As a result, the total expected regret is $\Omega\left(T\gamma^{k-1}\right) = \Omega(T(\alpha/\varepsilon)^{1-1/k})$.

Finally, for the CTL setting, the first step is the same and second step only differs in that there is no "contraction" effect as in the proof of Proposition 1. $\square$

C Proof of Proposition 4

Proof. Let us start with the LTC case. We divide the set of all sub-optimal arms $\mathcal{G}$ into two groups $\mathcal{G}_1$ and $\mathcal{G}_2 := \mathcal{G} \setminus \mathcal{G}_1$, where $\mathcal{G}_1 = \{a \in [K] \setminus a^* : c'\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} < \frac{1}{2}\Delta_a\}$ for some universal constant c' chosen later. Hence, $\mathcal{G}_2 = \{a \in [K] \setminus a^* : c'\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} \geq \frac{1}{2}\Delta_a\}$, which implies that the total expected regret from suboptimal arms in $\mathcal{G}_2$ is upper bounded by $O\left(T\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k}\right)$. Thus, it remains to bound the total expected regret of pulling suboptimal arms in $\mathcal{G}_1$. To this end, for each $i \in \mathcal{G}_1$, we aim to show that

$$\mathbb{E}\left[N_i(T)\right] \leq O\left(\frac{\log T}{\varepsilon^2(\Delta_i)^{\frac{2k}{k-1}}} + \frac{\log T}{\alpha}\right). \tag{9}$$

Let us first assume (9) holds and see how we can arrive at our claimed upper bound. By the definition of expected regret, we have

$$\begin{aligned} R(k,\varepsilon,\alpha,T) &= \sum_{i\in\mathcal{G}_1}\Delta_i\mathbb{E}\left[N_i(T)\right] + \sum_{i\in\mathcal{G}_2}\Delta_i\mathbb{E}\left[N_i(T)\right] \\ &\leq \sum_{i\in\mathcal{G}_1}\Delta_i\mathbb{E}\left[N_i(T)\right] + O\left(T\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k}\right), \end{aligned}$$

where inequality holds by the definition of $\mathcal{G}_2$. It remains to translate the first term into a problem-independent one. To this end, we further divide the arms in $\mathcal{G}_1$ into two groups: one group consists of all arms that satisfy $\Delta_i < \eta$ for some constant $\eta > 0$ and another one contains all arms that satisfy $\Delta_i \geq \eta$. Thus, by (9), we have

$$\sum_{i\in\mathcal{G}_1}\Delta_i\mathbb{E}\left[N_i(T)\right] \leq \eta T + O\left(\frac{K\log T}{\varepsilon^2\eta^{\frac{k+1}{k-1}}} + \frac{K\log T}{\alpha}\right).$$

Choosing $\eta = \left(\frac{K\log T}{\varepsilon^2 T}\right)^{\frac{k-1}{2k}}$, yields that the total expected regret satisfies

$$R(k,\varepsilon,\alpha,T) \leq O\left(\left(\frac{K\log T}{\varepsilon^2}\right)^{\frac{k-1}{2k}} T^{\frac{k+1}{2k}} + \frac{K\log T}{\alpha} + T\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k}\right).$$

Finally, for very small α, one can replace it with its upper bound $\bar{\alpha}$ to optimize the regret.

It remains to establish (9). First note that $O(\log T/\alpha)$ basically follows from the burn-in period. Thus, we only need to bound the total number of pulls after the burn-in period. We denote by $N'_i(t)$ the total number of by time t after the burn-in period, i.e., it is equal to $N_i(t)$ minus the total number of burn-in plays of arm i. In the following, we will show that

$$\mathbb{E}\left[N_i'(T)\right] \le C_1 \frac{\log T}{\varepsilon^2 (\Delta_i)^{\frac{2k}{k-1}}} + C_2, \tag{10}$$

for some constants C_1 and C_2.

To this end, for t that is after the burn-in period of arm $i \in \mathcal{G}_1$, if $a_t = i$, then one of the following must be true:

$$\mathrm{UCB}_{a^\star}(t) \le \mu(P_{a^\star}) \tag{11}$$

$$\widehat{\mu}_{i,N_i(t)} > \mu(P_i) + c\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} + c\left(\frac{1}{\varepsilon}\sqrt{\frac{\log(t^4)}{N_i(t)}}\right)^{1-1/k} \tag{12}$$

$$N_i'(t) < C\frac{\log T}{\varepsilon^2 (\Delta_i)^{\frac{2k}{k-1}}} \tag{13}$$

This is because if all three are not true, then we have

$$\begin{aligned}
\mathrm{UCB}_{a^\star}(t) &> \mu(P_{a^\star}) \\
&= \mu(P_i) + \Delta_i \\
&\overset{(a)}{\ge} \mu(P_i) + \frac{1}{2}\Delta_i + c'\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} \\
&\overset{(b)}{\ge} \mu(P_i) + 2c\left(\frac{1}{\varepsilon}\sqrt{\frac{\log(t^4)}{N_i'(t)}}\right)^{1-1/k} + c'\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} \\
&\overset{(c)}{\ge} \mu(P_i) + 2c\left(\frac{1}{\varepsilon}\sqrt{\frac{\log(t^4)}{N_i(t)}}\right)^{1-1/k} + c'\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} \\
&\overset{(d)}{\ge} \widehat{\mu}_{i,N_i(t)} + c\left(\frac{1}{\varepsilon}\sqrt{\frac{\log(t^4)}{N_i(t)}}\right)^{1-1/k} + c\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} \\
&= \mathrm{UCB}_i(t)
\end{aligned}$$

where (a) holds by the fact that $i \in \mathcal{G}_1$; (b) holds by choosing a large constant C in (13); (c) is true since $N_i(t) > N_i'(t)$; (d) holds by the inverse direction of (12) and choosing $c' = 2c$.

Let t' be the time just after the burn-in period, then we have

$$\begin{aligned}
\mathbb{E}\left[N_i'(T)\right] = \mathbb{E}\left[\sum_{t\ge t'} \mathbb{1}(a_t = i)\right] &\le C\frac{\log T}{\varepsilon^2 (\Delta_i)^{\frac{2k}{k-1}}} + \sum_{t \ge t'} \mathbb{E}\left[\mathbb{1}(a_t = i \text{ and (13) is false})\right] \\
&\overset{(a)}{\le} C\frac{\log T}{\varepsilon^2 (\Delta_i)^{\frac{2k}{k-1}}} + \sum_{t \ge t'} \mathbb{E}\left[\mathbb{1}(\text{(11) is true or (12) is true})\right]
\end{aligned}$$

where (a) holds by the above claim, i.e., if $a_t = i$ and (13) is false, then one of (11) and (12) must be true. Then, by our mean concentration result and union bounds, we can upper bound the second term above as

$$\mathbb{E}\left[\mathbb{1}(\text{(11) is true or (12) is true})\right] \le 2\sum_{s=1}^{t} \frac{1}{t^4} = \frac{2}{t^3}.$$

Putting them together, we have established (10), hence the result. The proof for CTL setting is essentially the same with the only difference in the definition of $\mathcal{G}_1$ and $\mathcal{G}_2$. $\square$

D Proof of Proposition 5

Proof. **Without corruption.** We consider two instances in $\mathrm{MAB}(\beta^\star, k)$. In particular, we consider two-arm MABs I and I':

$$\begin{aligned}&\text{For } I: \mu_1^I := \mu(P_1^I) = 1/2 \cdot \gamma^{k-1},\ \mu_2^I := \mu(P_2^I) = 1/4 \cdot \gamma^{k-1}\\ &\text{For } I': \mu_1^{I'} := \mu(P_1^I) = 1/2 \cdot \gamma^{k-1},\ \mu_2^{I'} := \mu(P_2^I) = 3/4 \cdot \gamma^{k-1}\end{aligned} \tag{14}$$

These distributions can be constructed in the same way as in the proof of Proposition 3 (cf. (7)). Moreover, for the behavior policy π, we have $\pi(2) = 1/\beta^\star$ and $\pi(1) = 1 - 1/\beta^\star$. We now verify that both (π, μ^I) and $(\pi, \mu^{I'})$ are in $\mathrm{MAB}(\beta^\star, k)$. By construction, each distribution is belonging to $\mathcal{P}_k$. It remains to verify that $1/\pi(a^\star) \le \beta^\star$. For I', we have $1/\pi(2) = \beta^\star$. And for I, we have $1/\pi(1) = 1/(1 - 1/\beta^\star) \le \beta^\star$ when $\beta^\star \ge 2$.

Now, we proceed to apply classic Le Cam's method. Let loss/sub-optimality of any final chosen arm $\widehat{a} \in \{1, 2\}$ under I and I' be $\ell(\widehat{a}; I)$, $\ell(\widehat{a}; I')$. Then, by our construction, we have

$$\ell(\widehat{a}; I) + \ell(\widehat{a}; I') \ge 1/4 \cdot \gamma^{k-1}.$$

Thus, by Le Cam's method and Bretagnolle–Huber inequality, we have

$$\mathrm{SubOpt}^*(\beta^\star, k, \varepsilon, \alpha, N) \ge \frac{\gamma^{k-1}}{16} \exp\left(-\mathrm{KL}\left(M_\pi^I \| M_\pi^{I'}\right)\right),$$

where $\mathrm{KL}\left(M_\pi^I \| M_\pi^{I'}\right)$ is the private KL divergence between two MAB instances. By chain rule of KL divergence and Theorem 1 in [DJW18], we have

$$\mathrm{KL}\left(M_\pi^I \| M_\pi^{I'}\right) \le \frac{N}{\beta^\star} 4(e^\varepsilon - 1)^2 \left(\mathrm{TV}\left(P_2^I, P_2^{I'}\right)\right)^2.$$

Thus, noting that for $\varepsilon \in [0,1]$, $e^\varepsilon - 1 \le 2\varepsilon$ and $\left(\mathrm{TV}\left(P_2^I, P_2^{I'}\right)\right)^2 = \frac{\gamma^{2k}}{4}$, we have that

$$\mathrm{SubOpt}^*(\beta^\star, k, \varepsilon, \alpha, N) \ge \frac{\gamma^{k-1}}{16} \exp\left(-\frac{\varepsilon^2 N \gamma^{2k}}{4\beta^\star}\right).$$

Finally, for a large enough N, choosing $\gamma = (\beta^\star/(\varepsilon^2 N))^{1/2k}$, yields that

$$\mathrm{SubOpt}^*(\beta^\star, k, \varepsilon, \alpha, N) \ge \Omega\left(\left(\frac{1}{\varepsilon}\sqrt{\frac{\beta^\star}{N}}\right)^{1-1/k}\right).$$

Corruption part. By our construction (cf. (14) (7), (8)) we have that $\mathrm{TV}\left(P_2^I, P_2^{I'}\right) = \frac{\gamma^k}{2}$. Then, a similar idea as in the proof of Proposition 1 applies here. That is, for the LTC setting, by the contraction of LDP, we can set $\gamma^k = \Theta(\frac{\alpha}{\varepsilon})$ so that $\mathrm{TV}\left(M_2^I, M_2^{I'}\right) \le \alpha$. Thus, one cannot distinguish I and I' under α-Huber model. Thus, one has to incur a sub-optimality gap as $\Omega(\gamma^k) = \left(\frac{\alpha}{\varepsilon}\right)^{1-1/k}$. In contrast, due to no contraction by LDP first, one can only set $\gamma^k = \Theta(\alpha)$, which leads to the final result. $\square$

E Proof of Proposition 6

Proof. We will focus on the LTC case, since the CTL case is nearly the same with a minor change in the confidence bound. Let $\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2$ where

$$\mathcal{E}_1 := \{\forall a \in [K], |\widehat{r}(a) - r(a)| \leq b(a)\}$$
$$\mathcal{E}_2 := \{N(a^\star) \geq \frac{1}{2} N\pi(a^\star)\}.$$

Let us first assume that $\mathbb{P}[\mathcal{E}] \geq 1 - 2\delta$ and see how we can prove the final result. Then, we will establish this high-probability event in the end. Hence, condition the event $\mathcal{E}$ and define LCB(a) $:= \widehat{r}(a) - b(a)$, we have

$$\begin{aligned} r(a^\star) - r(\widehat{a}) &= r(a^\star) - \mathrm{LCB}(a^\star) + \mathrm{LCB}(a^\star) - \mathrm{LCB}(\widehat{a}) + \mathrm{LCB}(\widehat{a}) - r(\widehat{a}) \\ &\leq 2b(a^\star) \\ &\leq 2c\left(\frac{\alpha}{\varepsilon}\right)^{1-1/k} + 2c\left(\frac{1}{\varepsilon}\sqrt{\frac{\log(2K/\delta)}{N_{a^\star}}}\right)^{1-1/k}. \end{aligned}$$

Then, by the definition of $\beta^\star$ and $\mathcal{E}$, we can further lower bound $N_{a^\star}$ by $\frac{N}{2\beta^\star}$. Then, by the bounded mean of each arm, and choosing $\delta = 1/N$, we have the claimed expected sub-optimality result.

It remains to bound the probability of $\mathcal{E}$. For $\mathcal{E}_2$, by standard Chernoff bound, we have $\mathbb{P}[\mathcal{E}_2] \geq 1 - \delta$ when $N \geq 8\beta^\star \log(1/\delta)$. For $\mathcal{E}_1$, we have the following argument. For any arm a such that N_a is larger than the burn-in threshold, the concentration in $\mathcal{E}_1$ follows from our high-probability mean estimation result. For all other arms, by construction and the condition that all arms have mean between $[-1, 1]$, we have

$$\widehat{r}(a) - b(a) = -1 \leq r(a) \leq \widehat{r}(a) + b(a) = 1,$$

which enables us to establish our claim $\mathbb{P}[\mathcal{E}] \geq 1 - 2\delta$.

□

F Online MABs under Strong Huber Corruption

In this scenario, our goal is to create an adversary strong Huber corruption for online MABs, where the agent cannot distinguish the distributions of two arms by utilizing the following probability distribution:

$$P(X = 1/\gamma) = \gamma^k, \quad P(X = 0) = 1 - \gamma^k$$
$$P'(X = 1/\gamma) = \gamma^k/2, \quad P'(X = 0) = 1 - \gamma^k/2$$

where γ adopts the form $c_1 \cdot (\alpha/\varepsilon)^{1/k}$ under LTC and $c_1 \cdot (\alpha)^{1/k}$ under CTL, with c_1 configured as 0.1 to ensure $\gamma^k \leq 1$ for an expansive α. As before, $P, P' \in \mathcal{P}_k$ for any $k > 1$ and $\mu(P) = \gamma^{k-1}$, $\mu(P) = \gamma^{k-1}/2$. Let P and P' represent the distributions for arms 0 and 1 respectively. We define the corruption distribution under CTL settings as:

Definition 10 (Strong Huber corruption under CTL Settings)**.**

$$C(X = 1/\gamma) = \gamma^k/2, \qquad C(X = 0) = 1 - \gamma^k/2$$
$$C'(X = 1/\gamma) = \gamma^k/(2\alpha), \qquad C'(X = 0) = 1 - \gamma^k/(2\alpha)$$

According to 2, it is apparent that the agent cannot differentiate between P and P' upon executing the operation:

$$(1-\alpha)P + \alpha C = (1-\alpha)P' + \alpha C'$$

This outcome emerges from the CTL's inherent nature of initially introducing contamination, which perseveres in maintaining indistinguishability, even post-transmission through the LDP channel and the Huber model.

In the context of LTC settings, the distinctiveness arises from the fact that the distributions of P and P' undergo alterations after passing through LDP, necessitating corresponding corruptions. Let R and R' be the post-LDP transformation distributions over variable Y, defined as:

$$R(Y=S) = \frac{1}{2} + \frac{\gamma^k}{2} \cdot \frac{e^\varepsilon - 1}{e^\varepsilon + 1}, \quad R(Y=-S) = \frac{1}{2} - \frac{\gamma^k}{2} \cdot \frac{e^\varepsilon - 1}{e^\varepsilon + 1}$$
$$R'(Y=S) = \frac{1}{2} + \frac{\gamma^k}{4} \cdot \frac{e^\varepsilon - 1}{e^\varepsilon + 1}, \quad R'(Y=-S) = \frac{1}{2} - \frac{\gamma^k}{4} \cdot \frac{e^\varepsilon - 1}{e^\varepsilon + 1}$$

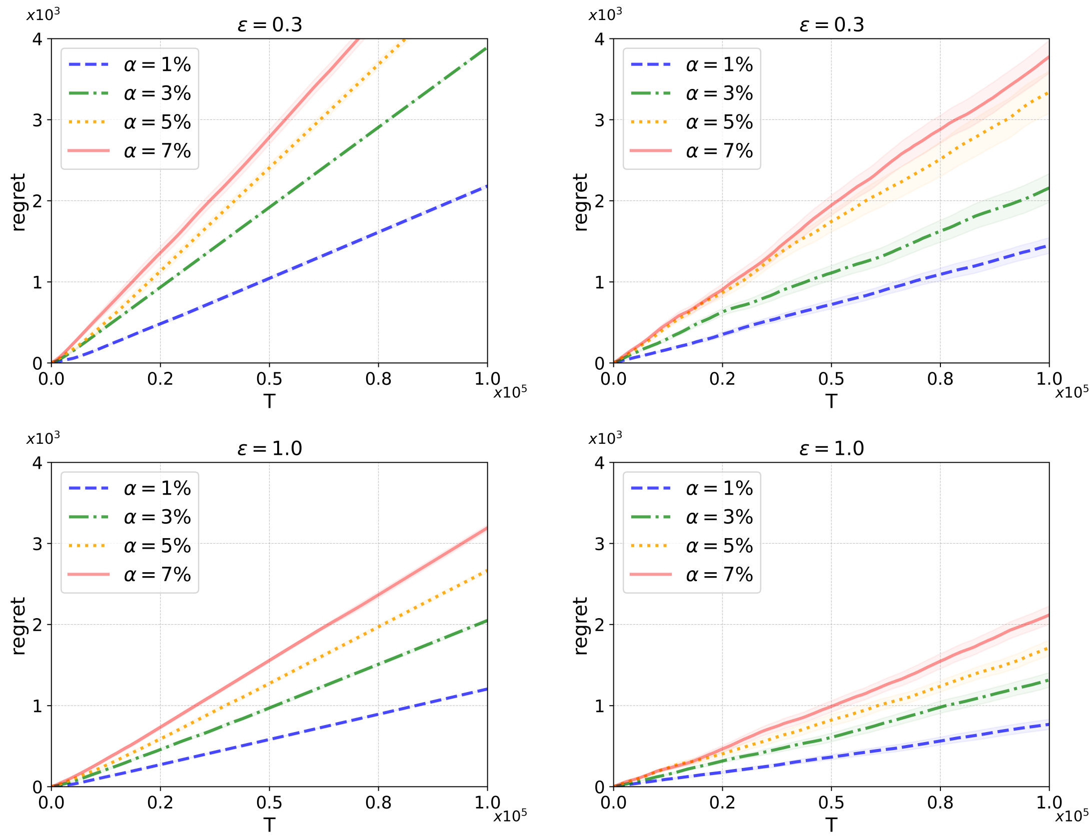
where $S = M \cdot \frac{e^\varepsilon+1}{e^\varepsilon-1}$. Additionally, we define the corruption distribution as:

Definition 11 (Strong Huber corruption under LTC Settings)**.**

$$N(Y=S) = \frac{\gamma^k}{4} \cdot \frac{e^\varepsilon - 1}{e^\varepsilon + 1}, \quad N(Y=-S) = 1 - \frac{\gamma^k}{4} \cdot \frac{e^\varepsilon - 1}{e^\varepsilon + 1}$$
$$N'(Y=S) = \frac{\gamma^k}{4} \cdot \frac{e^\varepsilon - 1}{e^\varepsilon + 1} \cdot \frac{1}{\alpha}, \quad N'(Y=-S) = 1 - \frac{\gamma^k}{4} \cdot \frac{e^\varepsilon - 1}{e^\varepsilon + 1} \cdot \frac{1}{\alpha}$$

Now we also have $(1-\alpha)R + \alpha N = (1-\alpha)R' + \alpha N'$, indicating our continued inability to distinguish between P and P' in the LTC setting.

Fig. 7 illustrates the regret performance of our proposed algorithm (i.e., Algorithm 3) for online MABs under LTC and CTL settings, with the strong Huber corruption in Definition 10 and Definition 11. A common pattern behind all the plots in Fig. 7 is that due to strong huber corruption, the agent cannot distinguish the distributions of two arms, hence linear regret. Based on the analysis above, we anticipate that the regret will scale linearly by a factor of c_1 with respect to our minimax clean regret and Fig. 7 aligned with our discussion. The two plots in column (a) capture the LTC setting while the two plots in column (b) denote the CTL setting. As expected, the regret performance under LTC is worse than that under CTL, highlighting separation results in terms of actual performance rather than only in terms of theoretical upper bounds.



(a) LTC, vary α

(b) CTL, vary α

Figure 7: Regret performance with strong Huber corruption given by Appendix F under LTC and CTL settings.