Stochastic Graphical Bandits with Heavy-Tailed Rewards (Supplementary Material)

Yutian Gou[1]   **Jinfeng Yi**[2]   **Lijun Zhang**[1*]

[1]National Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210023, China
[2]JD AI Research, Beijing 100176, China

1 OMITTED PROOFS FOR SECTION 3.1

In this section, we present the proofs for RUN-TEM, RUNE-TEM and RUNE-MoM omitted in the first section of our main results.

1.1 TECHNICAL LEMMAS

Lemma 1 (Bernstein's inequality for bounded distributions [Vershynin, 2018]) *Let* $X_1, \cdots, X_n$ *be* n *independent zero-mean random variables. Suppose* $|X_i| \leq K$ *and* $\mathbb{E}[X_i^2] \leq u$ *for all* $i \in [n]$. *Then for any* $t > 0$, *we have*

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}X_i \geq t\right] \leq \exp\left(\frac{nt^2}{2(u + Kt/3)}\right), \tag{1}$$

also,

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}X_i \leq -t\right] \leq \exp\left(\frac{nt^2}{2(u + Kt/3)}\right), \tag{2}$$

and therefore,

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}X_i \geq \sqrt{\frac{2ut}{n}} + \frac{2Kt}{3n}\right] \leq e^{-t}. \tag{3}$$

1.2 ANALYSIS FOR RUN-TEM

Proof [of Proposition 1] Here, we follow the same computing process as the proof of Lemma 1 in [Bubeck et al., 2013]. First of all, we will bound $\mathbb{E}[X_t^2 \mathbb{I}_{|X_t| \leq B_t}]$ by the following computation

$$
\begin{aligned}
\mathbb{E}[X_t^2 \mathbb{I}_{|X_t| \leq B_t}] &= \int_{x \sim X_t} x^2 \mathbb{I}_{|x| \leq B_t} \mathrm{d}x \\
&= \int_{x \sim X_t} |x|^{1+\epsilon} \mathbb{I}_{|x| \leq B_t} |x|^{1-\epsilon} \mathrm{d}x \\
&\leq B_t^{1-\epsilon} \int_{|x| \leq B_t} |x|^{1+\epsilon} \mathrm{d}x \\
&\leq B_t^{1-\epsilon} \mathbb{E}[|X_t|^{1+\epsilon}] \leq v B_t^{1-\epsilon}.
\end{aligned} \tag{4}
$$

*Corresponding author.

Also, we can bound $\mathbb{E}[X_t \mathbb{I}_{|X_t|>B_t}]$ by a similar calculation

$$
\begin{aligned}
\mathbb{E}[X_t \mathbb{I}_{|X_t|>B_t}] &= \int_{x \sim X_t} x \mathbb{I}_{|x|>B_t} \mathrm{d}x \\
&= \int_{x \sim X_t} \frac{|x|^{1+\epsilon}}{|x|^{\epsilon}} \mathbb{I}_{|x|>B_t} \mathrm{d}x \\
&< \frac{1}{B_t^{\epsilon}} \int_{|x|>B_t} |x|^{1+\epsilon} \mathrm{d}x \\
&\leq \frac{1}{B_t^{\epsilon}} \mathbb{E}[|X_t|^{1+\epsilon}] \leq \frac{v}{B_t^{\epsilon}} .
\end{aligned}
\tag{5}
$$

Then, by combining (4),(5) with Bernstein's inequality(3), we have that, with probability at least $1 - \delta$,

$$
\begin{aligned}
\mathbb{E}[X] - \frac{1}{n}\sum_{t=1}^{n} X_t \mathbb{I}_{|X_t| \leq B_t} &= \frac{1}{n}\sum_{t=1}^{n} (\mathbb{E}[X] - \mathbb{E}[X\mathbb{I}_{|X| \leq B_t}]) + \frac{1}{n}\sum_{t=1}^{n} (\mathbb{E}[X\mathbb{I}_{|X| \leq B_t}] - X_t \mathbb{I}_{|X_t| \leq B_t}) \\
&= \frac{1}{n}\sum_{t=1}^{n} \mathbb{E}[X\mathbb{I}_{|X| > B_t}] + \frac{1}{n}\sum_{t=1}^{n} (\mathbb{E}[X\mathbb{I}_{|X| \leq B_t}] - X_t \mathbb{I}_{|X_t| \leq B_t}) \\
&\leq \frac{1}{n}\sum_{t=1}^{n} \frac{v}{B_t^{\epsilon}} + \sqrt{\frac{2B_n^{1-\epsilon} v \log(1/\delta)}{n}} + \frac{2B_n \log(1/\delta)}{3n} .
\end{aligned}
\tag{6}
$$

Finally, substitute $B_t = (\frac{vt}{\log(1/\delta)})^{\frac{1}{1+\epsilon}} (\epsilon \in (0, 1])$ into the above inequality, we get that, with probability at least $1 - \delta$,

$$
\begin{aligned}
\mathbb{E}[X] - \frac{1}{n}\sum_{t=1}^{n} X_t \mathbb{I}_{|X_t| \leq B_t} &\leq \left(\frac{1}{n}\sum_{t=1}^{n} \left(\frac{n}{t}\right)^{\frac{\epsilon}{1+\epsilon}} + \sqrt{2} + \frac{2}{3} \right) v^{\frac{1}{1+\epsilon}} \left(\frac{\log(1/\delta)}{n} \right)^{\frac{\epsilon}{1+\epsilon}} \\
&\leq \left(\frac{1}{\sqrt{n}}\sum_{t=1}^{n} \frac{1}{\sqrt{t}} + \sqrt{2} + \frac{2}{3} \right) v^{\frac{1}{1+\epsilon}} \left(\frac{\log(1/\delta)}{n} \right)^{\frac{\epsilon}{1+\epsilon}} \\
&\leq \left(\frac{1}{\sqrt{n}}(2\sqrt{n} - 1) + \sqrt{2} + \frac{2}{3} \right) v^{\frac{1}{1+\epsilon}} \left(\frac{\log(1/\delta)}{n} \right)^{\frac{\epsilon}{1+\epsilon}} \\
&< 5 v^{\frac{1}{1+\epsilon}} \left(\frac{\log(1/\delta)}{n} \right)^{\frac{\epsilon}{1+\epsilon}} ,
\end{aligned}
\tag{7}
$$

where the last inequality use the following summation formula

$$
\sum_{t=1}^{n} \frac{1}{\sqrt{t}} \leq 1 + \int_{1}^{n} \frac{1}{\sqrt{x}} \mathrm{d}x = 2\sqrt{n} - 1 .
\tag{8}
$$

Now, have finished proving the first part of Proposition 1, and the second part follows the same scheme, so we omit it. ∎

Proof [of Theorem 1] Here, we use a proof sketch similar to [Caron et al., 2012]. First of all, we consider a clique covering $\mathcal{C}$ of $G = (V, E)$ and define the intra-clique regret $R_C(T)$ for any $C \in \mathcal{C}$ by

$$
R_C(T) = \sum_{i \in C}\sum_{t \leq T} \Delta_i \mathbb{I}_{\{I_t = i\}} .
\tag{9}
$$

By Definition 1, we have that $V = \cup_{C \in \mathcal{C}} C$ and furthermore $R_T \leq \sum_{C \in \mathcal{C}} R_C(T)$. From now on, we will focus on upper bounding the intra-clique regret for a given clique $C \in \mathcal{C}$ and abbreviate $R_C(T)$ to R_C.

Let $T_C(t) = \sum_{i \in C} T_i(t)$ denote the number of pulls for any arm in clique C up to time t. Then, for any positive integer ℓ_C,

$$\mathbb{E}[R_C] = \mathbb{E}\left[\sum_{i \in C}\sum_{t \le T}\Delta_i \mathbb{I}_{\{I_t=i\}}\right]$$

$$= \sum_{i \in C}\sum_{t \le T}\Delta_i \Pr[I_t = i]$$

$$= \sum_{i \in C}\sum_{t \le T}\Delta_i \Pr[I_t = i, T_C(t-1) < \ell_C] + \sum_{i \in C}\sum_{t \le T}\Delta_i \Pr[I_t = i, T_C(t-1) \ge \ell_C] \tag{10}$$

$$\le \ell_C \Delta_C^{\max} + \underbrace{\sum_{\substack{i \in C \\ t \le T}}\Delta_i \Pr\left[I_t = i, T_C(t-1) \ge \ell_C\right]}_{(a)}.$$

where $\Delta_C^{\max} := \max_{i \in C}\Delta_i$ is the maximum reward gap in clique C.

Now, we analyze term (a) using the similar tricks applied by [Auer et al., 2002]. Let $c_i(t) = 5v^{\frac{1}{1+\epsilon}}(\frac{4\log(t)}{O_i(t)})^{\frac{\epsilon}{1+\epsilon}}$ and $i^\star$ be the unique optimal arm, then (a) can be computed by

$$\sum_{\substack{i \in C \\ t < T}}\Delta_i \Pr\left[\widehat{\mu}_i(t) + c_i(t) \ge \widehat{\mu}_{i^\star}(t) + c_{i^\star}(t), T_C(t) \ge \ell_C\right] \le \sum_{\substack{i \in C \\ t < T}}\Delta_i \Pr\left[\max_{\ell_C \le s_i \le t}\widehat{\mu}_i(s_i) + c_i(s_i) \ge \min_{0 \le s \le t}\widehat{\mu}_{i^\star}(s) + c_{i^\star}(s)\right]$$

$$\le \sum_{\substack{i \in C \\ t < T}}\sum_{s_i = \ell_C}^{t}\sum_{s=0}^{t}\Delta_i \Pr\left[\widehat{\mu}_i(s_i) + c_i(s_i) \ge \widehat{\mu}_{i^\star}(s) + c_{i^\star}(s)\right]. \tag{11}$$

Next, observe that $\widehat{\mu}_i(s_i) + c_i(s_i) \ge \widehat{\mu}_{i^\star}(s) + c_{i^\star}(s)$ implies that at least one of the following three independent events must hold:

$$\widehat{\mu}_i(s_i) \ge \mu_i + c_i(s_i), \tag{12}$$

$$\widehat{\mu}_{i^\star}(s) \le \mu^\star - c_{i^\star}(s), \tag{13}$$

$$\mu^\star < \mu_i + 2c_{i,s_i}. \tag{14}$$

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

$$\widehat{\mu}_i(s_i) + c_i(s_i) < \mu_i + 2c_i(s_i) \le \mu^\star < \widehat{\mu}_{i^\star}(s) + c_{i^\star}(s), \tag{15}$$

which implies that $I_t \ne i$.

If we choose $\ell_C = \lceil\frac{40(10v)^{1/\epsilon}\log(T)}{(\Delta_C^{\min})^{(1+\epsilon)/\epsilon}}\rceil$, where $\Delta_C^{\min} = \min_{i \in C}\Delta_i$ is the minimum nonzero reward gap in clique C, then for $\forall i \ne i^\star \in C$ and $s_i \ge \ell_C$,

$$\mu^\star - \mu_i - 2c_i(s_i) = \Delta_i - 10v^{\frac{1}{1+\epsilon}}\left(\frac{4\log(T)}{s_i}\right)^{\frac{\epsilon}{1+\epsilon}} \ge 0, \tag{16}$$

which implies that (14) is always false. Thus, $\widehat{\mu}_i(s_i) + c_i(s_i) \ge \widehat{\mu}_{i^\star}(s) + c_{i^\star}(s)$ implies that (12) or (13) must hold.

Next, we can bound the probability of event (12) and (13) by using Proposition 1,

$$\Pr[(12)] = \Pr\left[\widehat{\mu}_i(s_i) \ge \mu_i + c_i(s_i)\right] \le \frac{1}{t^4},$$
$$\Pr[(13)] = \Pr\left[\widehat{\mu}_{i^\star}(s) \le \mu^\star - c_{i^\star}(s)\right] \le \frac{1}{t^4}. \tag{17}$$

Therefore, by using the union bound, we obtain that

$$
\begin{aligned}
(a) &\leq \sum_{\substack{i \in C \\ t < T}} \sum_{s_i = \ell_C}^{t} \sum_{s=0}^{t} \Delta_i \left(\Pr[(12)] + \Pr[(13)] \right) \\
&\leq \sum_{\substack{i \in C \\ t < T}} \Delta_i \left(\sum_{s_i = \ell_C}^{t} \sum_{s=0}^{t} \frac{2}{t^4} \right) \\
&\leq \sum_{\substack{i \in C \\ t < T}} \frac{2\Delta_i}{t^2} \leq \frac{\pi^2}{3} \sum_{i \in C} \Delta_i \ ,
\end{aligned}
\tag{18}
$$

substituting into (10) demonstrates that

$$
\begin{aligned}
\mathbb{E}[R_C(T)] &\leq \ell_C \Delta_C^{\max} + \frac{\pi^2}{3} \sum_{i \in C} \Delta_i \\
&\leq \frac{40(10v)^{1/\epsilon} \Delta_C^{\max}}{(\Delta_C^{\min})^{(1+\epsilon)/\epsilon}} \log T + \left(1 + \frac{\pi^2}{3} \right) \sum_{i \in C} \Delta_i \ ,
\end{aligned}
\tag{19}
$$

Finally, summing over all cliques in $\mathcal{C}$ and taking the infimum over all possible coverings $\mathcal{C}$ yields the aforementioned upper bound.of RUN-TEM. ∎

1.3 ANALYSIS FOR RUNE-TEM

Proof [of Theorem 2] Firstly, we try to bound the expectation of $R_C(T)$ by using the techniques similar to Hu et al. [2019], where we use the same symbols as the proof of Theorem 1.

For the sake of the simplicity, we assume $i^\star$ is the unique optimal arm with without loss of generality. Let $\Delta_C^{\max} = \max_{i \in C} \Delta_i, \Delta_C^{\min} = \min_{i \in C} \Delta_i, N_C := \max_{i \in C} |N(i)|^{\frac{1}{4}}$. Then, for any positive integer ℓ_C, we have

$$
\begin{aligned}
\mathbb{E}[R_C(T)] &= \sum_{\substack{i \in C \setminus \{i^\star\} \\ t \leq T}} \Delta_i \mathbb{E}[\mathbb{I}\{I_t = i\}] \\
&= \sum_{t \leq T} \Pr\left[\exists i \in C \setminus \{i^\star\} \text{ s.t. } I_t = i, T_C(t-1) < \ell_C \right] \cdot \Delta_C^{\max} \\
&\quad + \sum_{t \leq T} \Pr\left[\exists i \in C \setminus \{i^\star\} \text{ s.t. } I_t = i, T_C(t-1) \geq \ell_C \right] \cdot \Delta_C^{\max} \\
&\leq \ell_C \Delta_C^{\max} + \underbrace{\sum_{t \leq T} \Pr\left[\exists i \in C \setminus \{i^\star\} \text{ s.t. } I_t = i, T_C(t-1) \geq \ell_C \right]}_{(\alpha)} \cdot \Delta_C^{\max} \ .
\end{aligned}
\tag{20}
$$

Now, we analyze term (α) by using the similar tricks of Auer et al. [2002]. Let $c_i(t) = 5v^{\frac{1}{1+\epsilon}} \left(\frac{4 \log(|N(i)|^{1/4} t)}{O_i(t)} \right)^{\frac{\epsilon}{1+\epsilon}}$ and

$\tilde{c}_i(t) = 5v^{\frac{1}{1+\epsilon}} \left(\frac{4 \log(N_C t)}{O_i(t)} \right)^{\frac{\epsilon}{1+\epsilon}}$, then

$$
\begin{aligned}
(\alpha) &= \sum_{t \leq T} \Pr[\exists i \in C \setminus \{i^\star\} \text{ s.t. } I_t = i, T_C(t-1) \geq \ell_C] \\
&= \sum_{t < T} \Pr \left[\begin{array}{c} \max_{i \in C \setminus \{i^\star\}} \{\widehat{\mu}_i(t) + c_i(t)\} \geq \widehat{\mu}_{i^\star}(t) + c_{i^\star}((t), \\ T_C(t) \geq \ell_C \end{array} \right] \\
&\leq \sum_{t < T} \Pr \left[\max_{i \in C \setminus \{i^\star\}} \left\{ \max_{\ell_C \leq s_i \leq t} \widehat{\mu}_i(s_i) + c_i(s_i) \right\} \geq \min_{0 \leq s \leq t} \widehat{\mu}_{i^\star}(s) + c_{i^\star}(s) \right] \\
&\leq \sum_{t < T} \sum_{s_i = \ell_C}^{t} \sum_{s=0}^{t} \Pr \left[\max_{i \in C \setminus \{i^\star\}} \{\widehat{\mu}_i(s_i) + \tilde{c}_i(s_i)\} \geq \widehat{\mu}_{i^\star}(s) + c_{i^\star}(s) \right] .
\end{aligned}
\tag{21}
$$

Now, let $\mu_C^{\max} = \max_{i \in C} \mu_i$, we observe that $\max_{i \in C \setminus \{1\}} \{\widehat{\mu}_i(s_i) + \tilde{c}_i(s_i)\} \geq \widehat{\mu}_{i^\star}(s) + c_{i^\star}(s)$ implies that at least one of the following must hold:

$$
\max_{i \in C \setminus \{1\}} \{\widehat{\mu}_i(s_i)\} \geq \mu_C^{\max} + \tilde{c}_i(s_i) ,
\tag{22}
$$

$$
\widehat{\mu}_{i^\star}(s) \leq \mu^\star - c_{i^\star}(s) ,
\tag{23}
$$

$$
\mu^\star < \mu_C^{\max} + 2\tilde{c}_i(s_i) .
\tag{24}
$$

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

$$
\max_{i \in C \setminus \{1\}} \{\widehat{\mu}_i(s_i) + \tilde{c}_i(s_i)\} < \mu_C^{\max} + 2\tilde{c}_i(s_i) \leq \mu^\star < \widehat{\mu}_{i^\star}(s) + c_{i^\star}(s) ,
\tag{25}
$$

which implies that $I_t \neq i$.

If we choose $\ell_C = \lceil \frac{40(10v)^{1/\epsilon} \log(N_C T)}{(\Delta_C^{\min})^{(1+\epsilon)/\epsilon}} \rceil$, when $\ell_C \leq s_i \leq t$, then

$$
\mu^\star - \mu_C^{\max} - 2\tilde{c}_i(s_i) = \mu^\star - \mu_C^{\max} - 10v^{\frac{1}{1+\epsilon}} \left(\frac{4 \log(N_C t)}{s_i} \right)^{\frac{\epsilon}{1+\epsilon}} \geq 0 ,
\tag{26}
$$

which implies that (24) is always false. Thus, event (22) or event (23) must hold.

Next, we bound the probability of event (22) ,

$$
\begin{aligned}
\Pr[(22)] &= \Pr \left[\max_{i \in C \setminus \{i^\star\}} \{\widehat{\mu}_i(s_i)\} \geq \mu_C^{\max} + \tilde{c}_i(s_i) \right] \\
&\leq \sum_{i \in C \setminus \{i^\star\}} \Pr \left[\widehat{\mu}_i(s_i) \geq \mu_C^{\max} + \tilde{c}_i(s_i) \right] \\
&\leq \sum_{i \in C \setminus \{i^\star\}} \Pr \left[\widehat{\mu}_i(s_i) \geq \mu_i + \tilde{c}_i(s_i) \right] \\
&\leq |C| \cdot e^{-4 \log(N_C t)} = \frac{|C|}{N_C^4 t^4} \leq \frac{1}{t^4} ,
\end{aligned}
\tag{27}
$$

where the penultimate inequality is induced by Proposition 1 and subsequently a union bound on C. Similarly, we can bound the probability of (23) directly,

$$
\begin{aligned}
\Pr[(23)] &= \Pr \left[\widehat{\mu}_{i^\star}(s) \leq \mu^\star - c_{i^\star}(s) \right] \\
&\leq e^{-4 \log(|N(i^\star)|^{1/4} t)} = \frac{1}{|N(i^\star)| t^4} \leq \frac{1}{t^4} ,
\end{aligned}
\tag{28}
$$

By further computation, we get that

$$(\alpha) \leq \sum_{t<T} \sum_{s_i=\ell_C}^{t} \sum_{s=0}^{t} \Pr\left[\max_{i\in C\setminus\{i^\star\}} \{\widehat{\mu}_i(s_i) + \tilde{c}_i(s_i)\} \geq \widehat{\mu}_{i^\star}(s) + c_{i^\star}(s)\right]$$

$$\leq \sum_{t<T} \sum_{s_i=\ell_C}^{t} \sum_{s=0}^{t} (\Pr[(22)] + \Pr[(23)]) \tag{29}$$

$$\leq \sum_{t<T} \left(\sum_{s_i=\ell_C}^{t} \sum_{s=0}^{t} \frac{2}{t^4}\right) \leq \sum_{t<T} \frac{2}{t^2} \leq \frac{\pi^2}{3} \ ,$$

and furthermore, by (20), we have

$$\mathbb{E}[R_C(T)] \leq \ell_C \Delta_C^{\max} + (\alpha) \cdot \Delta_C^{\max}$$

$$\leq \ell_C \Delta_C^{\max} + \frac{\pi^2}{3} \Delta_C^{\max} \tag{30}$$

$$\leq 40 \frac{(10v)^{1/\epsilon} \Delta_C^{\max}}{(\Delta_C^{\min})^{(1+\epsilon)/\epsilon}} \log(N_C T) + \left(1 + \frac{\pi^2}{3}\right) \Delta_C^{\max} \ ,$$

Finally, summing over all cliques in $\mathcal{C}$ and taking the infimum over all possible coverings $\mathcal{C}$ yields the aforementioned upper bound of RUNE-TEM. ∎

1.4 ANALYSIS FOR RUNE-MOM

In this section, we present the proofs for RUNE-MoM and RUNE-TEM omitted in the first section of our main results. Before starting the proof, we need to state a lemma which has been proved by Bubeck et al. [2013].

Lemma 2 *Let $X_1, \cdots, X_n$ be a real i.i.d. sequence with finite mean μ. We assume that for some $\epsilon \in (0, 1]$ and $v \geq 0$, $\mathbb{E}[|X - \mu|^{1+\epsilon}] \leq v$. Let $\widehat{\mu}$ be the empirical mean:*

$$\widehat{\mu} = \frac{1}{n} \sum_{t=1}^{n} X_t \ ,$$

Then, for any $\eta > 0$, we have

$$\Pr[\mu - \widehat{\mu} > \eta] \leq \frac{3v}{n^\epsilon \eta^{1+\epsilon}} \ , \tag{31}$$

and also,

$$\Pr[\widehat{\mu} - \mu > \eta] \leq \frac{3v}{n^\epsilon \eta^{1+\epsilon}} \ . \tag{32}$$

Proof [of Proposition 2] We follow the same proof scheme as Bubeck et al. [2013]. Let $\eta = (12v)^{\frac{1}{1+\epsilon}} (\frac{1}{N})^{\frac{\epsilon}{1+\epsilon}} > 0$ and $Y_l = \mathbb{I}_{\widehat{\mu}_l < \mu - \eta}$ for $l \in [k]$. According to equation (31), Y_l has a Bernoulli distribution with parameter

$$p \leq \frac{3v}{N^\epsilon \eta^{1+\epsilon}} \leq \frac{1}{4} \ , \tag{33}$$

which implies that $\mathbb{E}[Y_l] = p \leq \frac{1}{4}$. Thus using Hoeffding's inequality [Hoeffding, 1963] for the tail of a binomial distribution, we get that

$$\Pr[\widehat{\mu}_M < \mu - \eta] \leq \Pr\left[\sum_{l=1}^{k} Y_l \geq \frac{k}{2}\right]$$

$$\leq \Pr\left[\sum_{l=1}^{k} Y_l - \mathbb{E}\left[\sum_{l=1}^{k} Y_l\right] \geq \frac{k}{4}\right] \tag{34}$$

$$\leq \exp\left(-\frac{2(k/4)^2}{k}\right) = \exp\left(-k/8\right)$$

where the first inequality holds because $\widehat{\mu}_M < \mu - \eta$ implies that at least half of $\widehat{\mu}_l$ for $l \in [k]$ satisfying $\mu > \widehat{\mu}_l + \eta$.

Substitute $k = \lceil 8 \log(1/\delta) \rceil$ and $N = \lceil n/k \rceil$ into the formula of η, we have that,

$$\eta = (12v)^{\frac{1}{1+\epsilon}} \left(\frac{1}{N} \right)^{\frac{\epsilon}{1+\epsilon}} \leq (12v)^{\frac{1}{1+\epsilon}} \left(\frac{k}{n} \right)^{\frac{\epsilon}{1+\epsilon}} \leq (12v)^{\frac{1}{1+\epsilon}} \left(\frac{8 \log(e^{1/8}/\delta)}{n} \right)^{\frac{\epsilon}{1+\epsilon}} , \tag{35}$$

Therefore, with probability at least $1 - \delta$,

$$\widehat{\mu}_M \geq \mu - (12v)^{\frac{1}{1+\epsilon}} \left(\frac{8 \log(e^{1/8}/\delta)}{n} \right)^{\frac{\epsilon}{1+\epsilon}} , \tag{36}$$

Now, we have finished proving the first part of Proposition 2, and the second part follows the same scheme, so we omit it. ∎

Proof [of Theorem 3] This proof is similar to the analysis of Theorem 2, so we only present some critical steps here.

First of all, we can get the same inequality for $\mathbb{E}[R_C(T)]$ as (20).

$$\mathbb{E}[R_C(T)] \leq \ell_C \Delta_C^{\max} + \underbrace{\sum_{t \leq T} \Pr\left[\exists i \in C\backslash\{i^\star\} \text{ s.t. } I_t = i, T_C(t-1) \geq \ell_C \right] \cdot \Delta_C^{\max}}_{(\alpha)} . \tag{37}$$

Then, we analyze term (α) by the same computation as (21)

$$(\alpha) \leq \sum_{t < T} \sum_{s_i = \ell_C}^{t} \sum_{s=0}^{t} \Pr\left[\max_{i \in C\backslash\{i^\star\}} \{\widehat{\mu}_i(s_i) + \tilde{c}_i(s_i)\} \geq \widehat{\mu}_{i^\star}(s) + c_{i^\star}(s) \right] . \tag{38}$$

where we redefine $c_i(t) = (12v)^{\frac{1}{1+\epsilon}} \left(\frac{32 \log(|N(i)|^{1/4}t)}{O_i(t)} \right)^{\frac{\epsilon}{1+\epsilon}}$ and $\tilde{c}_i(t) = (12v)^{\frac{1}{1+\epsilon}} \left(\frac{32 \log(N_C t)}{O_i(t)} \right)^{\frac{\epsilon}{1+\epsilon}}$.

Now, let $\mu_C^{\max} = \max_{i \in C} \mu_i$ and $\ell_C = \lceil \frac{64(24v)^{1/\epsilon} \log(N_C T)}{(\Delta_C^{\min})^{(1+\epsilon)/\epsilon}} \rceil$, then by using a similar trick to (22)-(24), when $\ell_C \leq s_i \leq t$, $\max_{i \in C\backslash\{1\}} \{\widehat{\mu}_i(s_i) + \tilde{c}_i(s_i)\} \geq \widehat{\mu}_{i^\star}(s) + c_{i^\star}(s)$ implies that at least one of the following must hold:

$$\max_{i \in C\backslash\{1\}} \{\widehat{\mu}_i(s_i)\} \geq \mu_C^{\max} + \tilde{c}_i(s_i) , \tag{39}$$

$$\widehat{\mu}_{i^\star}(s) \leq \mu^\star - c_{i^\star}(s) , \tag{40}$$

where the probabilities can be both bounded by $\frac{e^{1/8}}{t^4}$ by using Proposition 2 and similar analysis to (27) and (28).

By further computation, we get that

$$(\alpha) \leq \sum_{t < T} \sum_{s_i = \ell_C}^{t} \sum_{s=0}^{t} \Pr\left[\max_{i \in C\backslash\{i^\star\}} \{\widehat{\mu}_i(s_i) + \tilde{c}_i(s_i)\} \geq \widehat{\mu}_{i^\star}(s) + c_{i^\star}(s) \right]$$

$$\leq \sum_{t < T} \left(\sum_{s_i = \ell_C}^{t} \sum_{s=0}^{t} \frac{2e^{1/8}}{t^4} \right) \leq \sum_{t < T} \frac{2e^{1/8}}{t^2} \leq \frac{e^{1/8}\pi^2}{3} , \tag{41}$$

and furthermore, by (37), we have

$$\mathbb{E}[R_C(T)] \leq \ell_C \Delta_C^{\max} + (\alpha) \cdot \Delta_C^{\max}$$

$$\leq \ell_C \Delta_C^{\max} + \frac{\pi^2}{3} \Delta_C^{\max}$$

$$\leq 64 \frac{(24v)^{1/\epsilon} \Delta_C^{\max}}{(\Delta_C^{\min})^{(1+\epsilon)/\epsilon}} \log(N_C T) + \left(1 + \frac{e^{1/8}\pi^2}{3} \right) \Delta_C^{\max} , \tag{42}$$

Finally, summing over all cliques in $\mathcal{C}$ and taking the infimum over all possible coverings $\mathcal{C}$ yields the aforementioned upper bound of RUNE-MoM. ∎

2 OMITTED PROOFS FOR SECTION 3.2

In this section, we present the proofs for RAAE-TEM and RAAE-MoM omitted in the second section of our main results.

2.1 ANALYSIS FOR RAAE-TEM

In order to prove Theorem 4 we need one additional lemma, which is described as following. It shows that, after the active arm elimination phase of epoch r, the remained arms are at most $4\varepsilon_r$-suboptimal with high probability.

Lemma 3 *For all epoch r, with probability at least $1 - 1/T$ we have $\mu_i \leq \mu^\star + 4\varepsilon_r$ for all $i \in V_{r+1}$.*

Proof [of Lemma 3] In each epoch r, the player observed the reward of each arm in V_r for at least n_r times. Note that by Proposition 1 the rewards of the arms are distributed independently from the feedback graphs. Given an epoch r we denote by $\mathbb{E}_r$ the event that $|\widehat{\mu}_i(t-1) - \mu_i| \leq \varepsilon_r$ holds for all $i \in V_\epsilon$ and denote $\mathbb{E} = \cup_{r \geq 1} \mathbb{E}_r$. Therefore by Proposition 1

$$\Pr[\mathbb{E}_r] \geq 1 - \frac{1}{KT} \, , \tag{43}$$

by taking the union bound on each V_r, we have that

$$\Pr[\mathbb{E}] \geq 1 - \frac{1}{T} \, . \tag{44}$$

In the remainder of the proof, we assume that $\mathbb{E}$ holds. So for any epoch r and any arm $i \in V_r$, we have $|\widehat{\mu}(i, O_i(t-1)) - \mu_i| \leq \varepsilon_r$. As a result, for any epoch r and any two arms i and i', there must be

$$|(\widehat{\mu}_i(t-1) - \widehat{\mu}_{i'}(t-1)) - (\mu_i - \mu_{i'})| \leq 2\varepsilon_r \, . \tag{45}$$

Next, we argue that under $\mathbb{E}$ the optimal arm $i^\star$ is never eliminated. Indeed, for any epoch r, we denote the arm $i_r = \arg\max_{i \in V_r} \widehat{\mu}_i(t-1)$ and $\widehat{\mu}_r^\star = \widehat{\mu}_{i_r}(t-1)$, it is obvious to see that

$$\widehat{\mu}_r^\star - \widehat{\mu}_{i^\star}(t-1) \leq |(\widehat{\mu}_{i_r}(t-1) - \widehat{\mu}_{i^\star}(t-1)) - (\mu_{i_r} - \mu_{i^\star})| \leq 2c_r \, , \tag{46}$$

so the algorithm doesn't eliminate $i^\star$.

Next, we show that under $\mathbb{E}$, in any epoch r the algorithm eliminate all viable arms i with sub-optimal gap $\Delta_i > 4\varepsilon_r$. Thus, since arm $i^\star$ remains viable, we have that

$$\widehat{\mu}_{i^\star}(t-1) - \widehat{\mu}(i, O_i(t-1)) \geq \widehat{\mu}_r^\star - \widehat{\mu}_i(t-1) \geq \Delta_i - 2\varepsilon_r > 2\varepsilon_r \, , \tag{47}$$

which means arm i is eliminated by the algorithm. Thus, for all the rest arms i in V_{r+1}, we have that

$$\mu_i - \mu^\star \leq 4\varepsilon_r \, . \tag{48}$$

Until now, we have finished the proof of Lemma 3. ∎

Besides, we list the following two lemmas that will be used in the proof of Theorem 4. The first lemma shows that, the expected regret of the algorithm until the first arm in $V^{(\tau)}$ has been eliminated , is bounded by $O(\log(KT))$. Furthermore, the expected regret of the algorithm after that epoch is also bounded by $O(\log(KT))$. The second lemma shows that the expected regret of Algorithm 2 after we have eliminated all arms not in $V^{(\tau)}$, is at most $O(\log(KT))$.

Lemma 4 *Let $\bar{\Delta} = \max_{i \in V^{(\tau)}} \Delta_i$ be the largest gap of arms in $V^{(\tau)}$ and $\tilde{r} = \lfloor \log_2(4/\bar{\Delta})/\epsilon - 1 \rfloor$. Then, the expected regret of Algorithm 2 up to epoch $\tilde{r}$ is at most*

$$\frac{160\alpha(20v)^{1/\epsilon}}{\bar{\Delta}^{1/\epsilon}} \log(2KT). \tag{49}$$

Also, the expected regret of Algorithm 2 after epoch $\tilde{r}$ is at most

$$\sum_{i \in V^{(\tau)}} \frac{160(20v)^{1/\epsilon}}{\Delta_i^{1/\epsilon}} \log(2KT). \tag{50}$$

Proof [of Lemma 4] By Lemma 3, the instantaneous regret for each round on epoch r is at most $4\varepsilon_{r-1} = 8\varepsilon_r$. Note that each epoch spends at most α rounds. Then, the expected regret up to round $\tilde{r}$ is at most

$$
\begin{aligned}
\sum_{r=1}^{\tilde{r}} \alpha \cdot n_r \cdot 8\varepsilon_r &\leq 80\alpha(5v)^{1/\epsilon} \log(2KT) \sum_{r=1}^{\tilde{r}} \frac{1}{\varepsilon_r^{1/\epsilon}} \\
&= 80\alpha(5v)^{1/\epsilon} \log(2KT) \sum_{r=1}^{\tilde{r}} 2^{r+1} \\
&\leq 80\alpha(5v)^{1/\epsilon} \log(2KT) 2^{\tilde{r}+2} \leq \frac{160\alpha(20v)^{1/\epsilon} \log(2KT)}{\bar{\Delta}^{1/\epsilon}} \,,
\end{aligned}
\tag{51}
$$

where the last inequality comes from the fact $\tilde{r} = \lfloor \log_2(4/\bar{\Delta})/\epsilon - 1 \rfloor$.

Let $\tilde{r}_i = \lfloor \log_2(4/\bar{\Delta}_i)/\epsilon - 1 \rfloor$, the number of epochs until arm i is removed from V_r, which means that it takes $\tilde{r}_i + 1$ epochs in order for $4\varepsilon_{\tilde{r}_i+1} = 2^{2-\epsilon(\tilde{r}_i+2)} < \bar{\Delta}_i$. We shall assume that the algorithm finished after $\tilde{r}_u$ epochs, where u is the arm in V with the minimum nonzero gap. Note that each epoch spends at most $|V_r|$ rounds after we have eliminated all arms not in $V^{(\tau)}$. Thus, the expected regret for the remaining epochs is at most

$$
\begin{aligned}
\sum_{r=\tilde{r}+1}^{\tilde{r}_u} |V_r| \cdot n_r \cdot 8\varepsilon_r &\leq 80(5v)^{1/\epsilon} \log(2KT) \sum_{r=1}^{\tilde{r}} \frac{|V_r|}{\varepsilon_r^{1/\epsilon}} \\
&= 80(5v)^{1/\epsilon} \log(2KT) \sum_{i \in V^{(\tau)}} \sum_{r=\tilde{r}+1}^{\tilde{r}_i} \frac{1}{\varepsilon_r^{1/\epsilon}} \\
&\leq 80(5v)^{1/\epsilon} \log(2KT) \sum_{i \in V^{(\tau)}} \sum_{r=1}^{\tilde{r}_i} 2^{r+1} \\
&\leq \sum_{i \in V^{(\tau)}} \frac{160(20v)^{1/\epsilon} \log(2KT)}{\Delta_i^{1/\epsilon}} \,,
\end{aligned}
\tag{52}
$$

where the last inequality is induced by the same computation as (51). ∎

Lemma 5 *The expected regret of Algorithm 2 after we have eliminated all arms not in $V^{(\tau)}$, is at most*

$$
\sum_{i \in V^{(\tau)}} \frac{160(20v)^{1/\epsilon}}{\Delta_i^{1/\epsilon}} \log(2KT).
\tag{53}
$$

Proof [of Lemma 5] Let $\tilde{r}_i = \lfloor \log_2(4/\bar{\Delta}_i)/\epsilon - 1 \rfloor$, the number of epochs until arm i is removed from V_r, which means that it takes $\tilde{r}_i + 1$ epochs in order for $4\varepsilon_{\tilde{r}_i+1} = 2^{2-\epsilon(\tilde{r}_i+2)} < \bar{\Delta}_i$. We're going to talk about two categories.

When $\tau \leq K - 1$, the expected regret of Algorithm 2 after we have eliminated all arms not in $V^{(\tau)}$ equals that of Algorithm 2 after epoch $\tilde{r}$, then the conclusion is directly induced by the second part of Lemma 4.

When $\tau > K - 1$, then $V^{(\tau)} = V \backslash \{i^\star\}$, the expected regret equals that of the algorithm terminating with $V_r = \{i^\star\}$. We shall assume that the algorithm finished after $\tilde{r}_u$ epochs, where u is the arm in $V^{(\tau)}$ with the minimum gap. Note that each epoch spends at most $|V_r|$ rounds after we have eliminated all arms not in $V^{(\tau)}$. Thus, the expected regret for the remaining

epochs is at most

$$
\begin{aligned}
\sum_{r=1}^{\tilde{r}_u} |V_r| \cdot n_r \cdot 8\varepsilon_r &\le 80(5v)^{1/\epsilon} \log(2KT) \sum_{r=1}^{\tilde{r}} \frac{|V_r|}{\varepsilon_r^{1/\epsilon}} \\
&= 80(5v)^{1/\epsilon} \log(2KT) \sum_{i \in V^{(\tau)}} \sum_{r=1}^{\tilde{r}_i} \frac{1}{\varepsilon_r^{1/\epsilon}} \\
&= 80(5v)^{1/\epsilon} \log(2KT) \sum_{i \in V^{(\tau)}} \sum_{r=1}^{\tilde{r}_i} 2^{r+1} \\
&\le \sum_{i \in V^{(\tau)}} \frac{160(20v)^{1/\epsilon} \log(2KT)}{\Delta_i^{1/\epsilon}} ,
\end{aligned}
\tag{54}
$$

where the last inequality is induced by the same computation as (51). ∎

Proof [of Theorem 4] Here, we use a proof sketch similar to [Cohen et al., 2016]. Firstly, we need to bound R, which is defined as the total number of epochs that the algorithm runs for. As the fact that the number T_r of rounds elapsed during epoch r, is lower bounded by $n_r = \lceil (5(5v)^{1/\epsilon} \log(2KT)) \cdot (2^{(1+\epsilon)})^{r+1} \rceil$, there must be

$$
T \ge \sum_{r=1}^{R} T_r \ge (5(5v)^{1/\epsilon} \log(2KT)) \sum_{r=1}^{R} (2^{(1+\epsilon)})^{r+1} .
\tag{55}
$$

Then, by computing the RHS of the above inequality, we can obtain an upper bound

$$
R \le \bar{r} = \frac{1}{1+\epsilon} \log_2 \left(\frac{(2^{1+\epsilon} - 1)T}{20(5v)^{1/\epsilon} 4^\epsilon \log(2KT)} + 1 \right) .
\tag{56}
$$

Next, we attempt to bound the expected regret of the algorithm. By Lemma 3 and the union bound, the total failure probability of the robust mean estimators is at most $\bar{r}/T$. Then the expected regret of the algorithm is at most the expected regret conditioned on the success of the estimation of the means plus $(\bar{r}/T) \cdot (T\Delta_{\max}) = O(\Delta_{\max} \log T)$ by Eq. (56). Thus it remains to bound the conditioned expected regret.

In the remainder of the proof, we assume that the robust mean estimators always succeed. Therefore, the regret of the algorithm is at most the regret of an algorithm terminating with $V_r = \{i^\star\}$. In addition, we assume that T is large enough for this to happen.

Let $\bar{\Delta} = \max_{i \in V^{(\tau)}} \Delta_i$ be the largest gap of arms in $V^{(\tau)}$ and $\tilde{r} = \lfloor \log_2(4/\bar{\Delta})/\epsilon - 1 \rfloor$. Thus, it takes $\tilde{r} + 1$ epochs in order for $4\varepsilon_{\tilde{r}} = 2^{2-\epsilon(\tilde{r}+1)} < \bar{\Delta}$, which means that the algorithm eliminates all arms in $V \backslash V^{(\tau)}$ until epoch $\tilde{r}$. Now we will split in the following two cases:

If $\tau > K - 1$, then $V^{(\tau)} = V \backslash \{i^\star\}$. Thus the regret of the algorithm is upper bounded by the the regret until the algorithm finished running with $V_r = \{i^\star\}$, which can be directly obtained by (53) of Lemma 5.

If $\tau \le K - 1$, then splitting by epoch $\tilde{r}$, we can bound the expected regret before and after that, using (49) and (50) in Lemma 4 respectively. Since $\bar{\Delta} \ge \Delta_i$ for all $i \in V^{(\tau)}$, we have that

$$
\frac{\alpha}{\bar{\Delta}^{1/\epsilon}} \le \frac{\alpha}{|V^{(\tau)}|} \sum_{i \in V^{(\tau)}} \frac{1}{\Delta_i^{1/\epsilon}} = \sum_{i \in V^{(\tau)}} \frac{1}{\Delta_i^{1/\epsilon}} ,
\tag{57}
$$

by definition of $V^{(\tau)}$. Thus we notice that the expected regret of the algorithm up to epoch $\tilde{r}$ is at most the regret from round $\tilde{r} + 1$ thereafter. Then we can also bound the expected regret of the algorithm by using (50) of Lemma 4.

Finally, we obtain a regret $O(\sum_{i \in V^{(\tau)}} (\frac{v}{\Delta_i})^{1/\epsilon} \log T + \Delta_{\max} \log T)$ by the assumption $T \ge K$. ∎

2.2 ANALYSIS FOR RAAE-MOM

Following the same proof sketch of Theorem 4, we can obtain the property described in Lemma 3. It shows that, after the active arm elimination phase of epoch r, the remained arms are at most $4\varepsilon_r$-suboptimal with high probability.

Besides, we list the following two lemmas that will be used in the proof of Theorem 5. The first lemma shows that, the expected regret of the algorithm until the first arm in $V^{(\tau)}$ has been eliminated, is bounded by $O(\log(KT))$. Furthermore, the expected regret of the algorithm after that epoch is also bounded by $O(\log(KT))$. The second lemma shows that the expected regret of RUNE-MoM after we have eliminated all arms not in $V^{(\tau)}$, is at most $O(\log(KT))$.

Lemma 6 *Let $\bar{\Delta} = \max_{i \in V^{(\tau)}} \Delta_i$ be the largest gap of arms in $V^{(\tau)}$ and $\tilde{r} = \lfloor \log_2(4/\bar{\Delta})/\epsilon - 1 \rfloor$. Then, the expected regret of RAAE-MoM up to epoch $\tilde{r}$ is at most*

$$\frac{256\alpha(48v)^{1/\epsilon}}{\bar{\Delta}^{1/\epsilon}} \log(2KT). \tag{58}$$

Also, the expected regret of RAAE-MoM after epoch $\tilde{r}$ is at most

$$\sum_{i \in V^{(\tau)}} \frac{256(48v)^{1/\epsilon}}{\Delta_i^{1/\epsilon}} \log(2KT). \tag{59}$$

Proof [of Lemma 6] By Lemma 3, the instantaneous regret for each round on epoch r is at most $4\varepsilon_{r-1} = 8\varepsilon_r$. Note that each epoch spends at most α rounds. Then, the expected regret up to round $\tilde{r}$ is at most

$$
\begin{aligned}
\sum_{r=1}^{\tilde{r}} \alpha \cdot n_r \cdot 8\varepsilon_r &\le 128\alpha(12v)^{1/\epsilon} \log(2e^{1/8}KT) \sum_{r=1}^{\tilde{r}} \frac{1}{\varepsilon_r^{1/\epsilon}} \\
&= 128\alpha(12v)^{1/\epsilon} \log(2e^{1/8}KT) \sum_{r=1}^{\tilde{r}} 2^{r+1} \\
&\le 128\alpha(12v)^{1/\epsilon} \log(2e^{1/8}KT) 2^{\tilde{r}+2} \le \frac{256\alpha(48v)^{1/\epsilon} \log(2e^{1/8}KT)}{\bar{\Delta}^{1/\epsilon}},
\end{aligned} \tag{60}
$$

where the last inequality comes from the fact $\tilde{r} = \lfloor \log_2(4/\bar{\Delta})/\epsilon - 1 \rfloor$.

Let $\tilde{r}_i = \lfloor \log_2(4/\bar{\Delta}_i)/\epsilon - 1 \rfloor$, the number of epochs until arm i is removed from V_r, which means that it takes $\tilde{r}_i + 1$ epochs in order for $4\varepsilon_{\tilde{r}_i+1} = 2^{2-\epsilon(\tilde{r}_i+2)} < \bar{\Delta}_i$. We shall assume that the algorithm finished after $\tilde{r}_u$ epochs, where u is the arm in V with the minimum nonzero gap. Note that each epoch spends at most $|V_r|$ rounds after we have eliminated all arms not in $V^{(\tau)}$. Thus, the expected regret for the remaining epochs is at most

$$
\begin{aligned}
\sum_{r=\tilde{r}+1}^{\tilde{r}_u} |V_r| \cdot n_r \cdot 8\varepsilon_r &\le 128(12v)^{1/\epsilon} \log(2e^{1/8}KT) \sum_{r=1}^{\tilde{r}} \frac{|V_r|}{\varepsilon_r^{1/\epsilon}} \\
&= 128(12v)^{1/\epsilon} \log(2e^{1/8}KT) \sum_{i \in V^{(\tau)}} \sum_{r=\tilde{r}+1}^{\tilde{r}_i} \frac{1}{\varepsilon_r^{1/\epsilon}} \\
&\le 128(12v)^{1/\epsilon} \log(2e^{1/8}KT) \sum_{i \in V^{(\tau)}} \sum_{r=1}^{\tilde{r}_i} 2^{r+1} \\
&\le \sum_{i \in V^{(\tau)}} \frac{256(48v)^{1/\epsilon} \log(2e^{1/8}KT)}{\Delta_i^{1/\epsilon}},
\end{aligned} \tag{61}
$$

where the last inequality is induced by the same computation as (60). ∎

Lemma 7 *The expected regret of RAAE-MoM after we have eliminated all arms not in $V^{(\tau)}$, is at most*

$$\sum_{i \in V^{(\tau)}} \frac{256(48v)^{1/\epsilon}}{\Delta_i^{1/\epsilon}} \log(2e^{1/8}KT). \tag{62}$$

Proof [of Lemma 7] Let $\tilde{r}_i = \lfloor \log_2(4/\bar{\Delta}_i)/\epsilon - 1 \rfloor$, the number of epochs until arm i is removed from V_r, which means that it takes $\tilde{r}_i + 1$ epochs in order for $4\varepsilon_{\tilde{r}_i+1} = 2^{2-\epsilon(\tilde{r}_i+2)} < \bar{\Delta}_i$. We're going to talk about two categories.

When $\tau \leq K - 1$, the expected regret of RAAE-MoM after we have eliminated all arms not in $V^{(\tau)}$ equals that of RAAE-MoM after epoch $\tilde{r}$, then the conclusion is directly induced by the second part of Lemma 6.

When $\tau > K - 1$, then $V^{(\tau)} = V \setminus \{i^\star\}$, the expected regret equals that of the algorithm terminating with $V_r = \{i^\star\}$. We shall assume that the algorithm finished after $\tilde{r}_u$ epochs, where u is the arm in $V^{(\tau)}$ with the minimum gap. Note that each epoch spends at most $|V_r|$ rounds after we have eliminated all arms not in $V^{(\tau)}$. Thus, the expected regret for the remaining epochs is at most

$$
\begin{aligned}
\sum_{r=1}^{\tilde{r}_u} |V_r| \cdot n_r \cdot 8\varepsilon_r &\leq 128(12v)^{1/\epsilon} \log(2e^{1/8}KT) \sum_{r=1}^{\tilde{r}} \frac{|V_r|}{\varepsilon_r^{1/\epsilon}} \\
&= 128(12v)^{1/\epsilon} \log(2e^{1/8}KT) \sum_{i \in V^{(\tau)}} \sum_{r=1}^{\tilde{r}_i} \frac{1}{\varepsilon_r^{1/\epsilon}} \\
&= 128(12v)^{1/\epsilon} \log(2e^{1/8}KT) \sum_{i \in V^{(\tau)}} \sum_{r=1}^{\tilde{r}_i} 2^{r+1} \\
&\leq \sum_{i \in V^{(\tau)}} \frac{256(48v)^{1/\epsilon} \log(2e^{1/8}KT)}{\Delta_i^{1/\epsilon}} ,
\end{aligned}
\tag{63}
$$

where the last inequality is induced by the same computation as (60). ∎

Proof [of Theorem 5] Here, we use a proof sketch similar to [Cohen et al., 2016]. Firstly, we need to bound R, which is defined as the total number of epochs that the algorithm runs for. As the fact that the number T_r of rounds elapsed during epoch r, is lower bounded by $n_r = \lceil \frac{(8(12v)^{1/\epsilon} \log(2e^{1/8}KT))}{\varepsilon_r^{(1+\epsilon)/\epsilon}} \rceil$, there must be

$$
T \geq \sum_{r=1}^{R} T_r \geq (8(12v)^{1/\epsilon} \log(2e^{1/8}KT)) \sum_{r=1}^{R} (2^{(1+\epsilon)})^{r+1} .
\tag{64}
$$

Then, by computing the RHS of the above inequality, we can obtain an upper bound

$$
R \leq \bar{r} = \frac{1}{1+\epsilon} \log_2 \left(\frac{(2^{1+\epsilon} - 1)T}{48(12v)^{1/\epsilon}4^\epsilon \log(2e^{1/8}KT)} + 1 \right) .
\tag{65}
$$

Next, we attempt to bound the expected regret of the algorithm. By Lemma 3 and the union bound, the total failure probability of the robust mean estimators is at most $\bar{r}/T$. Then the expected regret of the algorithm is at most the expected regret conditioned on the success of the estimation of the means plus $(\bar{r}/T) \cdot (T\Delta_{\max}) = O(\Delta_{\max} \log T)$ by Eq. (65). Thus it remains to bound the conditioned expected regret.

In the remainder of the proof, we assume that the robust mean estimators always succeed. Therefore, the regret of the algorithm is at most the regret of an algorithm terminating with $V_r = \{i^\star\}$. In addition, we assume that T is large enough for this to happen.

Let $\bar{\Delta} = \max_{i \in V^{(\tau)}} \Delta_i$ be the largest gap of arms in $V^{(\tau)}$ and $\tilde{r} = \lfloor \log_2(4/\bar{\Delta})/\epsilon - 1 \rfloor$. Thus, it takes $\tilde{r} + 1$ epochs in order for $4\varepsilon_{\tilde{r}} = 2^{2-\epsilon(\tilde{r}+1)} < \bar{\Delta}$, which means that the algorithm eliminates all arms in $V \setminus V^{(\tau)}$ until epoch $\tilde{r}$. Now we will split in the following two cases:

If $\tau > K - 1$, then $V^{(\tau)} = V \setminus \{i^\star\}$. Thus the regret of the algorithm is upper bounded by the the regret until the algorithm finished running with $V_r = \{i^\star\}$, which can be directly obtained by (62) of Lemma 7.

If $\tau \leq K - 1$, then splitting by epoch $\tilde{r}$, we can bound the expected regret before and after that, using (58) and (59) in Lemma 4 respectively. Since $\bar{\Delta} \geq \Delta_i$ for all $i \in V^{(\tau)}$, we have that

$$
\frac{\alpha}{\bar{\Delta}^{1/\epsilon}} \leq \frac{\alpha}{|V^{(\tau)}|} \sum_{i \in V^{(\tau)}} \frac{1}{\Delta_i^{1/\epsilon}} = \sum_{i \in V^{(\tau)}} \frac{1}{\Delta_i^{1/\epsilon}} ,
\tag{66}
$$

by definition of $V^{(\tau)}$. Thus we notice that the expected regret of the algorithm up to epoch $\tilde{r}$ is at most the regret from round $\tilde{r}+1$ thereafter. Then we can also bound the expected regret of the algorithm by using (50) of Lemma 4.

Finally, we obtain a regret $O(\sum_{i \in V^{(\tau)}} (\frac{v}{\Delta_i})^{1/\epsilon} \log T + \Delta_{\max} \log T)$ by the assumption $T \geq K$. ∎