Cross-Frame Resource Allocation with Context-Aware QoE Estimation for 360° Video Streaming in Wireless Virtual Reality

Supplementary Document

Cheng-Yeh Chen and Hung-Yun Hsieh

In this supplementary document, we provide a complete proof for Theorem 1 aiming to bound the regret with delayed update in the online learning algorithm for context-aware QoE estimation.

Theorem 1. *(Regret bound with delayed update) Let* $K_t = t^{\frac{2\alpha}{3\alpha+D}} \log(t)$ *and* $h_d \leq \lceil T^{\frac{1}{3\alpha+D}} \rceil$, $\forall d = 1, \cdots, D$*. If there exists* $L > 0$ *and* $\alpha > 0$ *such that* $|\beta(\chi) - \beta(\chi')| \leq L\|\chi - \chi'\|^{\alpha}$ *for any two contexts* χ *and* χ'*, where* $\beta(\cdot)$ *is a function mapping each context to its QoE, then the complexity of the bound for* R_T *belongs to* $\mathcal{O}\left(W_{\max}^2 2^D \log(T) T^{\frac{2\alpha+D}{3\alpha+D}}\right)$.

Before proving Theorem 1, we first define an assumption for the relationship between QoE and context space. Based on this assumption, we prove a proposition on the sub-linearity of the proposed CMAB without considering delayed feedback and finally include the effect of delayed feedback to complete the proof for Theorem 1. To bound the regret of learning, our proposed QoE should be aware "enough" of context information. Assumption 1 assumes that statistically similar context information implies similar QoE.

Assumption 1. *(Hölder condition) Define* $\|\cdot\|$ *as the Euclidean norm. There exists* $L > 0$ *and* $\alpha > 0$ *such that*

$$|\beta_t(\chi) - \beta_t(\chi')| \leq L\|\chi - \chi'\|^{\alpha} \tag{1}$$

for any two contexts χ *and* χ'*, where* $\beta_t(\chi)$ *is a function mapping each context* χ *to its QoE* β_t.

Assumption 1 is required for the following proposition and theorem. Such an assumption could be satisfied since the defined QoE is bounded within 0 and 1. There must exist L and α meeting Assumption 1.

To proceed, we first prove the regret bound assuming all the updates of QoE estimation are instant (no delayed update is incurred) in Proposition 1.

Proposition 1. *(Regret bound) Let* $K_t = t^{\frac{2\alpha}{3\alpha+D}} \log(t)$ *and* $h_d \leq \lceil T^{\frac{1}{3\alpha+D}} \rceil$, $\forall d = 1, \cdots, D$. *If Assumption 1 holds, the complexity of* R_T *belongs to* $O\left(T^{\frac{2\alpha+D}{3\alpha+D}} \log(T)\right)$*, being sub-linear in* T.

Proof. For each hypercube $p \in \mathcal{P}_{\mathcal{X}}$, let $\bar{\beta}(p) = \sup_{\chi \in p} \beta(\chi)$ and $\underline{\beta}(p) = \inf_{\chi \in p} \beta(\chi)$ be the highest and lowest expected QoE over all context $\chi \in p$, where $\beta(\chi)$ denotes the expected QoE mapped from χ. Also let $\hat{\chi}_p$ be the context at the geometrical center of a hypercube p and $\hat{\beta}(p) = \beta(\hat{\chi}_p)$ be the corresponding expected QoE. Given $\chi(i,j,t')$ as the context of (i,j,t') and $\boldsymbol{p}_t = \{f_{\mathcal{X}}\big(\chi(i,j,t')\big) \mid \forall i,j,t'\}$ as the set of corresponding hypercubes, we define

$$\bar{\boldsymbol{\beta}}_t = \{\bar{\beta}(p) \mid \forall p \in \boldsymbol{p}_t\}, \tag{2}$$

$$\underline{\boldsymbol{\beta}}_t = \{\underline{\beta}(p) \mid \forall p \in \boldsymbol{p}_t\}, \tag{3}$$

$$\hat{\boldsymbol{\beta}}_t = \{\hat{\beta}(p) \mid \forall p \in \boldsymbol{p}_t\}. \tag{4}$$

For a scheduling decision at t, we define $\hat{\boldsymbol{\mathcal{E}}}_t$ as specific resource allocation satisfying

$$\hat{\boldsymbol{\mathcal{E}}}_t = \mathop{\arg\max}_{\boldsymbol{\mathcal{E}}_t \mid \mathbb{E}\{E_{t'}\} \leq B, \forall t' \in \{t, \cdots, t+W_t\}} r(\hat{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t). \tag{5}$$

$\hat{\boldsymbol{\mathcal{E}}}_t$ could help identifying the set of poor resource allocations:

$$\mathcal{L}_t = \left\{\boldsymbol{\mathcal{E}}_t^{\mathcal{L}} \mid \mathbb{E}\left\{E_{t'}^{\mathcal{L}}\right\} \leq B, \forall t' \in \{t, \cdots, t+W_t\}, r(\underline{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) - r(\bar{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \geq At^{\theta}\right\}, \tag{6}$$

which is the set of sub-optimal subsets of resource allocation for hypercube set $\boldsymbol{p}_t$ with $A > 0$ and $\theta < 0$ being the parameters only in the regret analysis. Note that $\boldsymbol{\mathcal{E}}_t^{\mathcal{L}}$ in $\mathcal{L}_t$ is sub-optimal for $\boldsymbol{p}_t$ since the sum of the lowest expected reward of $\hat{\boldsymbol{\mathcal{E}}}_t$ is higher than the best expected reward of $\boldsymbol{\mathcal{E}}_t^{\mathcal{L}}$ by at least At^{θ}. As a result, the resource allocation not in $\mathcal{L}_t$ could be treated as near-optimal candidate for $\boldsymbol{p}_t$.

The regret $R(T)$ is divided into three components:

$$R(T) = \mathbb{E}\left\{R_e(T)\right\} + \mathbb{E}\left\{R_s(T)\right\} + \mathbb{E}\left\{R_n(T)\right\}, \tag{7}$$

where $\mathbb{E}\left\{R_e(T)\right\}$ accounts for the regret incurred by exploration, $\mathbb{E}\left\{R_s(T)\right\}$ accounts for the regret incurred by the sub-optimal decisions in $\mathcal{L}_t$ during exploitation, and $\mathbb{E}\left\{R_n(T)\right\}$ accounts for the regret resulting from near-optimal decisions during exploitation. The regret bound could be given by the three seperate components one by one through Lemma 1, 2, and 3.

Lemma 1. *(Bound for $\mathbb{E}\left\{R_e(T)\right\}$) Given $K = t^z \log(t)$ and $h_d \leq \lceil T^\gamma \rceil, \forall d = 1, \cdots, D$, where $0 < z < 1$ and $0 < \gamma < \frac{1}{D}$, the regret $\mathbb{E}\left\{R_e(T)\right\}$ could be bounded by*

$$\mathbb{E}\left\{R_e(T)\right\} \leq W_{\max} 2^D (T^{z+\gamma D} \log(T) + T^{\gamma D}). \tag{8}$$

Proof. For exploration phase, there exists some under-explored hypercubes ($\exists\, p \in \mathcal{P}_{\mathcal{X}}$ s.t. $|\mathcal{A}_p| \leq K_t = t^z \log(t)$). There could be at most $\lceil T^z \log(T) \rceil$ exploration phases for each hypercube. Note that there are $\prod_{d=1}^{D} h_d$ hypercubes in the context space. The maximum number of exploration phases is $\left(\prod_{d=1}^{D} h_d\right) \lceil T^z \log(T) \rceil$. Since the maximum and minimum achievable QoE in the objective function are $W_{\max}$ and 0, the maximum regret of poor resource allocation in one exploration phase is bounded by $W_{\max}$. As a result, the overall regret for $\mathbb{E}\left\{R_e(T)\right\}$ is given by

$$\begin{aligned}\mathbb{E}\left\{R_e(T)\right\} \leq & W_{\max} \left(\prod_{d=1}^{D} h_d\right) \lceil T^z \log(T) \rceil \leq W_{\max} \lceil T^\gamma \rceil^D \lceil T^z \log(T) \rceil \\ \leq & W_{\max} 2^D T^{\gamma D} (T^z \log(T) + 1) = W_{\max} 2^D (T^{z+\gamma D} \log(T) + T^{\gamma D}),\end{aligned} \tag{9}$$

using the fact that $\lceil T^\gamma \rceil^D \leq (2T^\gamma)^D = 2^D T^{\gamma D}$. □

Lemma 2. *(Bound for $\mathbb{E}\left\{R_s(T)\right\}$) Given $K = t^z \log(t)$ and $h_d \leq \lceil T^\gamma \rceil, \forall d = 1, \cdots, D$, where $0 < z < 1$ and $0 < \gamma < \frac{1}{D}$, if Assumption 1 holds and $2M(t^{-z/2} + LD^\alpha h_d^{-\alpha}) \leq At^\theta$ holds for $1 \leq t \leq T$ where $M = \mathcal{N}_x \mathcal{N}_y W_{\max}$, the regret $\mathbb{E}\left\{R_e(T)\right\}$ could be bounded by*

$$\mathbb{E}\left\{R_s(T)\right\} \leq W_{\max} \sum_{k=1}^{M} \begin{pmatrix} M \\ k \end{pmatrix} M \frac{\pi^2}{3}. \tag{10}$$

Proof. For exploitation phase, no hypercube is under-explored, indicating that $|\mathcal{A}_p| > K_t = t^z \log(t), \forall\ p \in \mathcal{P}_{\mathcal{X}}$. Let G_t^s denote the event that time step t is in exploitation phase and $G_t^{\mathcal{L}}$ denote the the event that $\mathcal{E}_t^{\mathcal{L}} \in \mathcal{L}_t$ is selected in t. The regret for exploitation when $\mathcal{E}_t^{\mathcal{L}} \in \mathcal{L}_t$ is selected could be written as

$$\begin{aligned}\mathbb{E}\left\{R_s(T)\right\} = & \sum_{t=1}^{T} \sum_{\boldsymbol{\mathcal{E}}_t^{\mathcal{L}} \in \mathcal{L}_t} \mathbb{E}\left\{ I_{\{G_t^s, G_t^{\mathcal{L}}\}} \times \left(r(\boldsymbol{\mathcal{E}}_t^*) - r(\boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \right) \right\} \\ = & \sum_{t=1}^{T} \sum_{\boldsymbol{\mathcal{E}}_t^{\mathcal{L}} \in \mathcal{L}_t} \Pr\{G_t^s, G_t^{\mathcal{L}}\} \times \left(r(\boldsymbol{\mathcal{E}}_t^*) - r(\boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \right).\end{aligned} \tag{11}$$

Since the maximum regret of poor resource allocation is bounded by $W_{\max}$, it holds that

$$\mathbb{E}\left\{R_s(T)\right\} \leq W_{\max} \sum_{t=1}^{T} \sum_{\boldsymbol{\mathcal{E}}_t^{\mathcal{L}} \in \mathcal{L}_t} \Pr\{G_t^s, G_t^{\mathcal{L}}\}. \tag{12}$$

Let $\tilde{\boldsymbol{\beta}}_t$ denote the current QoE estimation. For the event $G_t^{\mathcal{L}}$ to happen, it must hold that $r(\tilde{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \geq r(\tilde{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t)$. Otherwise, $\boldsymbol{\mathcal{E}}_t^{\mathcal{L}}$ would not have been selected. Therefore, we have

$$\Pr\{G_t^s, G_t^{\mathcal{L}}\} \leq \Pr\left\{r(\tilde{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \geq r(\tilde{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t)\right\} \tag{13}$$

The event $\left\{r(\tilde{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \geq r(\tilde{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t)\right\}$ could be viewed as the subset of the union of the following three events:

$$\left\{r(\tilde{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \geq r(\tilde{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t)\right\} \subseteq G_1 \cup G_2 \cup G_3, \tag{14}$$

where

$$G_1 = \left\{r(\tilde{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \geq r(\bar{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) + H(t), G_t^s\right\}, \tag{15}$$

$$G_2 = \left\{r(\tilde{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) \leq r(\underline{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) - H(t), G_t^s\right\}, \tag{16}$$

$$\begin{aligned} G_3 = \Big\{ & r(\tilde{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \geq r(\tilde{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t), r(\tilde{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) < r(\bar{\boldsymbol{\beta}}_t), \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) + H(t), \\ & r(\tilde{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) > r(\underline{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) - H(t), G_t^s\Big\}. \end{aligned} \tag{17}$$

We will bound the probability of G_1, G_2, and G_3 step by step. For clarity, define $n = (i, j, t')$ to simplify the notation for a tile. We start from the upper bound of $\Pr\{G_1\}$. Previously, we define the highest expected QoE over hypercube p as $\bar{\beta}(p) = \sup_{\chi \in p} \beta(\chi)$. We further annotate $\bar{\beta}(p)$ by $\bar{\beta}(p_n) = \sup_{\chi \in p_n} \beta(\chi)$ where p_n denotes the hypercube corresponding to the tile n. The expected QoE for tile n is bounded by

$$\mathbb{E}\{\tilde{\beta}(p_n)\} = \mathbb{E}\left\{\frac{1}{|\mathcal{A}_{p_n}|} \sum_{\chi \in p_n} \beta_t(\chi)\right\} \leq \frac{1}{|\mathcal{A}_{p_n}|} \sum_{\chi \in p_n} \bar{\beta}(p_n) = \bar{\beta}(p_n). \tag{18}$$

Note that $\beta_t(\chi)$ denotes the actual QoE mapped from context χ. Also note that the final equivalence holds due to the fact that the number fo summation over $\chi \in p_n$ equals $|\mathcal{A}_{p_n}|$ by definition. By using the relation in (18), we have

$$\begin{aligned}
\Pr\{G_1\} =& \Pr\left\{ r(\tilde{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \geq r(\bar{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) + H(t), G_t^s \right\} \\
\leq& \Pr\left\{ \tilde{\beta}(p_n) \geq \bar{\beta}(p_n) + \frac{H(t)}{M}, \exists\; n \in \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}, G_t^s \right\} \\
\leq& \Pr\left\{ \tilde{\beta}(p_n) \geq \mathbb{E}\{\tilde{\beta}(p_n)\} + \frac{H(t)}{M}, \exists\; n \in \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}, G_t^s \right\} \\
=& \sum_{n \in \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}} \Pr\left\{ \tilde{\beta}(p_n) \geq \mathbb{E}\{\tilde{\beta}(p_n)\} + \frac{H(t)}{M}, G_t^s \right\}
\end{aligned} \tag{19}$$

Equation (19) could be further bounded by the Chernoff-Hoeffdling bound:

$$\begin{aligned}
\Pr\{G_1\} \leq& \sum_{n \in \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}} \Pr\left\{ \tilde{\beta}(p_n) \geq \mathbb{E}\{\tilde{\beta}(p_n)\} + \frac{H(t)}{M}, G_t^s \right\} \\
\leq& \sum_{n \in \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}} \exp\left(\frac{-2|A_{p_n}|H(t)^2}{M^2} \right) \\
\leq& \sum_{n \in \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}} \exp\left(\frac{-2t^z \log(t) H(t)^2}{M^2} \right).
\end{aligned} \tag{20}$$

The bound for $\Pr\{G_2\}$ can be proved similarly, leading to

$$\begin{aligned}
\Pr\{G_2\} \leq& \sum_{n \in \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}} \Pr\left\{ r(\tilde{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) \leq r(\underline{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) - H(t), G_t^s \right\} \\
\leq& \sum_{n \in \hat{\boldsymbol{\mathcal{E}}}_t} \exp\left(\frac{-2t^z \log(t) H(t)^2}{M^2} \right).
\end{aligned} \tag{21}$$

To prove the bound for $\Pr\{G_3\}$, we rewrite the current QoE estimation $\tilde{\boldsymbol{\beta}}(p)$ into

$$\tilde{\beta}(p) = \frac{1}{|\mathcal{A}_p|} \sum_{\chi \in p} \beta_t(\chi) = \frac{1}{|\mathcal{A}_p|} \sum_{\chi \in p} \beta(\chi) + \epsilon_\chi, \tag{22}$$

where ϵ_χ denotes the deviation between the actual QoE $\beta_t(\chi)$ and expected QoE $\beta(\chi)$ with context χ. Similar to the definition in (2), we define the best and worst context for a hypercube as $\chi^{\text{best}}(p) = \operatorname*{arg\,max}_{\chi \in p} \beta(\chi)$ and $\chi^{\text{worst}}(p) = \operatorname*{arg\,min}_{\chi \in p} \beta(\chi)$. The best and worst QoE estimation could be further written as

$$\beta^{\text{best}}(p) = \frac{1}{|\mathcal{A}_p|} \sum_{\chi \in p} \left(\beta(\chi^{\text{best}}(p)) + \epsilon_\chi \right), \tag{23}$$

$$\beta^{\text{worst}}(p) = \frac{1}{|\mathcal{A}_p|} \sum_{\chi \in p} \left(\beta(\chi^{\text{worst}}(p)) + \epsilon_\chi \right). \tag{24}$$

Similar to the definition of $\bar{\boldsymbol{\beta}}_t$ and $\underline{\boldsymbol{\beta}}$, we define $\boldsymbol{\beta}_t^{\text{best}} = \left\{\beta^{\text{best}}(p) \mid \forall p \in \boldsymbol{p}\right\}$ and $\boldsymbol{\beta}_t^{\text{worst}} = \{\beta^{\text{worst}}(p) \mid \forall p \in \boldsymbol{p}_t\}$. By Assumption 1, one could show that

$$\beta^{\text{best}}(p) - \tilde{\beta}(p) \leq LD^{\frac{\alpha}{2}} h_d^{-\alpha}, \forall d = 1, \cdots, D, \tag{25}$$

and

$$\tilde{\beta}(p) - \beta^{\text{worst}}(p) \leq LD^{\frac{\alpha}{2}} h_d^{-\alpha}, \forall d = 1, \cdots, D. \tag{26}$$

Applying (25) and (26) to the resource allocation $\boldsymbol{\mathcal{E}}_t^{\mathcal{L}}$ and $\hat{\boldsymbol{\mathcal{E}}}_t$, we have

$$r(\boldsymbol{\beta}_t^{\text{best}}, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) - r(\tilde{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \leq \sum_{n \in \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}} \left(\beta^{\text{best}}(p_n) - \tilde{\beta}(p_n)\right) \leq MLD^{\frac{\alpha}{2}} h_d^{-\alpha}, \forall d = 1, \cdots, D, \tag{27}$$

and

$$r(\tilde{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) - r(\boldsymbol{\beta}_t^{\text{worst}}, \hat{\boldsymbol{\mathcal{E}}}_t) \leq \sum_{n \in \hat{\boldsymbol{\mathcal{E}}}_t} \left(\tilde{\beta}(p_n) - \beta^{\text{worst}}(p_n)\right) \leq MLD^{\frac{\alpha}{2}} h_d^{-\alpha}, \forall d = 1, \cdots, D. \tag{28}$$

There are three components in the event of G_3. By the definition of (25) and (26), the first component in G_3 holds that

$$\left\{r(\tilde{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \geq r(\tilde{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t)\right\} \subseteq \left\{r(\tilde{\boldsymbol{\beta}}_t^{\text{best}}, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \geq r(\tilde{\boldsymbol{\beta}}_t^{\text{worst}}, \hat{\boldsymbol{\mathcal{E}}}_t)\right\}. \tag{29}$$

By (27), the second component in G_3 holds that

$$\begin{aligned} &\left\{r(\tilde{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) < r(\bar{\boldsymbol{\beta}}_t), \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) + H(t)\right\} \\ \subseteq &\left\{r(\boldsymbol{\beta}_t^{\text{best}}, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) - MLD^{\frac{\alpha}{2}} h_d^{-\alpha} < r(\bar{\boldsymbol{\beta}}_t), \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) + H(t)\right\} \\ = &\left\{r(\boldsymbol{\beta}_t^{\text{best}}, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) < r(\bar{\boldsymbol{\beta}}_t), \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) + MLD^{\frac{\alpha}{2}} h_d^{-\alpha} + H(t)\right\}. \end{aligned} \tag{30}$$

Similarly, the third component in G_3 holds that

$$\begin{aligned} &\left\{r(\tilde{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) > r(\underline{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) - H(t)\right\} \\ \subseteq &\left\{r(\boldsymbol{\beta}_t^{\text{worst}}, \hat{\boldsymbol{\mathcal{E}}}_t) + MLD^{\frac{\alpha}{2}} h_d^{-\alpha} > r(\underline{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) - H(t)\right\} \\ = &\left\{r(\boldsymbol{\beta}_t^{\text{worst}}, \hat{\boldsymbol{\mathcal{E}}}_t) > r(\underline{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) - MLD^{\frac{\alpha}{2}} h_d^{-\alpha} - H(t)\right\} \end{aligned} \tag{31}$$

Combining (29), (30), and (31), we could bound $\Pr\{G_3\}$ by

$$\begin{aligned} \Pr\{G_3\} \leq \Pr \Big\{ & r(\tilde{\boldsymbol{\beta}}_t^{\text{best}}, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) \geq r(\tilde{\boldsymbol{\beta}}_t^{\text{worst}}, \hat{\boldsymbol{\mathcal{E}}}_t), \\ & r(\boldsymbol{\beta}_t^{\text{best}}, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) < r(\bar{\boldsymbol{\beta}}_t), \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) + MLD^{\frac{\alpha}{2}} h_d^{-\alpha} + H(t), \\ & r(\boldsymbol{\beta}_t^{\text{worst}}, \hat{\boldsymbol{\mathcal{E}}}_t) > r(\underline{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) - MLD^{\frac{\alpha}{2}} h_d^{-\alpha} - H(t), G_t^s \Big\}. \end{aligned} \tag{32}$$

It can be shown that $\Pr\{G_3\} = 0$. Supposed $2H(t) + 2MLD^{\frac{\alpha}{2}}h_d^{-\alpha} \leq AT^{\theta}$ is satisfied. Since $\boldsymbol{\mathcal{E}}_t^{\mathcal{L}} \in \mathcal{L}_t$, we also have $r(\underline{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) - r(\bar{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t^{)} \geq At^{\theta}$. Combining these two equations yields

$$r(\underline{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) - r(\bar{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) - (2H(t) + 2MLD^{\frac{\alpha}{2}}h_d^{-\alpha}) \geq 0, \tag{33}$$

which can be written as

$$r(\underline{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) - H(t) - MLD^{\frac{\alpha}{2}}h_d^{-\alpha} \geq r(\bar{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) + H(t) + MLD^{\frac{\alpha}{2}}h_d^{-\alpha}. \tag{34}$$

However, if (34) is satisfied, the three components in (32) can not hold true simultaneously. The second and third component of (32) indicating that $r(\boldsymbol{\beta}_t^{\text{best}}, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) < r(\boldsymbol{\beta}_t^{\text{worst}}, \hat{\boldsymbol{\mathcal{E}}}_t)$, contradicting the first component of (32). As a result, $\Pr\{G_3\} = 0$ under condition (32).

By setting $H(t) = Mt^{-z/2}$, we can rewrite (20) and (21) into

$$\Pr\{G_2\} \leq M\exp\left(\frac{-2H(t)^2 t^z \log(t)}{M^2}\right) \leq M\exp(-2\log(t)) \leq Mt^{-2}. \tag{35}$$

and similarly,

$$\Pr\{G_2\} \leq Mt^{-2}. \tag{36}$$

In conclusion, starting from (13) and (14), under condition (32), we have

$$\begin{aligned}\Pr\{G_t^s, G_t^{\mathcal{L}}\} \leq& \Pr\{G_1 \cup G_2 \cup G_3\} \\ \leq& \Pr\{G_1\} + \Pr\{G_2\} + \Pr\{G_3\} \\ \leq& 2Mt^{-2}.\end{aligned} \tag{37}$$

Finally, the bound of $\mathbb{E}\left\{R_s(T)\right\}$ could be derived from (12)

$$\begin{aligned}\mathbb{E}\left\{R_s(T)\right\} \leq& W_{\max}\sum_{t=1}^{T}\sum_{\boldsymbol{\mathcal{E}}_t^{\mathcal{L}} \in \mathcal{L}_t} \Pr\{G_t^s, G_t^{\mathcal{L}}\} \leq W_{\max}|\mathcal{L}_t|\sum_{t=1}^{T} 2Mt^{-2} \\ \leq& W_{\max}|\mathcal{L}_t| 2M \sum_{t=1}^{\infty} t^{-2} \leq W_{\max}|\mathcal{L}_t| 2M\frac{\pi^2}{3} \\ \leq& W_{\max}\sum_{k=1}^{M}\left(\begin{array}{c} M \\ k \end{array}\right) M\frac{\pi^2}{3}.\end{aligned} \tag{38}$$

□

Lemma 3. *(Bound for $\mathbb{E}\left\{R_n(T)\right\}$) Given $K = t^z\log(t)$ and $h_d \leq \lceil T^{\gamma}\rceil, \forall d = 1, \cdots, D$, where $0 < z < 1$ and $0 < \gamma < \frac{1}{D}$, if Assumption 1 holds and $2M(t^{-z/2} + LD^{\alpha}h_d^{-\alpha}) \leq At^{\theta}$ holds for $1 \leq t \leq T$ where $M = \mathcal{N}_x\mathcal{N}_y W_{\max}$, the regret $\mathbb{E}\left\{R_n(T)\right\}$ could be bounded by*

$$\mathbb{E}\left\{R_n(T)\right\} \leq 3MLD^{\frac{\alpha}{2}}T^{1-\gamma\alpha} + \frac{A}{1+\theta}t^{1+\theta}. \tag{39}$$

Proof. Let $G_t^n = G_t^s \cap \{\boldsymbol{\mathcal{E}}_t \in \boldsymbol{\mathcal{H}}_t\}$ denote the event of selecting a near-optimal resource allocation, where

$$\mathcal{H}_t = \left\{\boldsymbol{\mathcal{E}}_t^{\mathcal{L}} \mid \mathbb{E}\left\{E_{t'}^{\mathcal{L}}\right\} \le B, \forall t' \in \{t, \cdots, t+W_t\}, r(\underline{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t) - r(\bar{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{L}}) < At^{\theta}\right\}, \tag{40}$$

For exploitation phase, the regret due to near-optimal resource allocation could be written as

$$\begin{aligned}
\mathbb{E}\left\{R_n(T)\right\} =& \sum_{t=1}^{T} \mathbb{E}\left\{I_{\{G_t^n \cap \{\boldsymbol{\mathcal{E}}_t^{\mathcal{H}} \in \boldsymbol{\mathcal{H}}_t\}\}} \times \left(r(\boldsymbol{\beta}_t, \boldsymbol{\mathcal{E}}_t^*) - r(\boldsymbol{\beta}_t, \boldsymbol{\mathcal{E}}_t \mathcal{H})\right)\right\} \\
=& \sum_{t=1}^{T} \Pr\{G_t^n\}\left(r(\boldsymbol{\beta}_t, \boldsymbol{\mathcal{E}}_t^*) - r(\boldsymbol{\beta}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{H}})\right) \\
\le& \sum_{t=1}^{T} \left(r(\boldsymbol{\beta}_t, \boldsymbol{\mathcal{E}}_t^*) - r(\boldsymbol{\beta}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{H}})\right).
\end{aligned} \tag{41}$$

By definition of (40), we have

$$r(\boldsymbol{\beta}_t, \boldsymbol{\mathcal{E}}_t^*) - r(\boldsymbol{\beta}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{H}}) < At^{\theta}. \tag{42}$$

Applying Assumption 1 multiple times, we have

$$\begin{aligned}
& r(\boldsymbol{\beta}_t, \boldsymbol{\mathcal{E}}_t^*) - r(\boldsymbol{\beta}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{H}}) \\
\le& r(\hat{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^*) + MLD^{\frac{\alpha}{2}} h_d^{-\alpha} - r(\boldsymbol{\beta}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{H}}) \\
\le& r(\hat{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t^*) + MLD^{\frac{\alpha}{2}} h_d^{-\alpha} - r(\boldsymbol{\beta}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{H}}) \\
\le& r(\underline{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t^*) + 2MLD^{\frac{\alpha}{2}} h_d^{-\alpha} - r(\boldsymbol{\beta}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{H}}) \\
\le& r(\underline{\boldsymbol{\beta}}_t, \hat{\boldsymbol{\mathcal{E}}}_t^*) + 3MLD^{\frac{\alpha}{2}} h_d^{-\alpha} - r(\bar{\boldsymbol{\beta}}_t, \boldsymbol{\mathcal{E}}_t^{\mathcal{H}}) \\
\le& 3MLD^{\frac{\alpha}{2}} h_d^{-\alpha} + At^{\theta}.
\end{aligned} \tag{43}$$

Setting $h_d^{-\alpha} = \lceil T^{\gamma} \rceil^{-\alpha} \le T^{-\gamma\alpha}$, we have

$$\begin{aligned}
\mathbb{E}\left\{R_n(T)\right\} \le& \sum_{t=1}^{T} \left(3MLD^{\frac{\alpha}{2}} h_d^{-\alpha} + At^{\theta}\right) \\
\le& \sum_{t=1}^{T} \left(3MLD^{\frac{\alpha}{2}} T^{-\gamma\alpha} + At^{\theta}\right) \\
\le& 3MLD^{\frac{\alpha}{2}} T^{1-\gamma\alpha} + \frac{A}{1+\theta} t^{1+\theta}.
\end{aligned} \tag{44}$$

The proof of proposition 1 could finally be completed by summing up the bound for each component in Lemma 1, Lemma 2, and Lemma 3:

$$\begin{aligned} R(T) \leq & W_{\max} 2^D (T^{z+\gamma D} \log(T) + T^{\gamma D}) + W_{\max} \sum_{k=1}^{M} \left(\begin{array}{c} M \\ k \end{array} \right) M \frac{\pi^2}{3} \\ & + 3MLD^{\frac{\alpha}{2}} T^{1-\gamma\alpha} + \frac{A}{1+\theta} t^{1+\theta}. \end{aligned} \tag{45}$$

To balance the leading orders, parameters like z, γ, A, and θ are specified as $z = \frac{2\alpha}{3\alpha+D} \in (0,1)$, $\gamma = \frac{z}{2\alpha} \in (0, \frac{1}{D})$, $\theta = -\frac{z}{2}$, and $A = 2M + 2MLD^{\frac{\alpha}{2}}$. The overall regret could now be balanced by

$$\begin{aligned} R(T) \leq & W_{\max} 2^D \left(\log(T) T^{\frac{2\alpha+D}{3\alpha+D}} + T^{\frac{D}{3\alpha+D}} \right) + W_{\max} \sum_{k=1}^{M} \left(\begin{array}{c} M \\ k \end{array} \right) M \frac{\pi^2}{3} \\ & + \left(3MLD^{\frac{\alpha}{2}} + \frac{(2M + 2MLD^{\frac{\alpha}{2}})(3\alpha + D)}{2\alpha + D} \right) T^{\frac{2\alpha+D}{3\alpha+D}}, \end{aligned} \tag{46}$$

with the leading order as $\mathcal{O}\left(W_{\max} 2^D \log(T) T^{\frac{2\alpha+D}{3\alpha+D}} \right)$.

□

Since Proposition 1 does not consider the mis-exploration caused by delayed update, the counter $|A_p|$ for each hypercube p may record incorrect number of exploration (less than it should be), leading to sub-optimal exploration decision. Considering the case that the prediction information for a future tile (i, j, t') being ahead of current time step t by $t' - t$, the maximum number of mis-exploration for (i, j, t') is given by $t' - t$ since the real QoE would be revealed after $t' - t$ time step. Since there exists a maximum prediction window $W_{\max}$, the maximum number of mis-exploration is bounded by $W_{\max}$ for all the hypercubes. Since the leading order of the bound for $R(T)$ without the consideration of delayed update belongs to the regret of exploration $R_e(T)$, we have the extended bound for regret $R(T)$ with delayed update as $\mathcal{O}\left(W_{\max}^2 2^D \log(T) T^{\frac{2\alpha+D}{3\alpha+D}} \right)$.

□