Appendices

A. Proof of Theorem 4

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Combining with Equ. 7, we have

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

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

□

B. Proof of Theorem 5

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

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

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

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

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

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

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

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

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

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

C. Proof of Theorem 6

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

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

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

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

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

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

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

- The probability of choosing a super arm $S_t = S$ in step t is dictated only by the algorithm $\pi(S|\mathcal{H}_{t-1})$.
- The distribution of rewards $\boldsymbol{X}_t$ in step t is $f_{S_t}^{\nu}$, which depends on S_t and conditionally independent on the history $\mathcal{H}_{t-1}$.
- In the case of local differential privacy, the algorithm cannot observe X_t directly, but a privated version of rewards $\boldsymbol{Z}_t$. $\boldsymbol{Z}_t$ only depends on X_t and is conditionally independent on the history $\mathcal{H}_{t-1}$. Denote the conditional distribution of $\boldsymbol{Z}$ as $M(\boldsymbol{Z}|\boldsymbol{X})$.

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

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

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

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

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

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

Proof.

$$\begin{aligned}
D\left(\mathbb{P}^T_{\pi\nu_1}\|\mathbb{P}^T_{\pi\nu_2}\right) &= \int_{\mathcal{H}_T} \ln\frac{\mathrm{d}\mathbb{P}^T_{\pi\nu_1}(H)}{\mathrm{d}\mathbb{P}^T_{\pi\nu_2}(H)}\mathrm{d}\mathbb{P}^T_{\pi\nu_1}(H) \\
&= \int_{\mathcal{H}_T} \sum_{t=1}^{T} \ln\frac{\pi(S_t|\mathcal{H}_{t-1},\nu_1)}{\pi(S_t|\mathcal{H}_{t-1},\nu_2)}\mathrm{d}\pi(S_t|\mathcal{H}_{t-1},\nu_1) + \int_{\mathcal{H}_T} \sum_{t=1}^{T} \ln\frac{g^{\nu_1}_{S_t}(\boldsymbol{Z})}{g^{\nu_2}_{S_t}(\boldsymbol{Z})}\mathrm{d}\left(g^{\nu_1}_{S_t}(\boldsymbol{Z})\right) \\
&= \sum_{t=1}^{T} \mathbb{E}_{\pi\nu_1}\left[D\left(\pi\left(S_t|\mathcal{H}_{t-1},\nu_1\right)\|\pi\left(S_t|\mathcal{H}_{t-1},\nu_2\right)\right)\right] + \sum_{S\in\mathcal{S}}\left[\sum_{t=1}^{T}\mathbb{E}_{\mathbb{P}^T_{\pi}\nu_1}\left[\mathbb{1}_{S_t=S}\right] D\left(g^{\nu_1}_S(\boldsymbol{Z})\|g^{\nu_2}_S(\boldsymbol{Z})\right)\right] \\
&= \sum_{t=1}^{T} \mathbb{E}_{\pi\nu_1}\left[D\left(\pi\left(S_t|\mathcal{H}_{t-1},\nu_1\right)\|\pi\left(S_t|\mathcal{H}_{t-1},\nu_2\right)\right)\right] + \sum_{S\in\mathcal{S}} \mathbb{E}_{\pi\nu_1}\left[N_S(T)\right] D\left(g^{\nu_1}_S\|g^{\nu_2}_S\right)
\end{aligned}$$

□

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

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

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

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

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

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

$$w(i) = \begin{cases} 0.5 & i \in S^* \\ 0.5 + \Delta/(B_1K) & i \in S^1 \\ 0.5 - \Delta/(B_1K) & \text{otherwise} \end{cases}$$

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

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

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

Combining these two inequality, we have

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

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

By lemma 3, we have

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

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

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

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

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

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

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

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

D. Omitted Proof of Theorem 8

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

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

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

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

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

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

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

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

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

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

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

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

Then we have,

$$\begin{aligned}Reg_{\mu,\alpha,\beta}(T) \leq & T\alpha\beta\,\mathrm{opt}_\mu - \mathbb{E}\sum_{t=1}^{T} r_\mu(S_t)\\ \leq & R_{\text{fail}} + T\alpha\beta\,\mathrm{opt}_\mu - \left(T\alpha\,\mathrm{opt}_\mu - \sum_{t\in[T]} \Delta_t \mathbb{1}\{\neg G_t\}\right)\\ \leq & \sum_{t\in[T]} \Delta_t \mathbb{1}\{\neg G_t\}\end{aligned}$$

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

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

$$\begin{aligned}Reg_{\mu,\alpha,\beta}(T) \leq & \Pr\{\Lambda_1 \cap \Lambda_2\}\hat{R}(T) + \sum_{i\in[m]} \Delta_{\min}^i\\ & + \Pr\{\neg\Lambda_1\}T\Delta_{\max} + \Pr\{\neg\Lambda_1\}T\Delta_{\max}\\ \leq & \hat{R}(T) + (m+2)\Delta_{\max}\end{aligned}$$

If event Λ_1 and Λ_2 happen, we have

$$\begin{aligned}|\tilde{\mu}_t(i) - \mu_i| = & \left|\frac{\mathrm{Sum}_{t,i}}{T_{t,i}} - \mu_i + \frac{\mathrm{Noise}_{t,i}}{T_{t,i}}\right|\\ \leq & \sqrt{\frac{4\ln T}{T_{t,i}}} + \frac{12K\ln^3 T}{T_{t,i}}\end{aligned}$$

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

$$\begin{aligned}\alpha r_\mu(S_\mu^*) - r_\mu(S_t) \leq & \alpha r_{\bar{\mu}_t}(S_\mu^*) - (r_{\bar{\mu}_t}(S_t) - B_1\|\bar{\mu}_t - \mu\|_1)\\ \leq & B_1\|\bar{\mu}_t - \mu\|_1\\ \leq & B_1(\|\bar{\mu}_t - \tilde{\mu}_t\|_1 + \|\tilde{\mu}_t - \mu\|_1)\\ \leq & B_1\sum_{i\in S_t}\left(4\sqrt{\frac{\ln T}{T_{t-1,i}}} + \frac{24K\ln^3 T}{T_{t-1,i}\varepsilon}\right)\end{aligned} \tag{11}$$

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

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

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

E. Proof of Theorem 9

Theorem 9. *For any m and K such that $m \geq 2K$,and any Δ satisfying $0 < \Delta/(B_1K) < 0.35$, the regret for any consistent ε-DP algorithm on the CSB problem with B_1 bounded smoothness is at least $\Omega\left(\frac{B_1^2 mK\ln T}{\Delta} + \frac{B_1 mK\ln T}{\varepsilon}\right)$.*

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

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

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

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

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

Lemma 7. *For any m and K such that $m \geq 2K$, and any Δ satisfying $0 < \Delta/(B_1K) < 0.35$, the regret for any consistent CSB algorithm guaranteeing ε-DP is at least $\Omega\left(\frac{B_1 mK \ln T}{\varepsilon}\right)$.*

Now we only need to prove Lemma 7.

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

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

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

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

Now we prove that $P_{\nu_1}(E_{S_1}) \leq 1/(2m)$. Our analysis is inspired by the work of Shariff & Sheffet (2018). Consider another CSB instance with all the setting the same as ν_1, except that the mean weights of base arms in S_1 are increased by $2\Delta/(B_1K)$ each. We denote this instance as ν_2. Consider the case that rewards are drawn from ν_2. Due to consistent property, the regret of the algorithm is at most $T^{3/4}m\Delta$. For sufficiently large T, we have

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

.

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

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

That is, for sufficiently large T,

$$\begin{aligned}
\mathbb{P}_{\nu_1}[E_{S_1}] \le& \exp\left(6\varepsilon t_S\cdot d_{\mathrm{TV}}(\nu_1,\nu_2)\right)\mathbb{P}_{\nu_2}[E_{S_1}] \\
\le& \exp\left(24\varepsilon t_S\cdot\frac{\Delta}{B_1K}\right)\mathbb{P}_{\nu_2}[E_{S_1}] \\
\le& \exp\left(0.06\ln T\right)\frac{mK}{T^{1/4}} \\
=& mKT^{-0.19} \le \frac{1}{2m}.
\end{aligned}$$

The second inequality is due to $d_{\mathrm{TV}}(\nu_1,\nu_2) \le \sqrt{\frac{D_{KL}(\nu_1\|\nu_2)}{2}} \le 4\Delta/(B_1K)$ by Pinsker's inequality and the setting that the public base arms are identical.

□