Suplement to Feedback-assisted Decentralized Dynamic Spectrum Access

APPENDIX

PROOF OF THEOREM 1

Before presenting the proof of Theorem 1, we first introduce the following concentration inequality [1], which provides an upper tail bound for the sum of possibly dependent bounded random variables.

Theorem 1. *([1] Theorem 1.8) Suppose that $X_1, \ldots, X_n$ are random variables such that $0 \leq X_i \leq 1$. Set $p = \frac{1}{n}\sum_{i=1}^{n} E[X_i]$ and fix a real number ϑ such that $np+1 < \vartheta < n$. If $\epsilon_0 > 0$ satisfies $\vartheta - 1 = np + np\epsilon_0$, then*

$$P\left(\sum_{i=1}^{n} X_i \geq \vartheta\right) \leq 2\exp\left(-nD(p(1+\epsilon_0)||p)\right),$$

where $D(q||p) = q\ln\frac{q}{p} + (1-q)\ln\frac{1-q}{1-p}$.

Now we are ready to proceed with the proof of Theorem 1. We start by relating separation to score differences. The event $p_i^t \neq p_{i'}^t$ occurs if the prototype maximizing $S_j(\mathcal{H}_i^t)$ differs from that maximizing $S_j(\mathcal{H}_{i'}^t)$. Define the score difference:

$$\Delta_k = S_j(\mathcal{H}_i^t) - S_j(\mathcal{H}_{i'}^t) = [D(p_j;\mathcal{H}_i^t) - D(p_j;\mathcal{H}_{i'}^t)]. \tag{1}$$

The event $p_i^t = p_{i'}^t$ implies that the maximum scores align, i.e., $\arg\max_j S_j(\mathcal{H}_i^t) = \arg\max_j S_j(\mathcal{H}_{i'}^t)$. A sufficient condition for separation is that for some j, $\Delta_j > 0$, and for another k, $\Delta_k < 0$. We focus on a prototype j expected to favor $\mathcal{H}_i^t$.

Since $\mathcal{D}_H(\mathcal{H}_i^t, \mathcal{H}_{i'}^t) \geq \delta$, the histories differ at some time steps. We select a prototype p_j for which the expected per-slot score difference satisfies $\mathbb{E}\left[d(s,p_j;\mathcal{H}_i^t) - d(s,p_j;\mathcal{H}_{i'}^t)\right] \geq \frac{\delta}{c}$, where $(s,p_j;\mathcal{H}_i^t)$ is defined as $d(s,p_j;\mathcal{H}_i^t) = w_a \cdot \mathbb{I}\{a_i^{s-1} = a_j\} + w_M \cdot \mathbb{I}\{M^s = M_j\} + w_m \cdot \mathbb{I}\{m_i^s = m_j\}$. Then define the per-slot score difference as

$$Y_s = d(s,p_j;\mathcal{H}_i^t) - d(s,p_j;\mathcal{H}_{i'}^t), \tag{2}$$

for $s = t-W, \ldots, t-1$. Since $d(s,p_j;\mathcal{H}) \in [0, w_{t-1-s}]$, we have, $Y_s \in [-w_{t-1-s}, w_{t-1-s}]$. Given $w_{t-1-s} \leq 1$, it follows that $Y_s \in [-1,1]$. To normalize this difference, let

$$X_s = \frac{1+Y_s}{2}, \tag{3}$$

so $X_s \in [0,1]$, which is useful for bounding the probability of separation in the next step.

Next, we compute the expectation of X_s. Let $\mu_s = \mathbb{E}[Y_s]$, where the expectation is taken over any randomness in the history components $(a_i^s, M^{s+1}, m_i^{s+1})$ such as stochastic actions or feedback and potentially the selection of p_j if it is random. Then, $\mathbb{E}[X_s] = \mathbb{E}\left[\frac{1+Y_s}{2}\right] = \frac{1+\mu_s}{2}$.

Set the average expectation across the window as

$$\begin{aligned} p &= \frac{1}{W}\sum_{s=t-W}^{t-1}\mathbb{E}[X_s] = \frac{1}{W}\sum_{s=t-W}^{t-1}\frac{1+\mu_s}{2} \\ &= \frac{1}{2} + \frac{1}{2W}\sum_{s=t-W}^{t-1}\mu_s. \end{aligned}$$

Let $\bar{\mu} = \frac{1}{W}\sum_{s=t-W}^{t-1}\mu_s$. Thus, we have $p = \frac{1}{2} + \frac{\bar{\mu}}{2}$. Now, since $\mu_s = \mathbb{E}[Y_s] = \mathbb{E}[d(s,p_j;\mathcal{H}_i^t) - d(s,p_j;\mathcal{H}_{i'}^t)]$, and we chose p_j such that

$$\mathbb{E}[d(s,p_j;\mathcal{H}_i^t) - d(s,p_j;\mathcal{H}_{i'}^t)] \geq \frac{\delta}{c}$$

, it follows that $\bar{\mu} = \frac{1}{W}\sum_{s=t-W}^{t-1}\mu_s \geq \frac{\delta}{c}$, assuming the bound holds uniformly across slots (or adjusting c if the expectation varies). Therefore, we have $p = \frac{1}{2} + \frac{\bar{\mu}}{2} \geq \frac{1}{2} + \frac{\delta}{2c}$.

These random variables and their expectations allow us to bound the probability that $\Delta_j > 0$ in the next step, leveraging the normalized form of X_s.

To establish a lower bound on $P(p_i^t \neq p_{i'}^t)$, consider the event that the score difference $\Delta_j = S_j(\mathcal{H}_i^t) - S_j(\mathcal{H}_{i'}^t)$ for the chosen prototype p_j is positive, where p_j satisfies $\mathbb{E}[d(s, p_j; \mathcal{H}_i^t) - d(s, p_j; \mathcal{H}_{i'}^t)] \geq \frac{\delta}{c}$. Since $p_i^t \neq p_{i'}^t$ occurs if the maximizing prototypes differ, we have

$$P(p_i^t \neq p_{i'}^t) \geq P(\Delta_j > 0). \tag{4}$$

Given $\Delta_j = \sum_{s=t-W}^{t-1} Y_s$, and using the normalized variables $X_s = \frac{1+Y_s}{2}$, we express the sum as:

$$\sum_{s=t-W}^{t-1} X_s = \frac{W + \sum_{s=t-W}^{t-1} Y_s}{2},$$

so that $\Delta_j > 0$ (i.e., $\sum_{s=t-W}^{t-1} Y_s > 0$) holds if and only if $\sum_{s=t-W}^{t-1} X_s > \frac{W}{2}$. Since $E\left[\sum_{s=t-W}^{t-1} X_s\right] = Wp$, where $p = \frac{1}{2} + \frac{\bar{\mu}}{2} \geq \frac{1}{2} + \frac{\delta}{2c}$, we have $Wp \geq W\left(\frac{1}{2} + \frac{\delta}{2c}\right) > \frac{W}{2}$, indicating that $\sum_{s=t-W}^{t-1} X_s > \frac{W}{2}$ occurs when the sum exceeds its mean. To bound this probability, define a threshold $\vartheta = Wp + \lambda W\frac{\delta}{c}$, where $\lambda > 0$ is a constant to be determined, ensuring $\vartheta > Wp$. We require $Wp + 1 < \vartheta < W$, which enable the application of Theorem 1 to bound the complementary event $P\left(\sum_{s=t-W}^{t-1} X_s \leq \frac{W}{2}\right)$, corresponding to $\Delta_j \leq 0$.

The lower bound $Wp + 1 < \vartheta$ simplifies to $Wp + 1 < Wp + \lambda W\frac{\delta}{c}$, which leads to the condition $\lambda > \frac{c}{W\delta}$, while the upper bound $\vartheta < W$ gives $Wp + \lambda W\frac{\delta}{c} < W$, which in turn implies $p + \lambda\frac{\delta}{c} < 1$.

Since $p \geq \frac{1}{2} + \frac{\delta}{2c}$, we deduce that $1 - p \leq \frac{1}{2} - \frac{\delta}{2c}$, and hence the condition $\lambda\frac{\delta}{c} < \frac{1}{2} - \frac{\delta}{2c}$ must hold, which yields $\lambda < \frac{c}{2\delta} - \frac{1}{2} < \frac{c}{2\delta}$. Then, for $W > 2$, consistent with the window size assumption, we obtain the range $\frac{c}{W\delta} < \lambda < \frac{c}{2\delta}$.

Next, define ϵ_0 such that $\vartheta - 1 = Wp(1 + \epsilon_0)$. By equating both expressions for ϑ, we obtain $Wp + \lambda W\frac{\delta}{c} - 1 = Wp(1 + \epsilon_0)$, which gives $\epsilon_0 = \frac{\lambda\frac{\delta}{c} - \frac{1}{W}}{p}$. Furthermore, we observe that $\epsilon_0 \geq 2\lambda\frac{\delta}{c}$.

Now apply Theorem 1 to bound the upper tail probability $P\left(\sum_{s=t-W}^{t-1} X_s \geq \vartheta\right)$, which requires the KL divergence $D(p(1+\epsilon_0)||p)$. For small ϵ_0, the divergence can be approximated as $D(p(1+\epsilon_0)||p) \geq \frac{p\epsilon_0^2}{2(1-p)}$. Substituting $p \geq \frac{1}{2}$ and $\epsilon_0 \geq 2\lambda\frac{\delta}{c}$, we get

$$D(p(1+\epsilon_0)||p) \geq 2\lambda^2\left(\frac{\delta}{c}\right)^2. \tag{5}$$

Thus, Theorem 1 yields, $P\left(\sum_{s=t-W}^{t-1} X_s \geq \vartheta\right) \leq 2\exp\left(-W \cdot 2\lambda^2\left(\frac{\delta}{c}\right)^2\right)$, implying $P\left(\sum_{s=t-W}^{t-1} X_s < \vartheta\right) \geq 1 - 2\exp\left(-W \cdot 2\lambda^2\left(\frac{\delta}{c}\right)^2\right)$.

Since $\vartheta = Wp + \lambda W\frac{\delta}{c} \geq W\left(\frac{1}{2} + \frac{\delta}{2c}\right) + \lambda W\frac{\delta}{c} > \frac{W}{2}$, the event $\sum_{s=t-W}^{t-1} X_s < \vartheta$ implies $\sum_{s=t-W}^{t-1} X_s \leq \frac{W}{2}$, so:

$$\begin{aligned} P\left(\sum_{s=t-W}^{t-1} X_s \leq \frac{W}{2}\right) &\leq P\left(\sum_{s=t-W}^{t-1} X_s < \vartheta\right) \\ &\leq 2\exp\left(-W \cdot 2\lambda^2\left(\frac{\delta}{c}\right)^2\right). \end{aligned} \tag{6}$$

Therefore, $P(\Delta_j \leq 0) \leq 2\exp\left(-W \cdot 2\lambda^2\left(\frac{\delta}{c}\right)^2\right)$, and from (4) and (6), we can write

$$\begin{aligned} P(p_i^t \neq p_{i'}^t) &\geq 1 - P(\Delta_j \leq 0) \\ &\geq 1 - 2\exp\left(-W \cdot 2\lambda^2\left(\frac{\delta}{c}\right)^2\right). \end{aligned}$$

Since $2e^{-x} \leq e^{-x/2}$ for all $x \geq 2\log(2)$ and let $c' = \lambda^2\frac{1}{c^2}$, we get

$$P(p_i^t \neq p_{i'}^t) \geq 1 - e^{-c' W \delta^2},$$

where λ satisfies the constraints $\frac{c}{W\delta} < \lambda < \frac{c}{2\delta}$. Which complete the proof

The bound on $P(\Delta_j > 0)$ for a single prototype p_j expected to favor $\mathcal{H}_i^t$ provides a lower bound on $P(p_i^t \neq p_{i'}^t)$. In reality, $p_i^t \neq p_{i'}^t$ occurs if the maximizer of $S_j(\mathcal{H}_i^t)$ differs from that of $S_j(\mathcal{H}_{i'}^t)$, which may involve multiple prototypes. However, since $\mathcal{D}_H \geq \delta$, there exists at least one prototype where the score difference is significant, and our bound captures this dominant effect. A more precise bound could use a union bound over all prototypes, but this suffices for the desired exponential form.

Proof of Theorem 2

Before presenting the proof of Theorem 2, we state a preliminary result on the convergence of the empirical Q-values under fixed policy to their true Q-values. To analyze the convergence of Algorithm 1, we define the true action-value function for user i's policy π_i as:

$$Q_i^{\pi_i}(s_i^{t_e}, a_i^{t_e}) = \mathbb{E}_{\pi_i, \boldsymbol{\pi}_{-i}}\left[\sum_{l=t_e}^{T_h} \gamma^{l-t_e} r_i^l \mid s_i^{t_e}, a_i^{t_e}\right],$$

and the joint action-value function for joint policy $\boldsymbol{\pi}$ as

$$Q_{\boldsymbol{\pi}}(\boldsymbol{s}, \boldsymbol{a}) = \frac{1}{N}\sum_{i=1}^{N} Q_i^{\pi_i}(s_i, a_i). \tag{7}$$

The empirical action-value function $\tilde{Q}_i(s_i^{t_e}, a_i^{t_e})$ is the average Q-value estimate over E episodes, and the empirical joint action-value function is:

$$\tilde{Q}_{\boldsymbol{\pi}}(\boldsymbol{s}, \boldsymbol{a}) = \frac{1}{N}\sum_{i=1}^{N} \tilde{Q}_i(s_i, a_i).$$

The reward $r_i^{t_e}$ is bounded by $r_{i,min}^e = 0$ and $r_{i,max}^e = \frac{1-\gamma^T}{1-\gamma}\max(1, \phi)$, based on (7). The joint reward bounds are $R_{min}^e = 0$, $R_{max}^e = \frac{1-\gamma^T}{1-\gamma}\max(1, \phi)$.

Lemma 1. *Let each user $i \in \mathcal{N}$ follow policy π_i^e in episode e, generating a joint trajectory $\boldsymbol{\tau}^e = \{(\boldsymbol{s}^{t_e}, \boldsymbol{a}^{t_e}, \boldsymbol{r}^{t_e})\}_{t=0}^{T-1}$ with valid actions. For a number of episodes $E > \frac{(R_{max}^e)^2}{2\epsilon^2}\log\left(\frac{2}{\sigma}\right)$, the empirical joint action-value function satisfies:*

$$P\left(\left|\tilde{Q}_{\boldsymbol{\pi}}(\boldsymbol{s}, \boldsymbol{a}) - Q_{\boldsymbol{\pi}}(\boldsymbol{s}, \boldsymbol{a})\right| \leq \epsilon\right) \geq 1 - \sigma,$$

with probability parameter $\sigma = 2\exp\left(-\frac{2\epsilon^2 E}{(R_{max}^e)^2}\right)$.

Proof. The empirical action-value function $\tilde{Q}_i(s_i^{t_e}, a_i^{t_e})$ for user i is computed as the average cumulative discounted reward over E episodes, where each episode e yields rewards $r_i^{t_e} \in [0, \max(1, \phi)]$. The cumulative reward per episode is bounded by:

$$r_{i,min}^e = 0, \quad r_{i,max}^e = \frac{1-\gamma^T}{1-\gamma}\max(1, \phi).$$

Assuming independent trajectories across episodes (due to the stationary environment and fixed policies π_i^e within each episode), we apply the Chernoff-Hoeffding bound to the empirical estimate $\tilde{Q}_i(s_i^{t_e}, a_i^{t_e})$:

$$\begin{aligned} P\Bigg(\left|\tilde{Q}_i(s_i^{t_e}, a_i^{t_e}) - Q_i^{\pi_i}(s_i^{t_e}, a_i^{t_e})\right| \geq \epsilon\Bigg) \leq \\ \exp\left(-\frac{2\epsilon^2 E}{(r_{i,max}^e - r_{i,min}^e)^2}\right). \end{aligned} \tag{8}$$

Thus:

$$P\left(\left|\tilde{Q}_i(s_i^{t_e}, a_i^{t_e}) - Q_i^{\pi_i}(s_i^{t_e}, a_i^{t_e})\right| < \epsilon\right) > 1 - \sigma_i,$$

where $\sigma_i = 2\exp\left(-\frac{2\epsilon^2 E}{(r_{i,max}^e)^2}\right)$, since $r_{i,min}^e = 0$.

For the joint action-value function, since $\tilde{Q}_{\boldsymbol{\pi}}(\boldsymbol{s},\boldsymbol{a}) = \frac{1}{N}\sum_{i=1}^{N}\tilde{Q}_i(s_i,a_i)$, the error is:

$$\begin{aligned}\left|\tilde{Q}_{\boldsymbol{\pi}}(\boldsymbol{s},\boldsymbol{a}) - Q_{\boldsymbol{\pi}}(\boldsymbol{s},\boldsymbol{a})\right| &= \frac{1}{N}\left|\sum_{i=1}^{N}\left(\tilde{Q}_i(s_i,a_i) - Q_i^{\pi_i}(s_i,a_i)\right)\right| \\ &\leq \frac{1}{N}\sum_{i=1}^{N}\left|\tilde{Q}_i(s_i,a_i) - Q_i^{\pi_i}(s_i,a_i)\right|.\end{aligned} \tag{9}$$

Assuming the same ϵ for all users, if $\left|\tilde{Q}_i(s_i,a_i) - Q_i^{\pi_i}(s_i,a_i)\right| \leq \epsilon$ for each i, then:

$$\left|\tilde{Q}_{\boldsymbol{\pi}}(\boldsymbol{s},\boldsymbol{a}) - Q_{\boldsymbol{\pi}}(\boldsymbol{s},\boldsymbol{a})\right| \leq \epsilon.$$

The joint error probability is bounded using the union bound and the fact that $r_{i,max}^e = R_{max}^e$:

$$P\left(\left|\tilde{Q}_{\boldsymbol{\pi}}(\boldsymbol{s},\boldsymbol{a}) - Q_{\boldsymbol{\pi}}(\boldsymbol{s},\boldsymbol{a})\right| \geq \epsilon\right) \leq 2\exp\left(-\frac{2\epsilon^2 E}{(R_{max}^e)^2}\right). \tag{10}$$

To ensure the error is at most ϵ with probability at least $1-\sigma$, set:

$$2\exp\left(-\frac{2\epsilon^2 E}{(R_{max}^e)^2}\right) \leq \sigma.$$

Solving for E:

$$E \geq \frac{(R_{max}^e)^2}{2\epsilon^2}\log\left(\frac{2}{\sigma}\right).$$

Thus, for $E > \frac{(R_{max}^e)^2}{2\epsilon^2}\log\left(\frac{2}{\sigma}\right)$, we have:

$$P\left(\left|\tilde{Q}_{\boldsymbol{\pi}}(\boldsymbol{s},\boldsymbol{a}) - Q_{\boldsymbol{\pi}}(\boldsymbol{s},\boldsymbol{a})\right| \leq \epsilon\right) \geq 1-\sigma,$$

with $\sigma = 2\exp\left(-\frac{2\epsilon^2 E}{(R_{max}^e)^2}\right)$. $\square$

We also state the following theorem on the almost sure convergence of sequences of random variables [2].

Theorem 2. *([2] Theorem 1, p. 253) Let $\{\xi_n\}_{n\geq 1}$ be a sequence of random variables, and let ξ be a random variable. Then, $\xi_n \to \xi$ almost surely (P-a.s.) if and only if*

$$\lim_{n\to\infty}\Pr\Bigl(\sup_{k\geq n}|\xi_k - \xi| \geq \varepsilon\Bigr) = 0.$$

for every $\varepsilon > 0$.

Now, we proceed with the proof of Theorem 2. From the results of Lemma 1, we establish concentration bounds for the value function in Phase 1 for $e < ((r_{max}^e - r_{min}^e)/\epsilon)^2\log(2/\sigma)$, there exist $\epsilon > 0$, $\rho > 0$ such that,

$$P\left(\sum_{t=0}^{n^e-1}\gamma^t r_i^{t_e} - \sum_{t=0}^{n^{e-1}-1}\gamma^t r_i^{t_{e-1}} \geq \epsilon\right) \geq \rho, \tag{11}$$

where $t_e = t + (e-1)T + 1$, $t_{e-1} = t + (e-2)T + 1$, $n^e = |\{t : M^{t_{e-1}} = 1\}|$, and $r_i^{t_e}$ is the reward from (7) in Phase 1. Let e_ϵ be the first episode for which the monotone nondecreasing sequence $\sum_{t=0}^{n^e-1}\gamma^t r_i^{t_e}$ becomes strictly larger than $Q_i^*(s_i^{t_e}, a_i^{t_e}) - \epsilon$.

$$\begin{aligned}&\text{If } \sum_{t=0}^{n^e-1}\gamma^t r_i^{t_e} \leq Q_i^*(s_i^{t_e}, a_i^{t_e}) - \epsilon, \text{ then} \\ &\mathbb{E}\left[\sum_{t=0}^{n^e-1}\gamma^t r_i^{t_e} - \sum_{t=0}^{n^{e-1}-1}\gamma^t r_i^{t_{e-1}}\right] \geq \epsilon\rho, \quad \forall e < e_\epsilon.\end{aligned} \tag{12}$$

From this and the fact that $Q_i^e(s_i^{t_e}, a_i^{t_e})$ is bounded above by $Q_i^*(s_i^{t_e}, a_i^{t_e})$, $\sum_{t=0}^{n^e-1} \gamma^t r_i^{t_e} - \sum_{t=0}^{n^{e-1}-1} \gamma^t r_i^{t_{e-1}} \geq 0$ when $n^e \geq n^{e-1}$, and $\sum_{t=0}^{n^e-1} \gamma^t r_i^{t_e} \geq 0$ with probability one, we have

$$\begin{aligned} Q_i^*(s_i^{t_e}, a_i^{t_e}) \geq \mathbb{E}\left[\lim_{e\to\infty} \sum_{t=0}^{n^e-1} \gamma^t r_i^{t_e}\right] = \mathbb{E}\left[\sum_{t=0}^{n^0-1} \gamma^t r_i^{t_0}\right] + \\ \mathbb{E}\left[\sum_{e=0}^{\infty}\left(\sum_{t=0}^{n^{e+1}-1} \gamma^t r_i^{t_{e+1}} - \sum_{t=0}^{n^e-1} \gamma^t r_i^{t_e}\right)\right] \\ \geq \mathbb{E}\left[\sum_{t=0}^{n^0-1} \gamma^t r_i^{t_0}\right] + \\ \mathbb{E}\left[\sum_{e=0}^{e_\epsilon}\left(\sum_{t=0}^{n^{e+1}-1} \gamma^t r_i^{t_{e+1}} - \sum_{t=0}^{n^e-1} \gamma^t r_i^{t_e}\right)\right] \\ \geq \mathbb{E}\left[\sum_{t=0}^{n^0-1} \gamma^t r_i^{t_0}\right] + \epsilon\rho\mathbb{E}\left[e_\epsilon\right]. \end{aligned} \tag{13}$$

From the expectation condition and the bound on e_ϵ, we get:

$$\mathbb{E}\left[e_\epsilon\right] \leq \frac{Q_i^*(s_i^{t_e}, a_i^{t_e}) - \mathbb{E}\left[\sum_{t=0}^{n^0-1} \gamma^t r_i^{t_0}\right]}{\epsilon\rho} < \infty. \tag{14}$$

It follows that,

$$\begin{aligned} &\lim_{e\to\infty} \Pr\left(\sup_{e'\geq e}\left|Q_i^*(s_i^{t_e}, a_i^{t_e}) - \sum_{t=0}^{n^{e'}-1} \gamma^t r_i^{t_{e'}}\right| \geq \epsilon\right) \\ &= \lim_{e\to\infty} \Pr\left(Q_i^*(s_i^{t_e}, a_i^{t_e}) - \sum_{t=0}^{n^e-1} \gamma^t r_i^{t_e} \geq \epsilon\right) \\ &= \lim_{e\to\infty} \Pr\left(e_\epsilon \geq e\right) = 0, \end{aligned} \tag{15}$$

where the first equality is due to the monotonicity of $\mathbb{E}[\sum_{t=0}^{n^e-1} \gamma^t r_i^{t_e}]$.

Using Theorem 2, we conclude that $Q_i^e(s_i^{t_e}, a_i^{t_e}) \to Q_i^*(s_i^{t_e}, a_i^{t_e})$ with probability one. Therefore, as $n^e \to T$, the cooperative phase dominates, and the Q-values converge to the optimal policy for each agent i.

References

[1] C. Pelekis and J. Ramon, "Hoeffding's inequality for sums of weakly dependent random variables," 2015. [Online]. Available: https://arxiv.org/abs/1507.06871

[2] A. N. Shiryaev, "Probability, volume 95 of," *Graduate texts in mathematics*, p. 81, 1996.