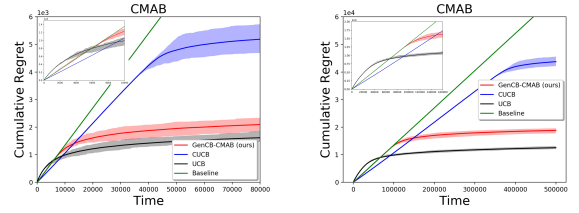
8 More Experimental Results

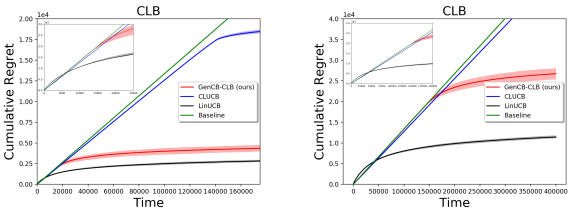
In this section, we present more experimental results for CMAB, CLB, CCCB and MV-CPB, which are shown in Figure 2, 3, 4 and 5, respectively. The parameter settings of our experiments are described in Section 5 of the main paper.



(a) $K = 24, \alpha = 0.05$ (b) $K = 144, \alpha = 0.05$

(c) $K = 72, \alpha = 0.1$ (d) $K = 72, \alpha = 0.15$

Figure 2: Experiments for CMAB.



(a) $d = 5, \alpha = 0.01$ (b) $d = 9, \alpha = 0.01$

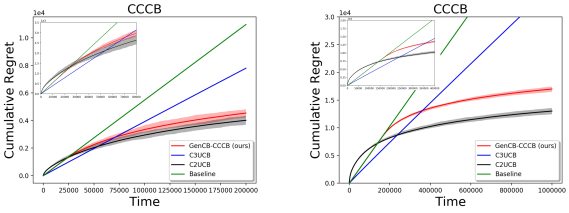
(c) $d = 5, \alpha = 0.02$ (d) $d = 5, \alpha = 0.03$

Figure 3: Experiments for CLB.

9 Technical Tools

We present some technical tools (Lemmas 1,2 and Facts 1,2) below.

Lemma 1. *For $m \geq 2$, $c_1 = 2$, $c_2 = 2\mathbb{E}[N_0(\tau - 1)] \geq 4$, $c_3 = \Delta_0 + \alpha\mu_0 \in (0, 1)$, $c_4 = 8H$ where $H \geq 2$, define function $g_1(m) = -c_3 m + c_1\sqrt{m}\ln(c_2 m) + c_4 \ln(m)$. Then,*



(a) $d = 5, \alpha = 0.01$ (b) $d = 9, \alpha = 0.01$

(c) $d = 5, \alpha = 0.02$ (d) $d = 5, \alpha = 0.03$

Figure 4: Experiments for CCCB.

$g_1(m)$ can be upper bounded by

$$g_1(m) \leq \frac{132 c_1 c_4}{c_3}\left[\ln\left(\frac{10\sqrt{c_2 c_4}}{c_3}\right)\right]^2.$$

Proof. Taking the derivative of $g_1(m)$, we obtain

$$g_1'(m) = -c_3 + \frac{c_1 \ln(c_2 m) + 2c_1}{2\sqrt{m}} + \frac{c_4}{m}$$

Let $\tilde{m}_1 = \frac{c_4}{c_3}$, $\tilde{m}_2 = \frac{100 c_4 [\ln(c_2 c_4 / c_3^2)]^2}{c_3^2}$. Then, we have

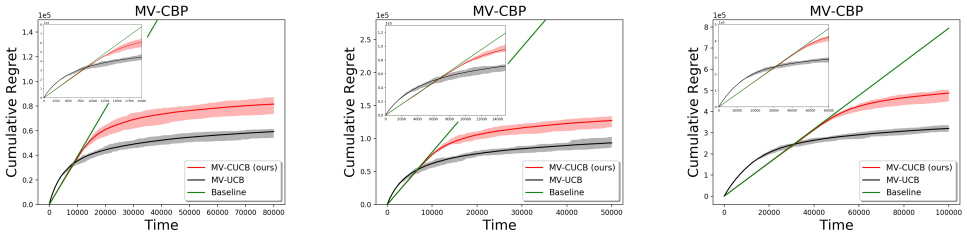
$$g_1'(\tilde{m}_1) = \frac{c_1 \ln(c_2 \tilde{m}_1) + 2c_1}{2\sqrt{\tilde{m}_1}} > 0$$

and

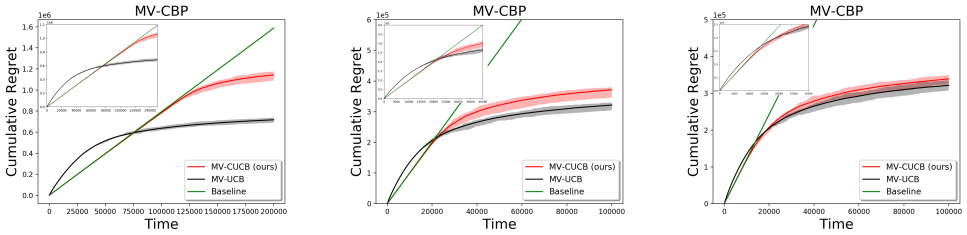
$$g_1'(\tilde{m}_2) = -c_3 + \frac{c_1 \ln(\frac{100 c_2 c_4 [\ln(c_2 c_4/c_3^2)]^2}{c_3^2}) + 2c_1}{10\sqrt{c_4}\ln(c_2 c_4/c_3^2)} \cdot \frac{c_3}{2} + \frac{c_3}{100[\ln(c_2 c_4/c_3^2)]^2} \cdot c_3.$$

Since

$$\begin{aligned}
&c_1 \ln(\frac{100 c_2 c_4 [\ln(c_2 c_4/c_3^2)]^2}{c_3^2}) + 2c_1 \\
=&c_1 \ln(\frac{100 c_2 c_4}{c_3^2}) + 2c_1 \ln(\ln(\frac{c_2 c_4}{c_3^2})) + 2c_1 \\
\leq&c_1 \ln(\frac{100 c_2 c_4}{c_3^2}) + 2c_1 \ln(\frac{c_2 c_4}{c_3^2}) + 2c_1 \\
\leq&3c_1 \ln(\frac{100 c_2 c_4}{c_3^2}) + 2c_1 \\
=&6\ln(\frac{c_2 c_4}{c_3^2}) + 6\ln(100) + 4 \\
<&10\sqrt{c_4}\ln(\frac{c_2 c_4}{c_3^2}),
\end{aligned}$$



(a) $K = 24, \alpha = 0.05, \rho = 30$ (b) $K = 24, \alpha = 0.05, \rho = 60$ (c) $K = 72, \alpha = 0.05, \rho = 60$



(d) $K = 144, \alpha = 0.05, \rho = 60$ (e) $K = 72, \alpha = 0.1, \rho = 60$ (f) $K = 72, \alpha = 0.15, \rho = 60$

Figure 5: Experiments for MV-CBP.

we have

$$\frac{c_1 \ln(\frac{100c_2c_4[\ln(c_2c_4/c_3^2)]^2}{c_3^2}) + 2c_1}{10\sqrt{c_4}\ln(c_2c_4/c_3^2)} < 1.$$

In addition, it is clear that $\frac{c_3}{100[\ln(c_2c_4/c_3^2)]^2} < \frac{1}{2}$. Thus, we have

$$\begin{aligned} g_1'(\tilde{m}_2) =& -c_3 + \frac{c_1 \ln(\frac{100c_2c_4[\ln(c_2c_4/c_3^2)]^2}{c_3^2}) + 2c_1}{10\sqrt{c_4}\ln(c_2c_4/c_3^2)} \cdot \frac{c_3}{2} \\ &+ \frac{c_3}{100[\ln(c_2c_4/c_3^2)]^2} \cdot c_3 < 0. \end{aligned}$$

Thus,

$$\begin{aligned} g_1(m) \le& -c_3\tilde{m}_1 + c_1\sqrt{\tilde{m}_2}\ln(c_2\tilde{m}_2) + c_4\ln(\tilde{m}_2) \\ \le& (c_1\sqrt{\tilde{m}_2} + c_4)\ln(c_2\tilde{m}_2) \\ \le& (\frac{10c_1\sqrt{c_4}}{c_3}\ln(\frac{c_2c_4}{c_3^2}) + c_4)\cdot \\ & \ln(\frac{100c_2c_4[\ln(c_2c_4/c_3^2)]^2}{c_3^2}) \\ \le& \frac{10c_1\sqrt{c_4} + c_4}{c_3} \cdot \ln(\frac{c_2c_4}{c_3^2}) \cdot 3\ln(\frac{100c_2c_4}{c_3^2}) \\ \le& \frac{33c_1c_4}{c_3}\left[\ln\left(\frac{100c_2c_4}{c_3^2}\right)\right]^2 \\ =& \frac{132c_1c_4}{c_3}\left[\ln\left(\frac{10\sqrt{c_2c_4}}{c_3}\right)\right]^2. \end{aligned}$$

□

Lemma 2. *For $m \ge 2$, $c_1 = 2(5+\rho)\sqrt{2K}$, $c_2 = 6K\mathbb{E}[N_0(\tau-1)]$, $c_3 = \Delta_0 + \alpha\mathbf{MV}_0 \in (2, \rho)$, $c_4 = 8\sqrt{2}K + 12(5+\rho)^2(H_1^{\mathrm{MV}} + 4H_2^{\mathrm{MV}}) > 8\sqrt{2}K + 12(5+\rho)(K-1) + 48(K-1)$ where $\rho > \frac{2}{\alpha\mu_0} > 2$, $c_4 > 3c_1$, $c_4 > 12c_3$, define function $g_2(m) = -c_3m + c_1\sqrt{m}\ln(c_2m) + c_4\ln(m)$. Then, $g_2(m)$ can be upper bounded by*

$$g_2(m) \le \frac{48c_1c_4}{c_3}\left[\ln(\frac{3\sqrt{c_2}c_4}{c_3})\right]^2.$$

Proof. Taking the derivative of $g_2(m)$, we obtain

$$g_2'(m) = -c_3 + \frac{c_1\ln(c_2m) + 2c_1}{2\sqrt{m}} + \frac{c_4}{m}$$

Let $\tilde{m}_1 = \frac{c_4}{c_3}, \tilde{m}_2 = \frac{9c_4^2[\ln(c_2c_4^2/c_3^2)]^2}{c_3^2}$. Then, we have

$$g_2'(\tilde{m}_1) = \frac{c_1\ln(c_2\tilde{m}_1) + 2c_1}{2\sqrt{\tilde{m}_1}} > 0$$

and

$$\begin{aligned} g_2'(\tilde{m}_2) =& -c_3 + \frac{c_1\ln(\frac{9c_2c_4^2[\ln(c_2c_4^2/c_3^2)]^2}{c_3^2}) + 2c_1}{3c_4\ln(c_2c_4^2/c_3^2)} \cdot \frac{c_3}{2} \\ &+ \frac{c_3}{9c_4[\ln(c_2c_4^2/c_3^2)]^2} \cdot c_3. \end{aligned}$$

Since

$$\begin{aligned} & c_1\ln(\frac{9c_2c_4^2[\ln(c_2c_4^2/c_3^2)]^2}{c_3^2}) + 2c_1 \\ =& c_1\ln(\frac{c_2c_4^2}{c_3^2}) + c_1\ln(9) + 2c_1\ln(\ln(\frac{c_2c_4^2}{c_3^2})) + 2c_1 \\ \le& c_1\ln(\frac{c_2c_4^2}{c_3^2}) + c_1\ln(9) + 2c_1\ln(\frac{c_2c_4^2}{c_3^2}) + 2c_1 \\ =& 3c_1\ln(\frac{c_2c_4^2}{c_3^2}) + (2+\ln(9))c_1 \\ <& 9c_1\ln(\frac{c_2c_4^2}{c_3^2}) \end{aligned}$$

$$<3c_4\ln(\frac{c_2c_4^2}{c_3^2}),$$

we have

$$\frac{c_1\ln(\frac{9c_2c_4^2[\ln(c_2c_4^2/c_3^2)]^2}{c_3^2})+2c_1}{3c_4\ln(c_2c_4^2/c_3^2)}<1.$$

In addition, it is clear that $\frac{c_3}{9c_4[\ln(c_2c_4^2/c_3^2)]^2}<\frac{1}{2}$. Thus, we have

$$\begin{aligned}g_2'(\tilde{m}_2)=&-c_3+\frac{c_1\ln(\frac{9c_2c_4^2[\ln(c_2c_4^2/c_3^2)]^2}{c_3^2})+2c_1}{3c_4\ln(c_2c_4^2/c_3^2)}\cdot\frac{c_3}{2}\\&+\frac{c_3}{9c_4[\ln(c_2c_4^2/c_3^2)]^2}\cdot c_3<0.\end{aligned}$$

Thus,

$$\begin{aligned}g_2(m)\leq&-c_3\tilde{m}_1+c_1\sqrt{\tilde{m}_2}\ln(c_2\tilde{m}_2)+c_4\ln(\tilde{m}_2)\\\leq&(c_1\sqrt{\tilde{m}_2}+c_4)\ln(c_2\tilde{m}_2)\\\leq&(\frac{3c_1c_4}{c_3}\ln(\frac{c_2c_4^2}{c_3^2})+c_4)\ln(\frac{9c_2c_4^2[\ln(c_2c_4^2/c_3^2)]^2}{c_3^2})\\\leq&\frac{4c_1c_4}{c_3}\ln(\frac{c_2c_4^2}{c_3^2})\cdot3\ln(\frac{9c_2c_4^2}{c_3^2})\\\leq&\frac{12c_1c_4}{c_3}\left[\ln(\frac{9c_2c_4^2}{c_3^2})\right]^2\\=&\frac{48c_1c_4}{c_3}\left[\ln(\frac{3\sqrt{c_2}c_4}{c_3})\right]^2.\end{aligned}$$

□

Fact 1 (Lemma 9 in (Kazerouni et al. 2017))**.** *For any $m\geq 2$ and $c_1,c_2,c_3>0$, the following holds*

$$-c_3m+c_1\sqrt{m}\ln(c_2m)\leq\frac{16c_1^2}{9c_3}\left[\ln\left(\frac{2c_1\sqrt{c_2}e}{c_3}\right)\right]^2.$$

Fact 2 (Lemma 10 in (Kazerouni et al. 2017))**.** *Let c_1 and c_2 be two positive constants such that $\ln(c_1c_2)\geq 1$. Then, any $z>0$ satisfying $z\leq c_1\ln(c_2z)$ also satisfies $z\leq 2c_1\ln(c_1c_2)$.*

10 Algorithm Pseudo-code and Proof for CMAB

Algorithm 3 presents the algorithm pseudo-code of GenCB-CMAB for CMAB, and we give the detailed proof of Theorem 2 in the following.

Proof. First, we prove that GenCB-CMAB satisfies the sample-path reward constraint Eq. (2) by induction. At timestep $t=1$, since the LHS of the if statement (in Line 3 of Algorithm 3) is zero and RHS is positive, GenCB-CMAB will pull the default arm x_0 and receive reward $\mu_0\geq(1-\alpha)\mu_0$, which satisfies the constraint. Suppose that the sample-path reward constraint holds at timestep $t-1$. At time step t, if GenCB-CMAB plays x_0, it is clear that the constraint still holds for t. If GenCB-CMAB plays a regular arm x_t, which implies $r_S(t-1)+N_0(t-1)\mu_0\geq(1-\alpha)\mu_0t$, the received cumulative reward is $r_S(t-1)+N_0(t-1)\mu_0+r_{t,x_t}\geq r_S(t-1)+N_0(t-1)\mu_0\geq(1-\alpha)\mu_0t$, and thus the constraint still holds for t.

Algorithm 3: GenCB-CMAB

Input: Reugular arms $[K]$, default arm x_0 with reward μ_0, parameter α.

1 $\forall t\geq0,\forall0\leq i\leq K,N_i(t)\leftarrow0.\ \forall t\geq0,r_S(t)\leftarrow0.$ $m\leftarrow0$;
2 **for** $t=1,2,\dots$ **do**
3 **if** $r_S(t-1)+N_0(t-1)\mu_0\geq(1-\alpha)\mu_0t$ **then**
4 $m\leftarrow m+1$;
5 $x_t\leftarrow\operatorname{argmax}_{i\in[K]}\left(\hat{\mu}_i+\sqrt{\frac{2\ln m}{N_i(t-1)}}\right)$;
6 Play arm x_t, observe the random reward r_{t,x_t} and update the empirical mean $\hat{\mu}_{x_t}$;
7 $N_{x_t}(t)\leftarrow N_{x_t}(t-1)+1$ and $\forall0\leq i\leq K,i\neq x_t,N_i(t)\leftarrow N_i(t-1)$;
8 $r_S(t)\leftarrow r_S(t-1)+r_{t,x_t}$;
9 **else**
10 Play x_0 and receive reward μ_0;
11 $N_0(t)\leftarrow N_0(t-1)+1$;
12 $r_S(t)\leftarrow r_S(t-1)$;

Recall that m_t denotes the number of times we played regular arms up to t, and $\mathcal{S}_t$ denotes the set of timesteps when we played regular arms up to t.

Fix a time horizon T. Let $\tau\leq T$ denote the last timestep when GenCB-CMAB played arm x_0. For timestep $1\leq n\leq\tau$, define event

$$\mathcal{E}_n:=\left\{\sum_{t=1}^n r_{t,x_t}\geq\sum_{t=1}^n\mu_{x_t}-2\sqrt{n\ln(\tau)}\right\}.$$

According to the Azuma-Hoeffding inequality, we have

$$\Pr[\bar{\mathcal{E}}_n]\leq\frac{1}{\tau^2}.$$

We further define event

$$\mathcal{E}:=\bigcap_{n=1}^{\tau}\mathcal{E}_n.$$

By a union bound over $1\leq n\leq\tau$, we have

$$\begin{aligned}\Pr[\bar{\mathcal{E}}]\leq&\sum_{n=1}^{\tau}\frac{1}{\tau^2}\\\leq&\tau\cdot\frac{1}{\tau^2}\\\leq&\frac{1}{\tau}.\end{aligned}$$

Thus, $\Pr[\mathcal{E}]>1-\frac{1}{\tau}$. In other words, with probability at least $1-\frac{1}{\tau}$, we have that for any $1\leq n\leq\tau$, $\sum_{t=1}^n r_{t,x_t}\geq\sum_{t=1}^n\mu_{x_t}-2\sqrt{n\ln(\tau)}$.

At the timestep τ, we have the following three equivalent inequalities:

$$\sum_{t\in\mathcal{S}_{\tau-1}} r_t + N_0(\tau-1)\mu_0 < (1-\alpha)\mu_0\tau$$

$$\sum_{t\in\mathcal{S}_{\tau-1}} r_t + N_0(\tau-1)\mu_0 < (1-\alpha)\mu_0(N_0(\tau-1)+m_{\tau-1}+1)$$

$$\alpha\mu_0 N_0(\tau-1) < (1-\alpha)\mu_0(m_{\tau-1}+1) - \sum_{t\in\mathcal{S}_{\tau-1}} r_t$$

In the standard multi-armed bandit problem, we define the pseudo-regret of the well-known UCB (Auer, Cesa-Bianchi, and Fischer 2002) algorithm for any time m as

$$\tilde{\mathcal{R}}_m(\mathsf{UCB}) = \mu_* m - \sum_{t=1}^{m} \mu_{x_t}.$$

Thus, we have

$$\sum_{t=1}^{m} \mu_{x_t} = \mu_* m - \tilde{\mathcal{R}}_m(\mathsf{UCB}).$$

Then, we have

$$\begin{aligned}
\alpha\mu_0 N_0(\tau-1) <& (1-\alpha)\mu_0(m_{\tau-1}+1) - \sum_{t\in\mathcal{S}_{\tau-1}} r_t \\
&- \tilde{\mathcal{R}}_{m_{\tau-1}}(\mathsf{UCB}) + \tilde{\mathcal{R}}_{m_{\tau-1}}(\mathsf{UCB}) \\
=& (1-\alpha)\mu_0(m_{\tau-1}+1) - \sum_{t\in\mathcal{S}_{\tau-1}} r_t - \mu_* m_{\tau-1} \\
&+ \sum_{t\in\mathcal{S}_{\tau-1}} \mu_{x_t} + \tilde{\mathcal{R}}_{m_{\tau-1}}(\mathsf{UCB}) \\
=& -(\mu_* - (1-\alpha)\mu_0)(m_{\tau-1}+1) + \sum_{t\in\mathcal{S}_{\tau-1}} \mu_{x_t} \\
&- \sum_{t\in\mathcal{S}_{\tau-1}} r_t + \tilde{\mathcal{R}}_{m_{\tau-1}}(\mathsf{UCB}) + \mu_* \\
=& -(\Delta_0 + \alpha\mu_0)(m_{\tau-1}+1) + \sum_{t\in\mathcal{S}_{\tau-1}} \mu_{x_t} \\
&- \sum_{t\in\mathcal{S}_{\tau-1}} r_t + \tilde{\mathcal{R}}_{m_{\tau-1}}(\mathsf{UCB}) + \mu_* \qquad (6)
\end{aligned}$$

In the following analysis, we assume $m_{\tau-1} \geq 1, N_0(\tau-1) \geq 2$, since otherwise the theorem trivially holds. Since $\tau = m_{\tau-1} + N_0(\tau-1) + 1 > 2$, we have $\mathbb{E}[m_{\tau-1}|\mathcal{E}] \leq \frac{\mathbb{E}[m_{\tau-1}]}{\Pr[\mathcal{E}]} \leq \frac{\mathbb{E}[m_{\tau-1}]}{1-\frac{1}{\tau}} < 2\mathbb{E}[m_{\tau-1}]$ and $\mathbb{E}[N_0(\tau-1)|\mathcal{E}] \leq \frac{\mathbb{E}[N_0(\tau-1)]}{\Pr[\mathcal{E}]} \leq \frac{\mathbb{E}[N_0(\tau-1)]}{1-\frac{1}{\tau}} < 2\mathbb{E}[N_0(\tau-1)]$. Taking expectation of both sides in (6), from $\mathbb{E}[\tilde{\mathcal{R}}_{m_{\tau-1}}(\mathsf{UCB})] \leq \mathbb{E}[8H\ln(m_{\tau-1}+1)+5K]$ and Jensen's inequality, we have

$$\begin{aligned}
\alpha\mu_0\mathbb{E}[N_0(\tau-1)] <& -(\Delta_0+\alpha\mu_0)\mathbb{E}[m_{\tau-1}+1] \\
&+ \mathbb{E}\left[\sum_{t\in\mathcal{S}_{\tau-1}} \mu_{x_t} - \sum_{t\in\mathcal{S}_{\tau-1}} r_t|\mathcal{E}\right]\Pr[\mathcal{E}] \\
&+ \mathbb{E}\left[\sum_{t\in\mathcal{S}_{\tau-1}} \mu_{x_t} - \sum_{t\in\mathcal{S}_{\tau-1}} r_t|\bar{\mathcal{E}}\right]\Pr[\bar{\mathcal{E}}] \\
&+ \mathbb{E}[\tilde{\mathcal{R}}_{m_{\tau-1}}(\mathsf{UCB})] + \mu_* \\
\leq& -(\Delta_0+\alpha\mu_0)\mathbb{E}[m_{\tau-1}+1] \\
&+ \mathbb{E}\left[2\sqrt{m_{\tau-1}\ln(\tau)}|\mathcal{E}\right] + \tau\cdot\frac{1}{\tau} \\
&+ \tilde{\mathcal{R}}(\mathbb{E}[m_{\tau-1}]) + \mu_* \\
<& -(\Delta_0+\alpha\mu_0)\mathbb{E}[m_{\tau-1}+1] \\
&+ 2\sqrt{2\mathbb{E}[m_{\tau-1}+1]}\ln(2\mathbb{E}[N_0(\tau-1)]\cdot \\
&\mathbb{E}[m_{\tau-1}+1]) + 8H\ln(\mathbb{E}[m_{\tau-1}+1]) \\
&+ 5K + 2 \qquad (7)
\end{aligned}$$

Let $m = \mathbb{E}[m_{\tau-1}+1] \geq 2$, $c_1 = 2\sqrt{2}$, $c_2 = 2\mathbb{E}[N_0(\tau-1)] \geq 2$, $c_3 = \Delta_0 + \alpha\mu_0 \in (0,1)$, $c_4 = 8H$ where $H \geq 2$. The RHS of Eq. (7) can be written as a constant term plus

$$g_1(m) = -c_3 m + c_1\sqrt{m}\ln(c_2 m) + c_4\ln(m).$$

According to Lemma 1, we have

$$g_1(m) \leq \frac{132 c_1 c_4}{c_3}\left[\ln\left(\frac{10\sqrt{c_2 c_4}}{c_3}\right)\right]^2.$$

Then, we have

$$\begin{aligned}
&\alpha\mu_0\mathbb{E}[N_0(\tau-1)] \\
<& \frac{2112\sqrt{2}H}{\Delta_0+\alpha\mu_0}\left[\ln\left(\frac{40\sqrt{H\mathbb{E}[N_0(\tau-1)]}}{\Delta_0+\alpha\mu_0}\right)\right]^2 + 5K + 2 \\
<& \frac{2994H}{\Delta_0+\alpha\mu_0}\left[\ln\left(\frac{40\sqrt{H\mathbb{E}[N_0(\tau-1)]}}{\Delta_0+\alpha\mu_0}\right)\right]^2.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
\mathbb{E}[N_0(\tau-1)] <& \frac{2994H}{\alpha\mu_0(\Delta_0+\alpha\mu_0)}\cdot \\
&\left[\ln\left(\frac{40\sqrt{H\mathbb{E}[N_0(\tau-1)]}}{\Delta_0+\alpha\mu_0}\right)\right]^2 \\
\sqrt{\mathbb{E}[N_0(\tau-1)]} <& \sqrt{\frac{2994H}{\alpha\mu_0(\Delta_0+\alpha\mu_0)}}\cdot \\
&\ln\left(\frac{40\sqrt{H\mathbb{E}[N_0(\tau-1)]}}{\Delta_0+\alpha\mu_0}\right)
\end{aligned}$$

According to Fact 2 (set $z = \sqrt{\mathbb{E}[N_0(\tau-1)]}, c_1 = \sqrt{\frac{2994H}{\alpha\mu_0(\Delta_0+\alpha\mu_0)}}, c_2 = \frac{40\sqrt{H}}{\Delta_0+\alpha\mu_0}$),

$$\begin{aligned}
\sqrt{\mathbb{E}[N_0(\tau-1)]} \leq& 2\sqrt{\frac{2994H}{\alpha\mu_0(\Delta_0+\alpha\mu_0)}}\cdot \\
&\ln\left(\sqrt{\frac{2994H}{\alpha\mu_0(\Delta_0+\alpha\mu_0)}}\cdot\frac{40\sqrt{H}}{\Delta_0+\alpha\mu_0}\right)
\end{aligned}$$

Algorithm 4: GenCB-CLB

Input: Reugular arms $\mathcal{X} \subseteq \mathbb{R}^d$, default arm x_0 with reward μ_0, parameter α, L, S, $\lambda \geq \max\{1, L^2\}$.

$\forall t \geq 0, N_0(t) \leftarrow 0, r_S(t) \leftarrow 0$. $m \leftarrow 0$. $V_0 \leftarrow \lambda I$. $b_0 \leftarrow \mathbf{0}^d$;

for $t = 1, 2, \ldots$ **do**

  if $r_S(t-1) + N_0(t-1)\mu_0 \geq (1-\alpha)\mu_0 t$ **then**

    $m \leftarrow m + 1$;

    $\mathcal{C}_t \leftarrow \{\theta \in \mathbb{R}^d : \|\theta - \hat{\theta}_{t-1}\|_{V_{t-1}} \leq \sqrt{d\ln(2m^2(1+mL^2/\lambda))} + \sqrt{\lambda}S\}$;

    $(x_t, \tilde{\theta}_t) \leftarrow \operatorname{argmax}_{(x,\theta)\in\mathcal{X}\times\mathcal{C}_t} x^\top\theta$;

    Play arm x_t and observe the random reward r_{t,x_t};

    $V_t \leftarrow V_{t-1} + x_t x_t^\top$, $b_t \leftarrow b_{t-1} + r_{t,x_t}x_t$;

    $\hat{\theta}_t \leftarrow V_t^{-1} b_t$;

    $r_S(t) \leftarrow r_S(t-1) + r_{t,x_t}$;

  else

    Play x_0 and receive reward μ_0;

    $N_0(t) \leftarrow N_0(t-1) + 1$;

    $r_S(t) \leftarrow r_S(t-1)$;

$$\mathbb{E}[N_0(\tau-1)] \leq \frac{11976H}{\alpha\mu_0(\Delta_0+\alpha\mu_0)} \cdot \left[\ln\left(\frac{2189H}{(\Delta_0+\alpha\mu_0)\sqrt{\alpha\mu_0(\Delta_0+\alpha\mu_0)}}\right)\right]^2$$

Thus,

$$\begin{aligned}\mathbb{E}[N_0(T)] =& \mathbb{E}[N_0(\tau)] \\ =& \mathbb{E}[N_0(\tau-1)] + 1 \\ =& O\left(\frac{H}{\alpha\mu_0(\Delta_0+\alpha\mu_0)}\left[\ln\left(\frac{H}{\alpha\mu_0(\Delta_0+\alpha\mu_0)}\right)\right]^2\right).\end{aligned}$$

Theorem 2 follows from $\mathbb{E}[\mathcal{R}_T(\text{GenCB-CMAB})] \leq \mathbb{E}[\mathcal{R}_T(\text{UCB})] + \mathbb{E}[N_0(T)]\Delta_0$. □

11 Algorithm Pseudo-code and Proof for CLB

Algorithm 4 presents the algorithm pseudo-code of GenCB-CLB for CLB, and we give the detailed proof of Theorem 3 in the following.

Proof. Since the proof of satisfaction on the performance constraint Eq. (2) is the same to Theorem 2, we mainly give the proof of regret bound here.

For timestep $1 \leq n \leq \tau$, define event

$$\mathcal{E}_n := \left\{\sum_{t=1}^n r_{x_t} \geq \sum_{t=1}^n \mu_{x_t} - 2\sqrt{n\ln(\tau)}\right\}.$$

According to the Azuma-Hoeffding inequality, we have

$$\Pr[\bar{\mathcal{E}}_n] \leq \frac{1}{\tau^2}.$$

We further define event

$$\mathcal{E} := \bigcap_{n=1}^{\tau} \mathcal{E}_n.$$

By a union bound over $1 \leq n \leq \tau$, we have

$$\begin{aligned}\Pr[\bar{\mathcal{E}}] \leq& \sum_{n=1}^{\tau} \frac{1}{\tau^2} \\ \leq& \tau \cdot \frac{1}{\tau^2} \\ \leq& \frac{1}{\tau}.\end{aligned}$$

Thus, $\Pr[\mathcal{E}] > 1 - \frac{1}{\tau}$. In other words, with probability at least $1 - \frac{1}{\tau}$, we have that for any $1 \leq n \leq \tau$, $\sum_{t=1}^n r_{x_t} \geq \sum_{t=1}^n x_t^\top\theta^* - 2\sqrt{n\ln(\tau)}$.

For the confidence ellipsoid in the LinUCB (Abbasi-yadkori, Pál, and Szepesvári 2011) algorithm, we set the confidence parameter $\delta_t = 1/(2t^2)$ and the confidence ellipsoid for timestep t is

$$\mathcal{C}_t = \{\theta \in \mathbb{R}^d : \|\theta - \hat{\theta}_{t-1}\|_{V_{t-1}} \leq \sqrt{d\ln(2m_t^2(1+m_tL^2/\lambda))} + \sqrt{\lambda}S\}.$$

Then, we can obtain

$$\begin{aligned}&\mathbb{E}[\tilde{\mathcal{R}}_{m_{\tau-1}}(\text{LinUCB})] \\ =& \mathbb{E}\left[\mu_* m_{\tau-1} - \sum_{t=1}^{m_{\tau-1}} \mu_{x_t}\right] \\ \leq& \mathbb{E}\left[4\sqrt{m_{\tau-1}d\ln\left(1+\frac{m_{\tau-1}L^2}{\lambda d}\right)} \cdot \left(\sqrt{\lambda}S + \sqrt{d\ln\left(2m_{\tau-1}^2 \cdot \left(1+\frac{m_{\tau-1}L^2}{\lambda}\right)\right)}\right) + \Delta_{\max}\right].\end{aligned}$$

Fix time horizon T. Recall that $\tau \leq T$ is the last timestep when we played x_0. Similar to the analysis in CMAB (Eq. (6)), at timestep τ, we have

$$\begin{aligned}\alpha\mu_0 N_0(\tau-1) <& -(\Delta_0+\alpha\mu_0)(m_{\tau-1}+1) + \sum_{t\in\mathcal{S}_{\tau-1}} \mu_{x_t} \\ & - \sum_{t\in\mathcal{S}_{\tau-1}} r_t + \tilde{\mathcal{R}}_{m_{\tau-1}}(\text{LinUCB}) + \mu_*.\end{aligned}$$

Taking expectation of both sides, we have

$$\begin{aligned}\alpha\mu_0\mathbb{E}[N_0(\tau-1)] <& -(\Delta_0+\alpha\mu_0)\mathbb{E}[m_{\tau-1}+1] \\ & + \mathbb{E}\left[\sum_{t\in\mathcal{S}_{\tau-1}} \mu_{x_t} - \sum_{t\in\mathcal{S}_{\tau-1}} r_t | \mathcal{E}\right] \Pr[\mathcal{E}] \\ & + \mathbb{E}\left[\sum_{t\in\mathcal{S}_{\tau-1}} \mu_{x_t} - \sum_{t\in\mathcal{S}_{\tau-1}} r_t | \bar{\mathcal{E}}\right] \Pr[\bar{\mathcal{E}}] \\ & + \mathbb{E}[\tilde{\mathcal{R}}_{m_{\tau-1}}(\text{LinUCB})] + \mu_*\end{aligned}$$

$$
\begin{aligned}
&\leq -(\Delta_0+\alpha\mu_0)\mathbb{E}[m_{\tau-1}+1] \\
&\quad + \mathbb{E}\left[2\sqrt{m_{\tau-1}\ln(\tau)}|\mathcal{E}\right] + \tau\cdot\frac{1}{\tau} \\
&\quad + \mathbb{E}[\tilde{\mathcal{R}}_{m_{\tau-1}}(\mathsf{LinUCB})] + \mu_* \\
&< -(\Delta_0+\alpha\mu_0)\mathbb{E}[m_{\tau-1}+1] \\
&\quad + 2\sqrt{2\mathbb{E}[m_{\tau-1}]\ln(2\mathbb{E}[N_0(\tau-1)]\mathbb{E}[m_{\tau-1}+1])} \\
&\quad + 4\sqrt{\mathbb{E}[m_{\tau-1}]d\ln\left(1+\frac{\mathbb{E}[m_{\tau-1}]L^2}{\lambda d}\right)}\cdot \\
&\left(\sqrt{\lambda}S+\sqrt{d\ln\left(2\mathbb{E}[m_{\tau-1}]^2\cdot\left(1+\frac{\mathbb{E}[m_{\tau-1}]L^2}{\lambda}\right)\right)}\right) \\
&\quad + \Delta_{\max}+2 \\
&< -(\Delta_0+\alpha\mu_0)\mathbb{E}[m_{\tau-1}+1] \\
&\quad + 19d\sqrt{\lambda}S\sqrt{(\mathbb{E}[m_{\tau-1}+1])}\cdot \\
&\quad \ln(2\mathbb{E}[N_0(\tau-1)]\cdot\mathbb{E}[m_{\tau-1}+1])+3,
\end{aligned}
$$

where $\Delta_{\max}=\max_{x\in\mathcal{X}}\Delta_x$.

Let $m=\mathbb{E}[m_{\tau-1}+1]$, $c_1=19d\sqrt{\lambda}S$, $c_2=2\mathbb{E}[N_0(\tau-1)]$, $c_3=\Delta_0+\alpha\mu_0$. According to Fact 1, we have

$$
\begin{aligned}
\alpha\mu_0\mathbb{E}[N_0(\tau-1)] <& \frac{5776d^2S^2\lambda}{9(\Delta_0+\alpha\mu_0)}\cdot \\
&\left[\ln\left(\frac{147d\sqrt{\lambda}S\sqrt{\mathbb{E}[N_0(\tau-1)]}}{\Delta_0+\alpha\mu_0}\right)\right]^2+3.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
\mathbb{E}[N_0(\tau-1)] <& \frac{645d^2S^2\lambda}{\alpha\mu_0(\Delta_0+\alpha\mu_0)}\cdot \\
&\left[\ln\left(\frac{147d\sqrt{\lambda}S\sqrt{\mathbb{E}[N_0(\tau-1)]}}{\Delta_0+\alpha\mu_0}\right)\right]^2 \\
\sqrt{\mathbb{E}[N_0(\tau-1)]} <& \frac{26d\sqrt{\lambda}S}{\sqrt{\alpha\mu_0(\Delta_0+\alpha\mu_0)}}\cdot \\
&\ln\left(\frac{147d\sqrt{\lambda}S\sqrt{\mathbb{E}[N_0(\tau-1)]}}{\Delta_0+\alpha\mu_0}\right)
\end{aligned}
$$

According to Fact 2 (set $z=\sqrt{\mathbb{E}[N_0(\tau-1)]}, c_1=\frac{26d\sqrt{\lambda}S}{\sqrt{\alpha\mu_0(\Delta_0+\alpha\mu_0)}}, c_2=\frac{147d\sqrt{\lambda}S}{\Delta_0+\alpha\mu_0}$),

$$
\begin{aligned}
\sqrt{\mathbb{E}[N_0(\tau-1)]} \leq& \frac{52d\sqrt{\lambda}S}{\sqrt{\alpha\mu_0(\Delta_0+\alpha\mu_0)}}\cdot \\
&\ln\left(\frac{3822d^2S^2\lambda}{(\Delta_0+\alpha\mu_0)^{\frac{3}{2}}\sqrt{\alpha\mu_0}}\right) \\
\mathbb{E}[N_0(\tau-1)] \leq& \frac{2704d^2S^2\lambda}{\alpha\mu_0(\Delta_0+\alpha\mu_0)}\cdot \\
&\left[\ln\left(\frac{3822d^2S^2\lambda}{(\Delta_0+\alpha\mu_0)^{\frac{3}{2}}\sqrt{\alpha\mu_0}}\right)\right]^2
\end{aligned}
$$

Algorithm 5: GenCB-CCCB

Input: Reugular arms (decision class) $\mathcal{X}$, base arms $x_1,\ldots,x_K\in\mathbb{R}^d$, default arm x_0 with reward μ_0, parameter α, L, S, $\lambda\geq\max\{1,L^2\}$.

$\forall t\geq 0, N_0(t)\leftarrow 0, r_S(t)\leftarrow 0$. $m\leftarrow 0$. $V_0\leftarrow\lambda I$, $b_0\leftarrow\mathbf{0}^d$;

for $t=1,2,\ldots$ **do**

  if $r_S(t-1)+N_0(t-1)\mu_0\geq(1-\alpha)\mu_0 t$ **then**

    $m\leftarrow m+1$;

    $\hat{w}_{t,e}\leftarrow x_e^\top\hat{\theta}_{t-1}, \forall e\in[K]$;

    $\bar{w}_{t,e}\leftarrow\hat{w}_{t,e}+(\sqrt{d\ln(2m^2(1+mKL^2/\lambda))}$

      $+\sqrt{\lambda}S)\|x_e\|_{V_{t-1}^{-1}}, \forall e\in[K]$;

    $A_t\leftarrow\operatorname{argmax}_{A\in\mathcal{X}}\bar{f}(A,\bar{\boldsymbol{w}})$;

    Play arm A_t and observe the random reward $w_{t,e}$ for all $e\in A_t$;

    $V_t\leftarrow\lambda I+\sum_{s=1}^t\sum_{e\in A_s}x_e x_e^\top$;

    $b_t\leftarrow\sum_{s=1}^t\sum_{e\in A_s}w_{s,e}x_e$;

    $\hat{\theta}_t\leftarrow V_t^{-1}b_t$;

    $r_S(t)\leftarrow r_S(t-1)+r_{t,A_t}$;

  else

    Play x_0 and receive reward μ_0;

    $N_0(t)\leftarrow N_0(t-1)+1$;

    $r_S(t)\leftarrow r_S(t-1)$;

Thus,

$$
\begin{aligned}
&\mathbb{E}[N_0(T)] \\
=&\mathbb{E}[N_0(\tau)] \\
=&\mathbb{E}[N_0(\tau-1)]+1 \\
=&O\left(\frac{d^2S^2\lambda}{\alpha\mu_0(\Delta_0+\alpha\mu_0)}\left[\ln\left(\frac{dS\sqrt{\lambda}}{\alpha\mu_0(\Delta_0+\alpha\mu_0)}\right)\right]^2\right).
\end{aligned}
$$

Theorem 3 follows from $\mathbb{E}[\mathcal{R}_T(\mathsf{GenCB\text{-}CLB})]\leq\mathbb{E}[\mathcal{R}_T(\mathsf{LinUCB})]+\mathbb{E}[N_0(T)]\Delta_0$. □

12 Algorithm Pseudo-code and Proof for CCCB

Algorithm 5 presents the algorithm pseudo-code of GenCB-CCCB for CCCB, and we give the detailed proof of Theorem 4 in the following.

Proof. Since the proof of satisfaction on the performance constraint Eq. (2) is the same to Theorem 2, we mainly give the proof of regret bound here.

For timestep $1\leq n\leq\tau$, define event

$$
\mathcal{E}_n:=\left\{\sum_{t=1}^n r_{t,A_t}\geq\sum_{t=1}^n f(A_t,\boldsymbol{w}^*)-2K\sqrt{n\ln(\tau)}\right\}.
$$

According to the Azuma-Hoeffding inequality, we have

$$
\Pr[\bar{\mathcal{E}}_n]\leq\frac{1}{\tau^2}.
$$

We further define event

$$\mathcal{E} := \bigcap_{n=1}^{\tau} \mathcal{E}_n.$$

By a union bound over $1 \leq n \leq T$, we have

$$\begin{aligned}\Pr[\bar{\mathcal{E}}] \leq & \sum_{n=1}^{\tau} \frac{1}{\tau^2} \\ \leq & \tau \cdot \frac{1}{\tau^2} \\ \leq & \frac{1}{\tau}.\end{aligned}$$

Thus, $\Pr[\bar{\mathcal{E}}] > 1 - \frac{1}{\tau}$. In other words, with probability at least $1 - \frac{1}{\tau}$, we have that for any $1 \leq n \leq \tau$, $\sum_{t=1}^{n} r_{t,A_t} \geq \sum_{t=1}^{n} f(A_t, \boldsymbol{w}^*) - 2K\sqrt{n \ln(\tau)}$.

For the confidence ellipsoid in the C2UCB (Qin, Chen, and Zhu 2014) algorithm, we set the confidence parameter $\delta_t = 1/(2t^2)$ and the confidence ellipsoid for timestep t is

$$\begin{aligned}\mathcal{C}_t = \big\{\theta \in \mathbb{R}^d : & \|\theta - \hat{\theta}_{t-1}\|_{V_{t-1}} \leq \\ & \sqrt{d \ln(2m^2(1 + mKL^2/\lambda))} + \sqrt{\lambda} S\big\}.\end{aligned}$$

Then, we can obtain

$$\begin{aligned}&\mathbb{E}[\mathcal{R}_{m_{\tau-1}}(\textsf{C2UCB})] \\ \leq & \mathbb{E}\Bigg[2P\sqrt{2dm_{\tau-1}\ln\left(1 + \frac{m_{\tau-1}KL^2}{\lambda d}\right)} \cdot \\ & \left(\sqrt{\lambda}S + \sqrt{d\ln\left(2m_{\tau-1}^2 \cdot \left(1 + \frac{m_{\tau-1}KL^2}{\lambda}\right)\right)}\right) + \Delta_{\max}\Bigg],\end{aligned}$$

where $\Delta_{\max} = \max_{A \in \mathcal{X}}(f(A_*, \boldsymbol{w}^*) - f(A, \boldsymbol{w}^*))$.

Fix time horizon T. Recall that $\tau \leq T$ is the last timestep when we played x_0. Similar to the conservative multi-armed bandit case (Eq. (6)), at timestep τ, we have

$$\begin{aligned}\alpha\mu_0 N_0(\tau - 1) < & -(\Delta_0 + \alpha\mu_0)(m_{\tau-1} + 1) \\ & + \sum_{t \in S_{\tau-1}} f(A_t, \boldsymbol{w}^*) - \sum_{t \in S_{\tau-1}} r_{t,A_t} \\ & + \tilde{\mathcal{R}}(m_{\tau-1}) + \mu_*.\end{aligned}$$

Taking expectation of both sides, we have

$$\begin{aligned}&\alpha\mu_0\mathbb{E}[N_0(\tau - 1)] \\ < & -(\Delta_0 + \alpha\mu_0)\mathbb{E}[m_{\tau-1} + 1] \\ & + \mathbb{E}\left[\sum_{t \in S_{\tau-1}} f(A_t, \boldsymbol{w}^*) - \sum_{t \in S_{\tau-1}} r_{t,A_t} | \mathcal{E}\right] \Pr[\mathcal{E}] \\ & + \mathbb{E}\left[\sum_{t \in S_{\tau-1}} f(A_t, \boldsymbol{w}^*) - \sum_{t \in S_{\tau-1}} r_{t,A_t} | \bar{\mathcal{E}}\right] \Pr[\bar{\mathcal{E}}] \\ & + \mathbb{E}[\tilde{\mathcal{R}}(m_{\tau-1})] + \mu_* \\ \leq & -(\Delta_0 + \alpha\mu_0)\mathbb{E}[m_{\tau-1} + 1] + \mathbb{E}\left[2K\sqrt{m_{\tau-1}\ln(\tau)} | \mathcal{E}\right] \\ & + \tau \cdot \frac{1}{\tau} + \mathbb{E}[\tilde{\mathcal{R}}(m_{\tau-1})] + \mu_* \\ < & -(\Delta_0 + \alpha\mu_0)\mathbb{E}[m_{\tau-1} + 1] \\ & + 2K\sqrt{2\mathbb{E}[m_{\tau-1}]\ln(2\mathbb{E}[N_0(\tau-1)]\mathbb{E}[m_{\tau-1}+1])} \\ & + 2P\sqrt{2d\mathbb{E}[m_{\tau-1}]\ln\left(1 + \frac{\mathbb{E}[m_{\tau-1}]KL^2}{\lambda d}\right)} \cdot \\ & \left(\sqrt{\lambda}S + \sqrt{d\ln\left(2\mathbb{E}[m_{\tau-1}]^2 \cdot \left(1 + \frac{\mathbb{E}[m_{\tau-1}]KL^2}{\lambda}\right)\right)}\right) \\ & + \Delta_{\max} + 2 \\ < & -(\Delta_0 + \alpha\mu_0)\mathbb{E}[m_{\tau-1} + 1] + (2\sqrt{2}K + 10P\sqrt{\lambda}Sd) \cdot \\ & \sqrt{(\mathbb{E}[m_{\tau-1} + 1])\ln(2K\mathbb{E}[N_0(\tau-1)] \cdot \mathbb{E}[m_{\tau-1} + 1])} + 3.\end{aligned}$$

Let $m = \mathbb{E}[m_{\tau-1} + 1]$, $c_1 = 2\sqrt{2}K + 10P\sqrt{\lambda}Sd$, $c_2 = 2K\mathbb{E}[N_0(\tau - 1)]$, $c_3 = \Delta_0 + \alpha\mu_0$. According to Fact 1, we have

$$\begin{aligned}\alpha\mu_0\mathbb{E}[N_0(\tau-1)] < & \frac{16(2\sqrt{2}K + 10P\sqrt{\lambda}Sd)^2}{9(\Delta_0 + \alpha\mu_0)} \cdot \\ & \left[\ln\left(\frac{8(2\sqrt{2}K + 10P\sqrt{\lambda}Sd)\sqrt{K\mathbb{E}[N_0(\tau-1)]}}{\Delta_0 + \alpha\mu_0}\right)\right]^2 + 3.\end{aligned}$$

Thus, we have

$$\begin{aligned}\mathbb{E}[N_0(\tau-1)] < & \frac{5(2\sqrt{2}K + 10P\sqrt{\lambda}Sd)^2}{\alpha\mu_0(\Delta_0 + \alpha\mu_0)} \cdot \\ & \left[\ln\left(\frac{8K(2\sqrt{2}K + 10P\sqrt{\lambda}Sd)\sqrt{\mathbb{E}[N_0(\tau-1)]}}{\Delta_0 + \alpha\mu_0}\right)\right]^2 \\ \sqrt{\mathbb{E}[N_0(\tau-1)]} < & \frac{3(2\sqrt{2}K + 10P\sqrt{\lambda}Sd)}{\sqrt{\alpha\mu_0(\Delta_0 + \alpha\mu_0)}} \cdot \\ & \ln\left(\frac{8K(2\sqrt{2}K + 10P\sqrt{\lambda}Sd)\sqrt{\mathbb{E}[N_0(\tau-1)]}}{\Delta_0 + \alpha\mu_0}\right)\end{aligned}$$

According to Fact 2 (set $z = \sqrt{\mathbb{E}[N_0(\tau-1)]}, c_1 = \frac{3(2\sqrt{2}K + 10P\sqrt{\lambda}Sd)}{\sqrt{\alpha\mu_0(\Delta_0 + \alpha\mu_0)}}, c_2 = \frac{8K(2\sqrt{2}K + 10P\sqrt{\lambda}Sd)}{\Delta_0 + \alpha\mu_0}$),

$$\begin{aligned}\sqrt{\mathbb{E}[N_0(\tau-1)]} \leq & \frac{12(2\sqrt{2}K + 10P\sqrt{\lambda}Sd)}{\sqrt{\alpha\mu_0(\Delta_0 + \alpha\mu_0)}} \cdot \\ & \ln\left(\frac{5K(2\sqrt{2}K + 10P\sqrt{\lambda}Sd)}{(\Delta_0 + \alpha\mu_0)\sqrt{\alpha\mu_0(\Delta_0 + \alpha\mu_0)}}\right) \\ \mathbb{E}[N_0(\tau-1)] \leq & \frac{144(2\sqrt{2}K + 10P\sqrt{\lambda}Sd)^2}{\alpha\mu_0(\Delta_0 + \alpha\mu_0)} \cdot \\ & \left[\ln\left(\frac{5K(2\sqrt{2}K + 10P\sqrt{\lambda}Sd)}{(\Delta_0 + \alpha\mu_0)\sqrt{\alpha\mu_0(\Delta_0 + \alpha\mu_0)}}\right)\right]^2\end{aligned}$$

Thus,

$$\begin{aligned}&\mathbb{E}[N_0(T)] \\ = & \mathbb{E}[N_0(\tau)]\end{aligned}$$

$$
\begin{aligned}
&=\mathbb{E}[N_0(\tau-1)]+1\\
&=O\left(\frac{(K+P\sqrt{\lambda}Sd)^2}{\alpha\mu_0(\Delta_0+\alpha\mu_0)}\left[\ln\left(\frac{K+P\sqrt{\lambda}Sd}{\alpha\mu_0(\Delta_0+\alpha\mu_0)}\right)\right]^2\right).
\end{aligned}
$$

Theorem 4 follows from $\mathbb{E}[\mathcal{R}_T(\textsf{GenCB-CCCB})] \leq \mathbb{E}[\mathcal{R}_T(\textsf{C2UCB})] + \mathbb{E}[N_0(T)]\Delta_0$. □

13 Proof for MV-CBP

We give the detailed proof of Theorem 5 below.

Proof. In order to prove that MV-CUCB satisfies the sample-path reward constraint Eq. (5), we give the following inequalities first. For any time horizon T,

$$
\begin{aligned}
&T\cdot\widehat{\mathrm{MV}}_T(\mathcal{A})\\
=&T\cdot\left(\rho\hat{\mu}_T(\mathcal{A})-\hat{\sigma}_T^2(\mathcal{A})\right)\\
=&T\cdot\frac{\rho}{T}\sum_{t=1}^{T}r_{t,x_t}-T\cdot\frac{1}{T}\sum_{t=1}^{T}r_{t,x_t}^2+T\cdot\left(\frac{\sum_{t=1}^{T}r_{t,x_t}}{T}\right)^2\\
=&\rho\sum_{t=1}^{T-1}r_{t,x_t}+\rho r_{T,x_T}-\sum_{t=1}^{T-1}r_{t,x_t}^2-r_{T,x_T}^2\\
&+\frac{1}{T}\left(\left(\sum_{t=1}^{T-1}r_{t,x_t}\right)^2+r_{T,x_T}^2+2\left(\sum_{t=1}^{T-1}r_{t,x_t}\right)r_{T,x_T}\right)\\
=&\rho\sum_{t=1}^{T-1}r_{t,x_t}-\sum_{t=1}^{T-1}r_{t,x_t}^2+\frac{1}{T-1}\left(\sum_{t=1}^{T-1}r_{t,x_t}\right)^2\\
&-\frac{1}{T-1}\left(\sum_{t=1}^{T-1}r_{t,x_t}\right)^2+\rho r_{T,x_T}-r_{T,x_T}^2\\
&+\frac{1}{T}\left(\left(\sum_{t=1}^{T-1}r_{t,x_t}\right)^2+r_{T,x_T}^2+2\left(\sum_{t=1}^{T-1}r_{t,x_t}\right)r_{T,x_T}\right)\\
\geq&(T-1)\cdot\widehat{\mathrm{MV}}_{T-1}(\mathcal{A})-1-\frac{1}{T(T-1)}\left(\sum_{t=1}^{T-1}r_{t,x_t}\right)^2\\
\geq&(T-1)\cdot\widehat{\mathrm{MV}}_{T-1}(\mathcal{A})-2 \qquad (8)
\end{aligned}
$$

Now we prove that MV-CUCB satisfies the sample-path reward constraint Eq. (5) by induction. At timestep $t=1$, since the LHS of the if statement (in Line 3 of Algorithm 3) is -2 and RHS is positive, MV-CUCB will pull the default arm x_0 and receive reward μ_0. Then, we have $\widehat{\mathrm{MV}}_1(\mathcal{A}) = \mathrm{MV}_0 \geq (1-\alpha)\mathrm{MV}_0$, which satisfies the constraint. Suppose that the sample-path reward constraint holds at timestep $t-1$. At time step t, if MV-CUCB plays x_0, since the exploration risk caused by one pull is bounded by 2 and $\alpha\mathrm{MV}_0 > 2$, the constraint still holds for t. If MV-CUCB plays a regular arm x_t, which implies $(t-1)\widehat{\mathrm{MV}}_{t-1}(\mathcal{A})-2 \geq (1-\alpha)\mathrm{MV}_0 t$, then from Eq. (8) we have $t\widehat{\mathrm{MV}}_t(\mathcal{A}) \geq (t-1)\widehat{\mathrm{MV}}_{t-1}(\mathcal{A})-2 \geq (1-\alpha)\mathrm{MV}_0 t$, and thus the constraint still holds for t.

Next, we prove the regret bound of the MV-CUCB algorithm. Fix a time horizon T. Let $\tau \leq T$ denote the last timestep when the algorithm pulled arm x_0. For ease of notation, we use $N_{i,\tau-1}$ as a shorthand for $N_i(\tau-1)$, $\forall 0 \leq i \leq K$. Define event

$$
\mathcal{F} := \Bigg\{\forall i=1,\ldots,K, \forall 1 \leq n \leq \tau,
$$

$$
|\hat{\mu}_{i,n}-\mu_i| \leq \sqrt{\frac{\ln(6K\tau^2)}{2n}} \text{ and } |\hat{\sigma}_{i,n}^2-\sigma_i^2| \leq 5\sqrt{\frac{\ln(6K\tau^2)}{2n}}\Bigg\}.
$$

Similar to Lemma 1 in (Sani, Lazaric, and Munos 2012), using Chernoff-Hoeffding inequality and a union bound over arms and observations, we have $\Pr[\bar{\mathcal{F}}] \leq \frac{1}{\tau}$. Thus, $\Pr[\mathcal{F}] > 1-\frac{1}{\tau}$.

Conditioning on $\mathcal{F}$, we have

$$
\begin{aligned}
&\sum_{i=1}^{K}N_{i,\tau-1}\mathrm{MV}_i-\frac{2}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\Gamma_{i,j}^2\\
\leq&\sum_{i=1}^{K}N_{i,\tau-1}\left(\widehat{\mathrm{MV}}_i+2(5+\rho)\sqrt{\frac{\log(6K\tau^2)}{2N_{i,\tau-1}}}\right)\\
&-\Bigg(\frac{2}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\Gamma_{i,j}^2\\
&+\frac{2\sqrt{2}}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{j,\tau-1}\log(6K\tau^2)\\
&+\frac{2\sqrt{2}}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}\log(6K\tau^2)\Bigg)\\
&+\frac{2\sqrt{2}}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{j,\tau-1}\log(6K\tau^2)\\
&+\frac{2\sqrt{2}}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}\log(6K\tau^2)\\
\leq&\sum_{i=1}^{K}N_{i,\tau-1}\widehat{\mathrm{MV}}_i+(5+\rho)\sum_{i=1}^{K}\sqrt{2N_{i,\tau-1}\log(6K\tau^2)}\\
&-\frac{1}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\cdot\\
&\left(|\Gamma_{i,j}|+\sqrt{\frac{\log(6K\tau^2)}{2N_{i,\tau-1}}}+\sqrt{\frac{\log(6K\tau^2)}{2N_{j,\tau-1}}}\right)^2\\
&+4\sqrt{2}K\log(6K\tau^2)\\
\leq&\sum_{i=1}^{K}N_{i,\tau-1}\widehat{\mathrm{MV}}_i-\frac{1}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\hat{\Gamma}_{i,j}^2
\end{aligned}
$$

$$+ (5+\rho)\sqrt{2Km_{\tau-1}N_{i,\tau-1}\log(6K\tau^2)}$$
$$+ 4\sqrt{2}K\log(6K\tau^2)$$

Let $L = 2, \text{MV}_* \le \rho, \Delta_{\max}^{\text{MV}} \le \frac{1}{4}+\rho$ and $\text{GAP}_{\max} \le \frac{5}{4}+\rho$. We set the confidence parameter $\delta_t = 1/(12Kt^3)$ in the MV-UCB (Sani, Lazaric, and Munos 2012) algorithm and use Jensen's inequality, and then we have

$$\begin{aligned}
&\mathbb{E}[m_{\tau-1}\tilde{\mathcal{R}}_{m_{\tau-1}}(\text{MV-UCB})]\\
\le&\mathbb{E}\Bigg[12(5+\rho)^2(H_1^{\text{MV}}+4H_2^{\text{MV}})\ln(6Km_{\tau-1})\\
&+288(5+\rho)^4H_3^{\text{MV}}\frac{\ln^2(6Km_{\tau-1})}{m_{\tau-1}}+9K+K\Delta_{\max}^{\text{MV}}\Bigg]\\
\le&12(5+\rho)^2(H_1^{\text{MV}}+4H_2^{\text{MV}})\ln(6K\mathbb{E}[m_{\tau-1}])\\
&+288(5+\rho)^4H_3^{\text{MV}}\frac{\ln^2(6K\mathbb{E}[m_{\tau-1}])}{\mathbb{E}[m_{\tau-1}]}+9K+K\Delta_{\max}^{\text{MV}}.
\end{aligned}$$

At timestep τ, we have

$$(\tau-1)\widehat{\text{MV}}_{\tau-1}(\mathcal{A}) - L < (1-\alpha)\text{MV}_0\tau.$$

Thus,

$$\begin{aligned}
&\sum_{i=0}^{K}N_{i,\tau-1}\widehat{\text{MV}}_i-\frac{1}{\tau-1}\sum_{i=0}^{K}\sum_{j\neq i}N_{i,\tau-1}N_{j,\tau-1}\hat{\Gamma}_{i,j}^2-L\\
&<(1-\alpha)\text{MV}_0(m_{\tau-1}+N_0(\tau-1)+1)
\end{aligned}$$

Rearranging the terms, we have

$$\begin{aligned}
&\alpha\text{MV}_0N_0(\tau-1)\\
\le&(1-\alpha)\text{MV}_0(m_{\tau-1}+1)\\
&-\left(\sum_{i=1}^{K}N_{i,\tau-1}\widehat{\text{MV}}_i-\frac{1}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\hat{\Gamma}_{i,j}^2\right)\\
&+\frac{2}{\tau-1}N_0(\tau-1)\sum_{i=1}^{K}N_{i,\tau-1}\hat{\Gamma}_{0,i}^2+L\\
\le&(1-\alpha)\text{MV}_0(m_{\tau-1}+1)\\
&-\left(\sum_{i=1}^{K}N_{i,\tau-1}\widehat{\text{MV}}_i-\frac{1}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\hat{\Gamma}_{i,j}^2\right)\\
&-m_{\tau-1}\tilde{\mathcal{R}}_{m_{\tau-1}}(\text{MV-UCB})+m_{\tau-1}\tilde{\mathcal{R}}_{m_{\tau-1}}(\text{MV-UCB})\\
&+2N_0(\tau-1)+L\\
\le&(1-\alpha)\text{MV}_0(m_{\tau-1}+1)\\
&-\left(\sum_{i=1}^{K}N_{i,\tau-1}\widehat{\text{MV}}_i-\frac{1}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\hat{\Gamma}_{i,j}^2\right)\\
&-\text{MV}_*m_{\tau-1}+\sum_{i=1}^{K}N_{i,\tau-1}\text{MV}_i\\
&-\frac{2}{m_{\tau-1}}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\Gamma_{i,j}^2\\
&+m_{\tau-1}\tilde{\mathcal{R}}_{m_{\tau-1}}(\text{MV-UCB})+2N_0(\tau-1)+L\\
\le&-(\text{MV}_*-(1-\alpha)\text{MV}_0)(m_{\tau-1}+1)\\
&-\left(\sum_{i=1}^{K}N_{i,\tau-1}\widehat{\text{MV}}_i-\frac{1}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\hat{\Gamma}_{i,j}^2\right)\\
&+\left(\sum_{i=1}^{K}N_{i,\tau-1}\text{MV}_i-\frac{2}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\Gamma_{i,j}^2\right)\\
&+m_{\tau-1}\tilde{\mathcal{R}}_{m_{\tau-1}}(\text{MV-UCB})+2N_0(\tau-1)+L+\text{MV}_*\\
\le&-(\Delta_0^{\text{MV}}+\alpha\text{MV}_0)(m_{\tau-1}+1)\\
&+\left(\sum_{i=1}^{K}N_{i,\tau-1}\text{MV}_i-\frac{2}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\Gamma_{i,j}^2\right)\\
&-\left(\sum_{i=1}^{K}N_{i,\tau-1}\widehat{\text{MV}}_i-\frac{1}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\hat{\Gamma}_{i,j}^2\right)\\
&+m_{\tau-1}\tilde{\mathcal{R}}_{m_{\tau-1}}(\text{MV-UCB})+2N_0(\tau-1)+L+\text{MV}_*
\end{aligned}$$

Taking expectation of both sides, we have

$$\begin{aligned}
&(\alpha\text{MV}_0-2)\mathbb{E}[N_0(\tau-1)]\\
\le&-(\Delta_0^{\text{MV}}+\alpha\text{MV}_0)\mathbb{E}[m_{\tau-1}+1]\\
&+\mathbb{E}\Bigg[\Bigg(\sum_{i=1}^{K}N_{i,\tau-1}\text{MV}_i\\
&-\frac{2}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\Gamma_{i,j}^2\Bigg)\\
&-\Bigg(\sum_{i=1}^{K}N_{i,\tau-1}\widehat{\text{MV}}_i\\
&-\frac{1}{\tau-1}\sum_{i=1}^{K}\sum_{\substack{j\neq i\\ j\neq 0}}N_{i,\tau-1}N_{j,\tau-1}\hat{\Gamma}_{i,j}^2\Bigg)|\mathcal{F}\Bigg]\Pr[\mathcal{F}]\\
&+\text{GAP}_{\max}\tau\Pr[\bar{\mathcal{F}}]+\mathbb{E}[m_{\tau-1}\tilde{\mathcal{R}}_{m_{\tau-1}}(\text{MV-UCB})]\\
&+L+\text{MV}_*\\
\le&-(\Delta_0^{\text{MV}}+\alpha\text{MV}_0)\mathbb{E}[m_{\tau-1}+1]\\
&+\mathbb{E}[(5+\rho)\sqrt{2Km_{\tau-1}\ln(6K\tau^2)}\\
&+4\sqrt{2}K\ln(6K\tau^2)|\mathcal{F}]\\
&+\text{GAP}_{\max}+\mathbb{E}[m_{\tau-1}\tilde{\mathcal{R}}_{m_{\tau-1}}(\text{MV-UCB})]\\
&+L+\text{MV}_*
\end{aligned}$$

$$
\begin{aligned}
&< -(\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0)\mathbb{E}[m_{\tau-1}+1] \\
&\quad +2(5+\rho)\cdot \\
&\quad \sqrt{2K\mathbb{E}[m_{\tau-1}+1]\ln(6K(\mathbb{E}[N_0(\tau-1)]\mathbb{E}[m_{\tau-1}+1]))} \\
&\quad +8\sqrt{2}K\ln(6K(\mathbb{E}[N_0(\tau-1)]\mathbb{E}[m_{\tau-1}+1])) \\
&\quad +12(5+\rho)^2(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})\ln(6K\mathbb{E}[m_{\tau-1}]) \\
&\quad +288(5+\rho)^4H_3^{\mathrm{MV}}\frac{\ln^2(6K\mathbb{E}[m_{\tau-1}])}{\mathbb{E}[m_{\tau-1}]} \\
&\quad +9K+K\Delta_{\max}^{\mathrm{MV}}+\mathrm{GAP}_{\max}+L+\mathrm{MV}_* \\
&< -(\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0)\mathbb{E}[m_{\tau-1}+1] \\
&\quad +2(5+\rho)\cdot \\
&\quad \sqrt{2K}\sqrt{\mathbb{E}[m_{\tau-1}+1]}\ln(6K\mathbb{E}[N_0(\tau-1)]\mathbb{E}[m_{\tau-1}+1]) \\
&\quad +(8\sqrt{2}K+12(5+\rho)^2(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}}))\cdot \\
&\quad \ln(6K\mathbb{E}[N_0(\tau-1)]\mathbb{E}[m_{\tau-1}])+864(5+\rho)^4KH_3^{\mathrm{MV}} \\
&\quad +(13+3\rho)K
\end{aligned}
$$

Let $m=\mathbb{E}[m_{\tau-1}+1]\geq 2$, $c_1=2(5+\rho)\sqrt{2K}$, $c_2=6K\mathbb{E}[N_0(\tau-1)]$, $c_3=\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0\in(2,\rho)$, $c_4=8\sqrt{2}K+12(5+\rho)^2(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})>8\sqrt{2}K+12(5+\rho)(K-1)+48(K-1)$ where $\rho>\frac{2}{\alpha\mu_0}>2$, $c_4>3c_1$, $c_4>12c_3$. The RHS of the above inequality can be written as a constant term plus

$$g_2(m)=-c_3m+c_1\sqrt{m}\ln(c_2m)+c_4\ln(m).$$

According to Lemma 2, we have

$$g_2(m)\leq\frac{48c_1c_4}{c_3}\left[\ln(\frac{3\sqrt{c_2}c_4}{c_3})\right]^2.$$

Then, we have

$$
\begin{aligned}
&(\alpha\mathrm{MV}_0-2)\mathbb{E}[N_0(\tau-1)] \\
\leq&\frac{48\cdot 2(5+\rho)\sqrt{2K}(8\sqrt{2}K+12(5+\rho)^2(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}}))}{\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0}\cdot \\
&\left[\ln\left(3\sqrt{6K\mathbb{E}[N_0(\tau-1)]}\cdot\right.\right. \\
&\left.\left.\left(8\sqrt{2}K+12(5+\rho)^2(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})\right)\right)\right. \\
&\left.-\ln(\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0)\right]^2+864(5+\rho)^4KH_3^{\mathrm{MV}} \\
&+(13+3\rho)K \\
\leq&\left(\frac{1536(5+\rho)K\sqrt{K}+1630(5+\rho)^3\sqrt{K}(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})}{\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0}\right. \\
&\left.+864(5+\rho)^4KH_3^{\mathrm{MV}}+(13+3\rho)K\right)\cdot\left[\ln\left(3\sqrt{6K}\cdot\right.\right. \\
&\left.\left.\left(8\sqrt{2}K+12(5+\rho)^2(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})\right)\sqrt{\mathbb{E}[N_0(\tau-1)]}\right)\right. \\
&\left.-\ln(\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0)\right]^2.
\end{aligned}
$$

Thus, we have

$$
\begin{aligned}
&\mathbb{E}[N_0(\tau-1)] \\
\leq&\left(\frac{1536(5+\rho)K\sqrt{K}+1630(5+\rho)^3\sqrt{K}(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})}{(\alpha\mathrm{MV}_0-2)(\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0)}\right. \\
&\left.+\frac{864(5+\rho)^4KH_3^{\mathrm{MV}}+(13+3\rho)K}{\alpha\mathrm{MV}_0-2}\right)\cdot \\
&\left[\ln\left(3\sqrt{6K}\left(8\sqrt{2}K+12(5+\rho)^2(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})\right)\cdot\right.\right. \\
&\left.\left.\sqrt{\mathbb{E}[N_0(\tau-1)]}\right)-\ln(\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0)\right]^2, \\
&\sqrt{\mathbb{E}[N_0(\tau-1)]} \\
\leq&\left(\frac{1536(5+\rho)K\sqrt{K}+1630(5+\rho)^3\sqrt{K}(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})}{(\alpha\mathrm{MV}_0-2)(\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0)}\right. \\
&\left.+\frac{864(5+\rho)^4KH_3^{\mathrm{MV}}+(13+3\rho)K}{\alpha\mathrm{MV}_0-2}\right)^{\frac{1}{2}}\cdot \\
&\left[\ln\left(3\sqrt{6K}\left(8\sqrt{2}K+12(5+\rho)^2(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})\right)\cdot\right.\right. \\
&\left.\left.\sqrt{\mathbb{E}[N_0(\tau-1)]}\right)-\ln(\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0)\right].
\end{aligned}
$$

According to Fact 2 with
$z=\sqrt{\mathbb{E}[N_0(\tau-1)]}$,
$c_1=(\frac{1536(5+\rho)K\sqrt{K}+1630(5+\rho)^3\sqrt{K}(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})}{(\alpha\mathrm{MV}_0-2)(\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0)}$
$+\frac{864(5+\rho)^4KH_3^{\mathrm{MV}}+(13+3\rho)K}{\alpha\mathrm{MV}_0-2})^{\frac{1}{2}}$ and
$c_2=\frac{3\sqrt{6K}(8\sqrt{2}K+12(5+\rho)^2(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}}))}{\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0}$,

$$
\begin{aligned}
&\sqrt{\mathbb{E}[N_0(\tau-1)]} \\
\leq&2\left(\frac{1536(5+\rho)K\sqrt{K}+1630(5+\rho)^3\sqrt{K}(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})}{(\alpha\mathrm{MV}_0-2)(\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0)}\right. \\
&\left.+\frac{864(5+\rho)^4KH_3^{\mathrm{MV}}+(13+3\rho)K}{\alpha\mathrm{MV}_0-2}\right)^{\frac{1}{2}}\cdot \\
&\ln\left(\left(\frac{1536(5+\rho)K\sqrt{K}+1630(5+\rho)^3\sqrt{K}(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})}{(\alpha\mathrm{MV}_0-2)(\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0)}\right.\right. \\
&\left.\left.+\frac{864(5+\rho)^4KH_3^{\mathrm{MV}}+(13+3\rho)K}{\alpha\mathrm{MV}_0-2}\right)^{\frac{1}{2}}\cdot\right. \\
&\left.\frac{3\sqrt{6K}(8\sqrt{2}K+12(5+\rho)^2(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}}))}{\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0}\right), \\
&\mathbb{E}[N_0(\tau-1)] \\
\leq&\left(\frac{6144(5+\rho)K\sqrt{K}+6520(5+\rho)^3\sqrt{K}(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})}{(\alpha\mathrm{MV}_0-2)(\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0)}\right.
\end{aligned}
$$

$$+\frac{3456(5+\rho)^4KH_3^{\mathrm{MV}}+4(13+3\rho)K}{\alpha\mathrm{MV}_0-2}\Bigg).$$

$$\left[\ln\left(\left(\frac{1536(5+\rho)K\sqrt{K}+1630(5+\rho)^3\sqrt{K}(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})}{(\alpha\mathrm{MV}_0-2)(\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0)}\right.\right.\right.$$

$$\left.+\frac{864(5+\rho)^4KH_3^{\mathrm{MV}}+(13+3\rho)K}{\alpha\mathrm{MV}_0-2}\right)^{\frac{1}{2}}\cdot$$

$$\left.\left.\left.\frac{3\sqrt{6K}(8\sqrt{2}K+12(5+\rho)^2(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}}))}{\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0}\right)\right]^2.$$

Thus,

$$\begin{aligned}
&\mathbb{E}[N_0(\tau)]\\
=&\mathbb{E}[N_0(\tau-1)]+1\\
=&O\Bigg(\frac{\rho^3\sqrt{K}(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})+(\rho^4KH_3^{\mathrm{MV}}+\rho K)\tilde{\Delta}_0^{\mathrm{MV}}}{(\alpha\mathrm{MV}_0-2)\tilde{\Delta}_0^{\mathrm{MV}}}\cdot\\
&\left[\ln\left(\frac{\rho^3\sqrt{K}(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})+(\rho^4KH_3^{\mathrm{MV}}+\rho K)\tilde{\Delta}_0^{\mathrm{MV}}}{(\alpha\mathrm{MV}_0-2)\tilde{\Delta}_0^{\mathrm{MV}}}\cdot\right.\right.\\
&\left.\left.\frac{K\sqrt{K}+\rho^2\sqrt{K}(H_1^{\mathrm{MV}}+4H_2^{\mathrm{MV}})}{\tilde{\Delta}_0^{\mathrm{MV}}}\right)\right]^2\Bigg),
\end{aligned}$$

where $\tilde{\Delta}_0^{\mathrm{MV}}=\Delta_0^{\mathrm{MV}}+\alpha\mathrm{MV}_0$.

Theorem 5 follows from $\mathbb{E}[\mathcal{R}_T(\textsf{MV-CUCB})] \leq \mathbb{E}[\mathcal{R}_T(\textsf{MV-UCB})]+\frac{\mathbb{E}[N_0(T)]}{T}\Delta_0^{\mathrm{MV}}$. □