A. Sub-Routines used in the Algorithms

Algorithm 3 INDEX-ESTIMATION()

```
Index ← N
Arm ← 1
for $1 \leq t \leq N-1$ do
    Play Arm labeled Arm
    if Matched AND Arm == 1 then
        Index ← $t$
        Arm ← 2
    end if
end for
Return Index
```

In Algorithm 3, we give a simple algorithm by which every agent in a decentralized fashion, can estimate an unique rank. As the arms are labeled, the agents agree to the protocol, that at the beginning of the game, they will play arm labeled 1, until it gets matched for the first time. The index of the agent is the time at which it matches with arm 1. Subsequently, the agent will play arm 2 in the remaining time. Thus, the id estimated by an agent is its relative rank at arm 1, i.e. $\mathsf{Index}(j) = rank(j, 1)$ which is unique among the agents as for any arm there is no tie in its preference.

Algorithm 4 COMMUNICATION(i, $\mathcal{O}_j[i]$) for Agent j

```
Input: Phase number $i \in \{1, 2, \ldots, \}$, and max played arm $\mathcal{O}_j[i]$
for $t = 1$ to $NK - 1$ do
    if $K(\mathsf{Index}(j) - 1) \leq t \leq K\mathsf{Index}(j) - 1$ then
        Play arm $P_j(t) = (t \mod K) + 1$
        if Collision Occurs then
            $\mathcal{C} \leftarrow \mathcal{C} \cup \{P_j(t)\}$
        end if
    else
        Play arm $P_j(t) = \mathcal{O}_j[i]$
    end if
end for
Return $\mathcal{C}$
```

Algorithm 4 allows each agent j to learn the dominated arms – the arms which are most played by at least one agent $j' \neq j$ such that $j' >_{\mathcal{O}_{j'}[i]} j$. We argue that if there exists such an agent j' then $\mathcal{C} \ni \mathcal{O}_{j'}[i]$. Assume an arbitrary such agent j'. At the time $t = (K(\mathsf{Index}(j) - 1) + \mathcal{O}_{j'}[i] - 1)$ the agent j' and j plays arm $\mathcal{O}_{j'}[i]$, but agent j collides as $j' >_{\mathcal{O}_{j'}[i]} j$. This ensures $\mathcal{C} \subseteq \{\mathcal{O}_{j'}[i] : j' >_{\mathcal{O}_{j'}[i]} j\}$. Also, if there is no j' such that $j' >_{\mathcal{O}_{j'}[i]} j$, then there is no collision during $K(\mathsf{Index}(j) - 1) \leq t \leq K\mathsf{Index}(j) - 1$ when $\mathcal{C}$ is updated (there can be collision in other times). So we have $\mathcal{C} = \{\mathcal{O}_{j'}[i] : j' >_{\mathcal{O}_{j'}[i]} j\}$ which proves the correctness for the Algorithm 4.

Algorithm 5 GALE-SHAPELY at the Agents

```
Input: The preference over arms for all agents, the preference over agents for all arms
$P_j \leftarrow -1$, for all $j$. {All agents are unmatched to begin with}
while $\exists 1 \leq j \leq N$, such that $P_j = -1$, i.e., there exists an un-matched agent do
    for $1 \leq N$ in parallel do
        Agent $j$ proposes to its highest ranked arm that has not yet rejected it.
    end for
    An agent is matched to its arm if and only if it is the highest ranked agent proposing to the arm at time $t$.
end while
Return $\mathcal{C}$
```

In Algorithm 5, first introduced in (Gale & Shapley, 1962), is used as a sub-routine in Algorithm 1.

B. Additional Related Work on Multi-Agent Bandits

A popular line of work in competitive multi-agent bandits (as in this paper) is the *colliding bandits model*, where if two or more agents play the same arm in the same round, then *all* of them get blocked. Such models study spectrum sharing in wireless networks (Kalathil et al., 2014; Rosenski et al., 2016; Bistritz & Leshem, 2020), and recently sophisticated centralized and de-centralized algorithms were developed to obtain the right $O(\log(T))$ regret (Boursier & Perchet, 2019; Mehrabian et al., 2020). Our model fundamentally differs from the colliding bandits, because if multiple agents play the same arm simultaneously, one of them receives a reward while the others do not. Thus, the developments therin (Boursier & Perchet, 2019; Mehrabian et al., 2020) are inapplicable to our problem. From an application side, bandits have been used for resource allocation in networks (Darak & Hanawal, 2019; Larrnaaga et al., 2016; Avner & Mannor, 2016). This line of work does not fall under our matching bandit model and are mostly centralized systems. Another related line of work is that of collaborative multi-agent bandits, where agents are not competing for resources, but aim to maximize group reward by minimal communications (Kolla et al., 2018; Chawla et al., 2020; Sankararaman et al., 2019; Buccapatnam et al., 2015; Landgren et al., 2021; Dubey & Pentland, 2020). In a recent work, (Dai & Jordan, 2020) consider the problem of preference learning in decentralized matching markets. In their setup, arms have a noisy preference over the agents and the goal is to learn this preference via repeated interaction. The authors provide an algorithm that asymptotically converges to a matching that is stable and fair. Although related, the model and the goal of the problem is significantly different from that considered in this paper.

C. Proof of Theorem 1

Proof of Theorem 1. We use the Hoeffding's bound and linearity of expectation to establish this result. We collect some useful observations from the description of the algorithm.

- The total number of phases until time T is at-most $\lceil \log_2(T) \rceil + 1$.
- The total number of explore samples by any agent after the explore part of phase i is at-least $\frac{(i-1)^{1+\varepsilon}}{1+\varepsilon} - i$.
- Each agent experiences no more than N^2 collisions in a phase.

The first point follows from the fact that phase i lasts for 2^i rounds. The second point follows from the fact that in any phase i. the first i^ε rounds are used for exploration. The last point follows from classical results on Gale-Shapley matching, which takes at-most N^2 rounds.

In any phase i, denote by the event $\mathcal{E}_i$ to be the one in which, every agent correctly estimated the ordering of its arms, at the end of the explore portion of phase i. The following two propositions help us prove the regret bound

Proposition 1.

$$R_T^{(j)} \leq \sum_{i=0}^{\lceil \log_2(T) \rceil + 1} 2^i \mathbb{P}(\mathcal{E}_i^\complement) + K \frac{(\log_2(T)+2)^{1+\varepsilon}}{1+\varepsilon} + (N^2+K)(\log_2(T)+2).$$

Proposition 2.

$$\mathbb{P}[\mathcal{E}_i^\complement] \leq 2NK \exp\left(- \left(\frac{(i-1)^{1+\varepsilon}}{1+\varepsilon} - i \right) \frac{\Delta^2}{2} \right). \tag{1}$$

Before giving the proofs of these propositions, we show how these aid in bounding the regret. Denote by i_Δ as

$$i_\Delta := \min \left\{ u : \forall i \geq u, \left(\frac{(i-1)^{1+\varepsilon}}{1+\varepsilon} - i \right) \frac{\Delta^2}{4} \geq i \right\}.$$

It is easy to verify that

$$i_\Delta \leq \left(\frac{8}{\Delta^2} \right)^{\frac{1}{\varepsilon}} 4^{\frac{1+\varepsilon}{\varepsilon}}. \tag{2}$$

Now, from Propositions 1 and 2, we get the following

$$R_T^{(j)} \le 2NK \sum_{i=0}^{\lceil \log_2(T) \rceil + 1} 2^i e^{-2\left(\frac{(i-1)^{1+\varepsilon}}{1+\varepsilon} - i\right)\frac{\Delta^2}{4}} + K\frac{(\log_2(T)+2)^{1+\varepsilon}}{1+\varepsilon} + (N^2+K)(\log_2(T)+2). \tag{3}$$

From the definition of i_Δ, we have

$$\begin{aligned} \sum_{i=0}^{\lceil \log_2(T) \rceil + 1} 2^i e^{-2\left(\frac{(i-1)^{1+\varepsilon}}{1+\varepsilon} - i\right)\frac{\Delta^2}{4}} &\le \sum_{i=0}^{i_\Delta} 2^i + \sum_{i \ge 0} \left(\frac{2}{e}\right)^i, \\ &\le 2^{i_\Delta + 1} + \frac{e}{e-2}. \end{aligned} \tag{4}$$

Substituting Equation (4) into Equation (5), and bounding i_Δ with Equation (2), yields the result. □

Proof of Proposition 1.

$$\begin{aligned} R_T^{(j)} &= \sum_{t=1}^{T} \mathbb{P}[I_j(t) \neq k_j^*], \\ &\le \sum_{i=0}^{\lceil \log_2(T) \rceil + 1} \sum_{n=1}^{2^i} \mathbb{P}[I_j(\max(T, n + 2^i - 1) \neq k_j^*], \\ &\overset{(a)}{\le} \sum_{i=0}^{\lceil \log_2(T) \rceil + 1} \left(\sum_{n=1}^{K\lfloor i^\varepsilon \rfloor + N^2} 1 + \sum_{n=K\lfloor i^\varepsilon \rfloor + N^2 + 1}^{2^i} \mathbb{P}[\mathcal{E}_i^{\complement}] \right), \\ &\le \sum_{i=0}^{\lceil \log_2(T) \rceil + 1} (K(i^\varepsilon + 1) + N^2) + \sum_{i=0}^{\lceil \log_2(T) \rceil + 1} 2^i \mathbb{P}[\mathcal{E}_i^{\complement}], \\ &\le \int_{i=0}^{\lceil \log_2(T) \rceil + 1} (Ki^\varepsilon + K + N^2) + \sum_{i=0}^{\lceil \log_2(T) \rceil + 1} 2^i \mathbb{P}[\mathcal{E}_i^{\complement}], \\ &\le K\frac{(\log_2(T)+2)^{1+\varepsilon}}{1+\varepsilon} + (N^2+K)(\log_2(T)+2) + \sum_{i=0}^{\lceil \log_2(T) \rceil + 1} 2^i \mathbb{P}(\mathcal{E}_i^{\complement}). \end{aligned}$$

In step (a), we use the Gale Shapley property, that if all agents correctly estimate their arm-ranking, then all agents will play the agent-optimal stable match arm in the exploit part of the phase (after accounting for the at-most N^2 rounds needed for the Gale-Shapley matching to converge). □

Proof of Proposition 2. A sufficient condition for the event $\mathcal{E}_i$ to hold, is if all agents, learn of all their respective arm-means, to a resolution of within $\Delta/2$. This will automatically imply that the empirical rankings of all agents will be identical to the truth. For notational simplicity, denote by $\widetilde{\mu}_j k^{(i)}$ to be the empirical mean of arm k, computed by agent j, using all the explore samples upto and including phase i. We have from definition of $\mathcal{E}_i$,

$$\mathcal{E}_i \supseteq \bigcap_{j=1}^{N} \bigcap_{k=1}^{K} \left\{ |\widetilde{\mu}_{jk}^{(i)} - \mu_{jk}| < \frac{\Delta}{2} \right\}.$$

Thus, we have from the union bound that

$$\mathbb{P}[\mathcal{E}_i^{\complement}] \le \sum_{j=1}^{N} \sum_{k=1}^{K} \mathbb{P}\left[|\widetilde{\mu}_{jk}^{(i)} - \mu_{jk}| \ge \frac{\Delta}{2} \right]. \tag{5}$$

Since $\widetilde{\mu}_{jk}^{(i)}$ is the empirical mean estimated using at-least $\frac{(i-1)^{1+\varepsilon}}{1+\varepsilon} - i$ i.i.d. samples, we have from Hoeffding's bound

$$\mathbb{P}\left[|\widetilde{\mu}_{jk}^{(i)} - \mu_{jk}| \ge \frac{\Delta}{2} \right] \le 2\exp\left(-2\left(\frac{(i-1)^{1+\varepsilon}}{1+\varepsilon} - i \right) \frac{\Delta^2}{4} \right). \tag{6}$$

The result follows by substituting Equation (6) into Equation (5). □

D. Proof of Regret Upper Bound Under SPC Condition

D.1. Notation and Definition:

We next set up the notations required for the proof of the main results. We denote by $\mathbb{N}$ the set of natural numbers and by $\mathbb{R}_+$ the set of non-negative real numbers.

Ranks: We define by $>_k$ the preference order (a.k.a. rank) of arm k, for any arm $k \in [K]$, where if $j >_k j'$ then arm k prefers agent j over agent $j^{'}$. Recall, under the SPC condition we assume the common order (among agents and arms) is identity without loss of generality, for the ease of exposition.

Phases: Our algorithm works in phases. By S_i we denote the starting round for phase i. We have $S_1 = R + 1$ and for $i > 1$, $S_i = R + \sum_{i'=1}^{i-1}(C + 2^{i'-1}) = R + C(i-1) + 2^i$, where R is the time required for the Ranking period which runs once in the beginning, and C for the communication phase which is used each phase once.

Arm Classification: For each agent j, let the set of *dominated arms* be $\mathcal{D}_j := \{k^*_{j'} : j' = 1 \ldots, j-1\}$ the stable matching arm of the agents ranked higher than j in the SPC order. Further, for each arm $k \notin \mathcal{D}_j$, we also define the *blocking agents* for arm k and agent j as $\mathcal{B}_{jk} = \{j' : j' >_k j\}$, the set of agents preferred by arm k over agent j. We define the arms as *hidden arms* $\mathcal{H}_j := \{k : k \notin D_j, \mathcal{B}_{jk} \neq \emptyset\}$.

Gaps: Let the stable matching pair of agent j be arm k^*_j for any $j \in [N]$. Let $\Delta_{jk} = \mu_{jk} - \mu_{jk^*_j}$ be the gap for arm $k \in [K]$. Let $\Delta_{\min} = \min_{jk}\{\Delta_{jk} : k \notin \mathcal{D}_j \cup k^*_j\}$. Recall our assumption that, for every agent, no two arm means are the same implies that $\Delta_{min} > 0$, is strictly greater than 0.

Number of Plays and Attempts: For each agent $j \in [N]$ and arm $k \in [K]$, denote by $N_{jk}(t)$ as the number of times agent j has successfully matched with arm k (i.e., without colliding) up to time t for any t. For all $i \geq 1$, we denote by $N_{jk}[i] = N_{jk}(S_{i+1} - 1)$ to be the same quantity as above at the end of phase i. We denote by $I_j(t) \in [K] \cup \emptyset$ as the arm sampled by agent j on round t (here $I_j(t) = \emptyset$ denotes the agent j collides in time t). Let $G_j[i]$ denote the set of globally deleted arms for agent j at the beginning of phase i, and $L_j[i]$ denote the set of locally deleted arms for agent j at the end of phase i. We note that N_{jk}, $I_j(t-1)$, $G_j[i]$, and $L_j[i]$ all are random variables adapted to the filtration constructed by the history of agent j at different points in time.

Critical Phases:

- The phase i for agent j, for some $j \in [N]$, is a *Good Phase* if the following are true:

  1. The dominated arms are globally deleted, i.e. $G_j[i] = \mathcal{D}_j$.
  2. For each arm $k \notin \mathcal{D}_j \cup k^*_j$ (not globally deleted), in phase i arm k is successfully played (a.k.a. sampled) by agent j at most $\frac{10\gamma i}{\Delta^2_{jk}}$ times.
  3. The stable match pair arm k^*_j is sampled the most number of times in phase i.

  The *good phase* definition is identical to the definition in Sankararaman et al. (Sankararaman et al., 2021).

- We further define a phase i for agent j, for some $j \in [N]$ to be a *Low Collision Phase* if the following are true:

  1. Phase i is a good phase for agent 1 to j.
  2. Phase i is a good phase for all agent $j' \in \cup_{k \in \mathcal{H}_j} \mathcal{B}_{jk}$.

For notational ease, let $\mathbb{I}_G[i,j]$ be the indicator that phase i is a good phase for agent j. Similarly, let $\mathbb{I}_{LC}[i,j]$ be the indicator that phase i is a low collision phase for agent j.

Let $i_1 = \left((N-1)\frac{10\gamma}{\Delta^2_{min}}\right)^{\frac{1}{\beta-1}} + 1$. We define the *Freezing* phase for each agent j as the phase on and after which all phases are good phases for agents 1 to $(j-1)$.

$$F_j = \max\left(i_1, \min\left(\{i : \prod_{j'=1}^{(j-1)} \prod_{i' \geq i} \mathbb{I}_G[i',j'] = 1\} \cup \{\infty\}\right)\right).$$

Further, the *Vanishing* phase for each agent j as the phase on and after which all phases are low collision phases for agent j

$$V_j = \max\left(i_1, \min\left(\{i : \prod_{i' \geq i} \mathbb{I}_{LC}[i', j] = 1\} \cup \{\infty\}\right)\right).$$

It is easy to see, that $F_j = \max\left(F_{j'} : 1 \leq j' \leq (j-1)\right)$ and $V_j = \max\left(F_{j+1}, \cup_{k \in \mathcal{H}_j} \cup_{j' \in \mathcal{B}_{jk}} F_{j'}\right)$ from the definition of low collision phase.

D.2. Proof of main result

We begin the proof with the simple result that arm j and agent j form a stable match pair for any $j \leq N$

Proposition 3. *If a system satisfies SPC then $k_j^* = j$ for all $j \in [N]$. Furthermore, if a system satisfies α-condition then we have $k_j^* = j$ for all $1 \leq j \leq N$, and $j_{a_k}^* = A_k$ and $k_{A_j}^* = a_j$ for all $1 \leq k, j \leq N$.*

Proof. We note for the SPC condition, for $j = 1$ we have $\mu_{1k_1^*} > \mu_{1k}$ for all $k > 1$ implying $k_1^* = 1$. For $j = 2$ we see $\mu_{2k_2^*} > \mu_{2k}$ for all $k > 2$. So $k_2^* \in \{1, 2\}$. But $k_1^* = 1$, thus $k_2^* = 2$. This logic can be extended to prove that for all $1 \leq j \leq N$ $k_j^* = j$. □

We now prove that under low collision phase there is no local deletion for agent j in the following lemma.

Lemma 1. *If a phase $i \geq i_1 = \min\{i : (N-1)\frac{10\gamma i}{\Delta_{min}^2} < \beta 2^{(i-1)}\}$, is a Low Collision phase for agent j, for any $j \in [N]$, then $L_j[i] = \emptyset$.*

Proof. As phase i is Low Collision we know phase i is good for the agent j. Therefore, the arms $k \in \mathcal{D}_j$ are deleted in the beginning of the phase and no collision is encountered. So $\forall k \in \mathcal{D}_j, k \notin L_j[i]$.

Further, as phase i is a good phase for all agent $j' \in \cup_{k \in \mathcal{H}_j} \mathcal{B}_{jk}$ (by definition), for each $k \in \mathcal{H}_j$ the maximum number of collision $(N-1)\frac{10\gamma i}{\Delta_{min}^2}$. This is true as in a good phase any agent $j' \in \mathcal{B}_{jk}$ plays k at most $\frac{10\gamma i}{\Delta_{min}^2}$ times as $k \notin \mathcal{D}_{j'} \cup k_{j'}^*$. Therefore, for $i \geq i_1$ we have local deletion threshold $\beta 2^{(i-1)} > (N-1)\frac{10\gamma i}{\Delta_{min}^2}$. Thus, $\forall k \in \mathcal{H}_j, k \notin L_j[i]$.

Finally, for each arm $k \notin \mathcal{H}_j \cup \mathcal{D}_j$, we have for all agent $j' \neq j, j' >_k j$. Therefore, $\forall k \notin \mathcal{H}_j \cup \mathcal{D}_j$, there is no collision experienced by agent j. Therefore, $k \notin L_j[i]$ and we conclude that $L_j[i] = \emptyset$. □

Lemma 2. *If a phase $i \geq i_1$ (i_1 as defined in Lemma 1), is a good phase for all agents 1 to $(j-1)$, for any $j \in [N]$, then $k_j^* \notin L_j[i]$.*

Proof. Due to SPC we have $j >_{k_j^*} j'$ for all $j' > j$. Therefore, the arm k_j^* can be locally deleted ($k_j^* \in L_j[i]$) in phase i only if the total collisions from agents 1 to $(j-1)$ is greater than i^β. But the total collisions from agents 1 to $(j-1)$ is at most $\frac{10(j-1)\gamma i}{\Delta_{min}^2}$ as phase i is a good phase for all agents 1 to $(j-1)$. Also, $i^\beta > (N-1)\frac{10\gamma i}{\Delta_{min}^2}$ for all $i \geq i_1$. Therefore, k_j^* can not be locally deleted in the phase i as mentioned in the lemma. □

We now decompose the regret and provide a regret upper bound.

Lemma 3. *The expected regret for agent j can be upper bounded as*

$$\begin{aligned}\mathbb{E}[R_j(T)] \leq\, & \mathbb{E}[S_{F_j}] + \min(1, \beta|\mathcal{H}_j|)\mathbb{E}[S_{V_j}] + \left((K - 1 + |\mathcal{B}_{jk_j^*}|)\log_2(T) + NK\mathbb{E}[V_j]\right) \\ & + \sum_{k \notin \mathcal{D}_j} \sum_{j' \in \mathcal{B}_{jk} : k \notin \mathcal{D}_{j'}} \frac{8\gamma\mu_{k_j^*}}{\Delta_{j'k}^2}\left(\log(T) + \sqrt{\tfrac{\pi}{\gamma}\log(T)}\right) + \sum_{k \notin \mathcal{D}_j \cup k_j^*} \frac{8\gamma}{\Delta_{jk}}\left(\log(T) + \sqrt{\tfrac{\pi}{\gamma}\log(T)}\right) \\ & + NK\left(1 + (\psi(\gamma) + 1)\frac{8\gamma}{\Delta_{\min}^2}\right)\end{aligned}$$

Proof. We now decompose the expected regret as follows

$$
\begin{aligned}
&\mathbb{E}[R_j(T)]\\
&\le \mathbb{E}[S_{F_j}] + \mathbb{E}\left[\sum_{i=(F_j+1)}^{V_j}\sum_{k\notin\mathcal{D}_j}\beta 2^{i-1}\right] + \big(c_j\log_2(T) + NK\mathbb{E}[F_{(j+1)}]\big)\\
&+\mathbb{E}\left[\sum_{k\notin\mathcal{D}_j}\sum_{j'\in\mathcal{B}_{jk}:k\notin\mathcal{D}_{j'}}\mu_{k_j^*}(N_{j'k}(T)-N_{j'k}(S_{V_j}))\right]\\
&+\mathbb{E}\left[\sum_{k\notin\mathcal{D}_j\cup k_j^*}\Delta_{jk}(N_{jk}(T)-N_{jk}(S_{F_j}))\right]\\
&\le \mathbb{E}[S_{F_j}] + \min(1,\beta|\mathcal{H}_j|)\mathbb{E}\left[S_{V_j}\right] + \big(c_j\log_2(T) + NK\mathbb{E}[F_{(j+1)}]\big)\\
&+\sum_{k\notin\mathcal{D}_j}\sum_{j'\in\mathcal{B}_{jk}:k\notin\mathcal{D}_{j'}}\mu_{k_j^*}\mathbb{E}\left[(N_{j'k}(T)-N_{j'k}(S_{V_j}))\right]\\
&+\sum_{k\notin\mathcal{D}_j\cup k_j^*}\Delta_{jk}\mathbb{E}\left[(N_{jk}(T)-N_{jk}(S_{F_j}))\right]\\
&\le \mathbb{E}[S_{F_j}] + \min(1,\beta|\mathcal{H}_j|)\mathbb{E}\left[S_{V_j}\right] + \big(c_j\log_2(T) + NK\mathbb{E}[F_{(j+1)}]\big)\\
&+\sum_{k\notin\mathcal{D}_j}\sum_{j'\in\mathcal{B}_{jk}:k\notin\mathcal{D}_{j'}}\mu_{k_j^*}\left(\psi(\gamma)\tfrac{8}{\Delta_{j'k}^2}+1+\tfrac{8}{\Delta_{j'k}^2}\left(\gamma\log(T)+\sqrt{\pi\gamma\log(T)}+1\right)\right)\\
&+\sum_{k\notin\mathcal{D}_j\cup k_j^*}\Delta_{jk}\left(\psi(\gamma)\tfrac{8}{\Delta_{jk}^2}+1+\tfrac{8}{\Delta_{jk}^2}\left(\gamma\log(T)+\sqrt{\pi\gamma\log(T)}+1\right)\right).
\end{aligned}
$$

In the first inequality, follows due to the following reasons.

- We upper bound the regret till the end of phase F_j by S_{F_j} as regret per round is at most $\mu_{k_j^*}\le 1$.
- **Local deletion:** Next from phase (F_j+1) upto phase V_j (both inclusive), we upper bound the regret due to collision by $\sum_{i=(F_j+1)}^{V_j}\sum_{k\in\mathcal{H}_j}\beta 2^{(i-1)}$ as in each round i at most $\beta 2^{(i-1)}$ collisions are possible when pulling an arm from the set $|\mathcal{H}_j|$ in phase (F_j+1) to V_j. This is true as all the arms in $\mathcal{D}_j$ are globally deleted from phase (F_j+1) onwards.
- **Communication phase:** The best arm for agent j is not played in all but $(K-1)$ number of steps for each communication phase after phase F_{j+1}, and other agents $j'\in\mathcal{B}_{jk_j^*}$ collide at most once after phase V_j (as each of them enter good phase). Thus, beyond phase V_j we have at most $(K-1+|\mathcal{B}_{jk_j^*}|)$ regret due to collision, and there are at most $\log_2(T)$ communication phases. This limits the regret due to communicationat $\Big((K-1+|\mathcal{B}_{jk_j^*}|)\log_2(T)+NKV_j\Big)$.
- **Collision:** From phase (V_j+1) (inclusive) onwards only an agent $j'\in\mathcal{B}_{jk}$ collides with k only if $k\notin\mathcal{D}_{j'}$, because (1) agent j' deletes all arms in $\mathcal{D}_{j'}$ from (V_j+1) (inclusive) onwards, and (2) all $j'\notin\mathcal{B}_{jk}$ and $j'\neq j$, $j>_k j'$ (agent j is preferred by k over agent j'). This amounts to $\sum_{k\notin\mathcal{D}_j}\sum_{j'\in\mathcal{B}_{jk}:k\notin\mathcal{D}_{j'}}\mu_{k_j^*}(N_{j'k}(T)-N_{j'k}(S_{V_j}))$ regret.
- **Suboptimal play:** Finally, from phase (F_j+1) (inclusive) onwards till the last phase agent j incurs regret Δ_{jk} each time the agent j successfully plays arm $k\notin\mathcal{D}_j\cup k_j^*$. Thus she incurs total $\Delta_{jk}(N_{jk}(T)-N_{jk}(S_{F_j}))$ regret for each such k.

The validity of the second inequality is easy to see. We now come to the last inequality. We know that $(N_{j'k}(T)-N_{j'k}(S_{V_j}))\le(N_{j'k}(T)-N_{j'k}(S_{F_{j'}}))$ (almost surely) as $V_j\ge F_{j'}$ almost surely from the definition of V_j, for all $j'\in\mathcal{B}_{jk}$. Thus, the final inequality follows by substituting the bounds from Lemma 4. □

We now prove the upper bound on the expected number of times a sub-optimal arm is played by an agent j after the Global deletion freezes.

Lemma 4. *For any $j\in[N]$, $k\notin\mathcal{D}_j\cup k_j^*$, for $\gamma>1$,*

$$\mathbb{E}\left[(N_{jk}(T)-N_{jk}(S_{F_j}))\right]\le\psi(\gamma)\tfrac{8}{\Delta_{jk}^2}+1+\tfrac{8}{\Delta_{jk}^2}\left(\gamma\log(T)+\sqrt{\pi\gamma\log(T)}+1\right).$$

Proof. We have for any $k \notin \mathcal{D}_j \cup k_j^*$ and $\epsilon > 0$

$$\begin{aligned}
(N_{jk}(T) - N_{jk}(S_{F_j})) &= \sum_{t=S_{F_j}+1}^{T} \mathbb{I}(I_j(t) = k) \\
&\leq \sum_{t=S_{F_j}+1}^{T} \Big(\mathbb{I}(u_{jk}(t-1) \geq \mu_{jk_j^*} - \epsilon \wedge I_j(t) = k) + \mathbb{I}(u_{jk_j^*}(t-1) \leq \mu_{jk_j^*} - \epsilon) \Big) \qquad (7)
\end{aligned}$$

The inequality is true because phase $(F_j + 1)$ onwards (inclusive) the arm k_j^* is neither globally deleted (by definition of F_j) or locally deleted as shown in Lemma 2. Therefore, we have

$$\{I_j(t) = k \wedge t > S_{F_j}\} \subseteq \{I_j(t) = k \wedge u_{jk}(t-1) \geq \mu_{jk_j^*} - \epsilon \wedge t > S_{F_j}\} \cup \{u_{jk_j^*}(t-1) \leq \mu_{jk_j^*} - \epsilon \wedge t > S_{F_j}\}.$$

Note, before the phase $(F_j + 1)$ it is not true that any arm k' better than k, in particular arm k_j^*, survives the global and local deletion. The rest of the proof of this lemma is fairly standard. However, we present it for completeness.

We next bound the expectation of the second term in a standard way as follows (c.f. (Lattimore & Szepesvári, 2020) Theorem 8.1 proof)

$$\begin{aligned}
&\mathbb{E}[\sum_{t=S_{F_j}+1}^{T} \mathbb{I}(u_{jk_j^*}(t-1) \leq \mu_{jk_j^*} - \epsilon)] \\
&= \mathbb{E}[\sum_{t=1}^{T} \mathbb{I}(u_{jk_j^*}(t-1) \leq \mu_{jk_j^*} - \epsilon \wedge t > S_{F_j})] \\
&\leq \mathbb{E}[\sum_{t=1}^{T} \mathbb{I}(u_{jk_j^*}(t-1) \leq \mu_{jk_j^*} - \epsilon)] \\
&\leq \sum_{t=1}^{T} \sum_{s=1}^{T} \mathbb{P}(\hat{\mu}_{jk_j^*,s} + \sqrt{\tfrac{2\gamma \log(t)}{s}} \leq \mu_{jk_j^*} - \epsilon) \\
&\leq \sum_{t=1}^{T} \sum_{s=1}^{T} \exp\left(-\tfrac{s}{2} (\sqrt{\tfrac{2\gamma \log(t)}{s}} + \epsilon)^2 \right) \\
&\leq \sum_{t=1}^{T} t^{-\gamma} \sum_{s=1}^{T} \exp(-\tfrac{s\epsilon^2}{2}) \leq \psi(\gamma) \tfrac{2}{\epsilon^2}
\end{aligned}$$

Here, $\psi()$ is the Riemann zeta function. Note, the first inequality is valid as

$$\mathbb{I}(u_{jk_j^*}(t-1) \geq \mu_{jk_j^*} - \epsilon \wedge t > S_{F_j}) \leq \mathbb{I}(u_{jk_j^*}(t-1) \geq \mu_{jk_j^*} - \epsilon) \text{ a.s.}$$

Finally, we bound the expectation of the first term also in a standard way ((c.f. (Lattimore & Szepesvári, 2020) Lemma 8.2))

$$\begin{aligned}
&\mathbb{E}\left[\sum_{t=S_{F_j}+1}^{T} \mathbb{I}(u_{jk}(t-1) \geq \mu_{jk_j^*} - \epsilon \wedge I_j(t) = k) \right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}(\hat{\mu}_{jk}(t-1) + \sqrt{\tfrac{2\gamma \log(t)}{N_{jk}(t-1)}} \geq \mu_{jk_j^*} - \epsilon \wedge I_j(t) = k) \right] \\
&\leq \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{I}(\hat{\mu}_{jk}(t-1) + \sqrt{\tfrac{2\gamma \log(T)}{N_{jk}(t-1)}} \geq \mu_{jk_j^*} - \epsilon \wedge I_j(t) = k) \right] \\
&\leq \mathbb{E}\left[\sum_{s=1}^{T} \mathbb{I}(\hat{\mu}_{jk,s} + \sqrt{\tfrac{2\gamma \log(T)}{s}} \geq \mu_{jk} + \Delta_{jk} - \epsilon) \right]
\end{aligned}$$

$$\leq 1 + \frac{2}{(\Delta_{jk} - \epsilon)^2}\left(\gamma \log(T) + \sqrt{\pi\gamma \log(T)} + 1\right)$$

We combine the two above bounds and pick $\epsilon = \Delta_{jk}/2$ (for simplicity, this can be tightened with some effort) we obtain the following bound for $\gamma > 1$,

$$\mathbb{E}\left[(N_{jk}(T) - N_{jk}(S_{F_j}))\right] \leq \psi(\gamma)\tfrac{8}{\Delta_{jk}^2} + 1 + \tfrac{8}{\Delta_{jk}^2}\left(\gamma \log(T) + \sqrt{\pi\gamma \log(T)} + 1\right).$$

□

We now have to provide an upper bound on the moments of V_j and mean of S_{F_j} to complete the proof of the regret bound. As V_j is a function of $F_{j'}$ for $j' \in [N]$ we need to derive bounds for moments and exponent of $F_{j'}$ for all j'. The key idea is to show that once the Global deletion has settled for agents 1 to $(j-1)$ (recall the agents are ordered according to the SPC order) the agent j enters Good phase with high probability.

Lemma 5. *For any agent j and any phase $i \geq i^* = \max\{8, i_1, i_2\}$ and $\gamma > 1$,*

$$\mathbb{P}[\mathbb{I}_G[i,j] = 0 \wedge i \geq F_j + 1] \leq (K - 1 - |\mathcal{D}_j|)\left(1 + \tfrac{64}{\Delta_{\min}^2}\right)2^{-i(\gamma-1)},$$

where $i_1 = \min\{i : (N-1)\frac{10\gamma i}{\Delta_{min}^2} < \beta 2^{(i-1)}\}$ and $i_2 = \min\{i : (R - 1 + C(i-1)) \leq 2^{i+1}\}$.

Proof. Let us recall that the phase i is a Good phase for agent j if and only if (1) the dominated arms $\mathcal{D}_j$ are deleted in global deletion, and (2) each arm $k \notin \mathcal{D}_j \cup k_j^*$ is sampled by agent j at most $\frac{10\gamma i}{\Delta_{jk}^2}$ times, and (3) arm k_j^* is matched the most number of times.

Given $\{i \geq (F_j + 1)\}$ in phase i condition (1) is satisfied for any $i \geq 1$. For $i \geq i_1$, as in Lemma 1, we can show that condition (1) and (2) implies condition (3) holds true. So we need to bound the probability that given $\{i \geq F_j + 1\}$ the condition (2) holds. This follows from the properties of UCB as we show below. We have for any j and $\epsilon > 0$,

$$\begin{aligned}
&\mathbb{P}[\mathbb{I}_G[i,j] = 0 \wedge i \geq (F_j + 1)]\\
&\leq \mathbb{P}\left[\cup_{k \notin \mathcal{D}_j \cup k_j^*}\{(N_{jk}[i] - N_{jk}[i-1]) > \tfrac{10\gamma i}{\Delta_{jk}^2}\} \wedge i \geq (F_j+1)\right]\\
&\leq \sum_{k \notin \mathcal{D}_j \cup k_j^*} \mathbb{P}\left[\cup_{t \in S_i}^{(S_{i+1}-1)} N_{jk}(t) = \tfrac{10\gamma i}{\Delta_{jk}^2} \wedge I_j(t) = k \wedge i \geq (F_j+1)\right]\\
&\overset{(i)}{\leq} \sum_{k \notin \mathcal{D}_j \cup k_j^*} \sum_{t \in S_i}^{(S_{i+1}-1)} \mathbb{P}\left[N_{jk}(t) = \tfrac{10\gamma i}{\Delta_{jk}^2} \wedge u_{jk}(t-1) > u_{jk_j^*}(t-1)\right]\\
&\overset{(ii)}{\leq} \sum_{k \notin \mathcal{D}_j \cup k_j^*} \sum_{t \in S_i}^{(S_{i+1}-1)} \mathbb{P}\left[\{N_{jk}(t) = \tfrac{10\gamma i}{\Delta_{jk}^2} \wedge u_{jk}(t-1) \geq \mu_{jk_j^*} - \epsilon\} \cup \{u_{jk_j^*}(t-1) \leq \mu_{jk_j^*} - \epsilon\}\right]\\
&\overset{(iii)}{\leq} \sum_{k \notin \mathcal{D}_j \cup k_j^*} \sum_{t \in S_i}^{(S_{i+1}-1)} \mathbb{P}\left[N_{jk}(t) = \tfrac{10\gamma i}{\Delta_{jk}^2} \wedge \hat{\mu}_{jk}(t-1) + \Delta_{jk}\sqrt{\tfrac{2\gamma \log(t)}{10\gamma i}} \geq \mu_{jk} + \Delta_{jk} - \epsilon\right] +\\
&+ \sum_{k \notin \mathcal{D}_j \cup k_j^*} \sum_{t \in S_i}^{(S_{i+1}-1)} \sum_{s=1}^{t-1} \mathbb{P}\left[\hat{\mu}_{jk_j^*,s} + \sqrt{\tfrac{2\gamma \log(t)}{s}} \leq \mu_{jk_j^*} - \epsilon\right]\\
&\overset{(iv)}{\leq} \sum_{k \notin \mathcal{D}_j \cup k_j^*} \sum_{t \in S_i}^{(S_{i+1}-1)} \left(\exp\left(-\tfrac{5\gamma i}{\Delta_{jk}^2}(\tfrac{1}{2}\Delta_{jk} - \epsilon)^2\right) + \sum_{s=1}^{t-1}\exp\left(-\tfrac{s}{2}(\sqrt{\tfrac{2\gamma \log(t)}{s}} + \epsilon)^2\right)\right)\\
&\overset{(v)}{\leq} \sum_{k \notin \mathcal{D}_j \cup k_j^*} (S_{i+1} - S_i)2^{-i\gamma}\left(1 + \tfrac{64}{\Delta_{jk}^2}\right)
\end{aligned}$$

$$\overset{(vi)}{\leq} (K-j)2^{-i(\gamma-1)}\left(1+\tfrac{64}{\Delta_{\min}^2}\right)$$

Inequality (i) relates the event of playing arm k to the UCB bounds along with the fact that k_j^* is present after phase $i \geq (F_j+1)$. The inequalities (ii) and (iii) follow similar logic as in Lemma 4. Here for inequality (iv) we use large enough i such that $(1-\sqrt{\frac{\log(S_{i+1}-1)}{5i}}) \geq 1/2$, and for (v) we use small enough ϵ such that $\frac{5}{\log(2)}(\frac{1}{2}-\frac{\epsilon}{\Delta_{jk}})^2 \geq 1$. We also use the fact $S_i \geq 2^i$ for (v). The above are satisfied when $\epsilon = \Delta_{jk}/8$ and for all $i \geq \max\{8, i_2\}$. The latter is true because for $i_2 = \min\{i : (R-1+C(i-1)) \leq 2^{i+1}\}$ we have $\log(S_{i+1}-1) \leq i+2$, and for $i \geq 8$, $1-\sqrt{\frac{i+2}{5i}} \geq 1/2$. Finally, (vi) simply uses minimum gap over all arms and agents (for simplicity) and $|\mathcal{D}_j| = (j-1)$. □

To complete the proof we need to upper bound of the moments, and exponents of F_j in an inductive manner similar to Sankararaman et al. (Sankararaman et al., 2021).

Lemma 6. *For any $j \in [N]$ and $m \geq 1$, the following hold with i^* as defined in Lemma 5*

$$\mathbb{E}[F_j^m] \leq i_1 + (j-1)(i^*)^m + (j-1)(K-j/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}$$

$$\mathbb{E}[2^{F_j}] \leq i_1 + (j-1)2^{i^*} + (j-1)(K-j/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}.$$

Proof. Let $g : \mathbb{R} \to \mathbb{R}_+$ be any monotonically increasing and continuous (hence invertible) function. We have that $F_0 = i_1$ almost surely by definition (this accounts for the max with i_1 in the definitino of F_j). The inductive hypothesis is

$$\mathbb{E}[g(F_j)] \leq i_1 + (j-1)g(i^*) + (j-1)(K-j/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}.$$

We calculate the expectation for agent j as

$$\begin{aligned}
\mathbb{E}[g(F_j)] &= \sum_{x\geq 0}\mathbb{P}[g(F_j)\geq x] \\
&= \sum_{x\geq 0}\mathbb{P}[F_j \geq g^{-1}(x)] \\
&= \sum_{x\geq 0}\left(\mathbb{P}[F_j \geq g^{-1}(x), F_{j-1}\geq g^{-1}(x)] + \mathbb{P}[F_j \geq g^{-1}(x), F_{j-1} < g^{-1}(x)]\right) \\
&\leq \sum_{x\geq 0}\mathbb{P}[F_{j-1}\geq g^{-1}(x)] + \sum_{x\geq 0}\mathbb{P}[F_j \geq g^{-1}(x), F_{j-1} < g^{-1}(x)] \\
&\leq \mathbb{E}[g(F_{j-1})] + \sum_{x=0}^{g(i^*)-1} 1 + \sum_{x\geq g(i^*)}\mathbb{P}[F_j \geq g^{-1}(x), F_{j-1} < g^{-1}(x)] \\
&\leq \mathbb{E}[g(F_{j-1})] + g(i^*) + \sum_{i\geq i^*}\mathbb{P}[F_j \geq i, F_{j-1} < i] \\
&\leq \mathbb{E}[g(F_{j-1})] + g(i^*) + \sum_{i\geq i^*}\mathbb{P}[\{\exists i' \geq i, \mathbb{I}_G[i',j]=0\}, F_{j-1}+1 \leq i] \\
&\leq \mathbb{E}[g(F_{j-1})] + g(i^*) + \sum_{i\geq i^*}\sum_{i'\geq i}\mathbb{P}[\mathbb{I}_G[i',j]=0, F_{j-1}+1 \leq i'] \\
&\leq \mathbb{E}[g(F_{j-1})] + g(i^*) + \sum_{i'\geq i^*}(i'-i^*+1)\mathbb{P}[\mathbb{I}_G[i',j]=0, F_{j-1}+1 \leq i'] \\
&\overset{(i)}{\leq} \mathbb{E}[g(F_{j-1})] + g(i^*) + (K-j)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\sum_{i'\geq i^*}(i'-i^*+1)2^{-i'(\gamma-1)} \\
&\leq (j-2)g(i^*) + (j-2)(K-j/2+1/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2} \\
&+ g(i^*) + (K-j)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}
\end{aligned}$$

$$\leq i_1 + (j-1)g(i^*) + (j-1)(K-j/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}.$$

The inequality (i) follows due to Lemma 5, while the rest are standard. □

To finalize the regret upper bound proof we note that the following holds. For the expected rounds upto the end of phase F_j is upper bounded as

$$\begin{aligned}
&\mathbb{E}[S_{F_j}] = \mathbb{E}[R + C(F_j - 1) + 2^{F_j}]\\
&\leq R + C(i_1 + (j-1)i^* + (j-1)(K-j/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2} - 1)\\
&+ (j-1)2^{i^*} + (j-1)(K-j/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}\\
&= R + C(i_1-1) + C(j-1)i^* + (j-1)2^{i^*} + (C+1)(j-1)(K-j/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}
\end{aligned}$$

Let us define $J_{\max}(j) = \max\left(j+1, \{j' : \exists k \in \mathcal{H}_j, j' \in \mathcal{B}_{jk}\}\right)$. Then as $F_j \geq F_{j'}$ almost surely for all $j \geq j'$ by definition, we have

$$V_j = \max\left(F_{j+1}, \cup_{k\in\mathcal{H}_j} \cup_{j'\in\mathcal{B}_{jk}} F_{j'}\right) = F_{J_{\max}(j)}.$$

Thus, for to upper bound the regret upto the end of the phase when the local deletion vanishes is given as $\mathbb{E}[S_{V_j}] \leq \mathbb{E}[S_{F_{J_{\max}(j)}}]$. Combining the above two inequalities with the result in Lemma 4 we obtain the final regret bound as

$$\begin{aligned}
&\mathbb{E}[R_j(T)]\\
&\leq R + C(i_1-1) + C(j-1)i^* + (j-1)2^{i^*} + (C+1)(j-1)K\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2} + \min(1,\beta|\mathcal{H}_j|) \times \dots\\
&\dots \times \left(R + C(i_1-1) + C(J_{\max}(j)-1)i^* + (J_{\max}(j)-1)2^{i^*} + (C+1)(J_{\max}(j)-1)K\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}\right)\\
&+ (K-1+|\mathcal{B}_{jk_j^*}|)\log_2(T) + NK\left(i_1 + (j-1)i^* + (J_{\max}(j)-1)K\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}\right)\\
&+ \sum_{k\notin\mathcal{D}_j}\sum_{j'\in\mathcal{B}_{jk}:k\notin\mathcal{D}_{j'}} \tfrac{8\gamma\mu_{k_j^*}}{\Delta_{j'k}^2}\left(\log(T) + \sqrt{\tfrac{\pi}{\gamma}\log(T)}\right) + \sum_{k\notin\mathcal{D}_j\cup k_j^*} \tfrac{8\gamma}{\Delta_{jk}}\left(\log(T) + \sqrt{\tfrac{\pi}{\gamma}\log(T)}\right)\\
&+ NK\left(1 + (\psi(\gamma)+1)\frac{8\gamma}{\Delta_{\min}^2}\right)\\
&\leq \sum_{k\notin\mathcal{D}_j}\sum_{j'\in\mathcal{B}_{jk}:k\notin\mathcal{D}_{j'}} \tfrac{8\gamma\mu_{k_j^*}}{\Delta_{j'k}^2}\left(\log(T) + \sqrt{\tfrac{\pi}{\gamma}\log(T)}\right) + \sum_{k\notin\mathcal{D}_j\cup k_j^*} \tfrac{8\gamma}{\Delta_{jk}}\left(\log(T) + \sqrt{\tfrac{\pi}{\gamma}\log(T)}\right)\\
&+ (K-1+|\mathcal{B}_{jk_j^*}|)\log_2(T) + O\left(\tfrac{N^2K^2}{\Delta_{\min}^2} + (\min(1,\beta|\mathcal{H}|_j)J_{\max}(j) + j - 1)2^{i^*} + N^2Ki^*\right)
\end{aligned}$$

This gives us the following rerget bound under the SPC setting

Theorem 4. *For a stable matching instance satisfying α-condition (Definition 3), suppose each agent follows* UCB-D4 *(Algorithm 2) with $\gamma > 1$ and $\beta \in (0, 1/K)$, then the regret for an agent $j \in [N]$ is upper bounded by*

$$\begin{aligned}
\mathbb{E}[R_j(T)] \leq &\underbrace{\sum_{k\notin\mathcal{D}_j\cup k_j^*} \tfrac{8\gamma}{\Delta_{jk}}\left(\log(T) + \sqrt{\tfrac{\pi}{\gamma}\log(T)}\right)}_{\textit{sub-optimal match}} + \underbrace{\sum_{k\notin\mathcal{D}_j}\sum_{j'\in\mathcal{B}_{jk}:k\notin\mathcal{D}_{j'}} \tfrac{8\gamma\mu_{k_j^*}}{\Delta_{j'k}^2}\left(\log(T) + \sqrt{\tfrac{\pi}{\gamma}\log(T)}\right)}_{\textit{collision}}\\
&+ \underbrace{(K-1+|\mathcal{B}_{jk_j^*}|)\log_2(T)}_{\textit{communication}} + + \underbrace{O\left(\tfrac{N^2K^2}{\Delta_{\min}^2} + (\beta|\mathcal{H}_j|J_{\max}(j) + j - 1)2^{i^*}\right)}_{\textit{transient phase, independent of } T},
\end{aligned}$$

where

$i^* = \max\{8, i_1, i_2\}, i_1 = \min\{i : (N-1)\frac{10\gamma i}{\Delta_{min}^2} < \beta 2^{(i-1)}\},$ *and* $i_2 = \min\{i : (N-1+NK(i-1)) \leq 2^{i+1}\}.$

E. Proof of Regret Upper Bound under α-Condition

In this section we prove our main result for the instances satisfying α-condition. We will present a short note on the main proof idea, while pointing out why the proof in the previous section does not go through. Next we present the necessary notations before going into the proof of the results. The proof structure, and some parts of the proof remain closely related to that of the previous section. Therefore, we mainly focus the new parts of the proof, while referring to the parts related to SPC we present proof sketch.

E.1. Main Proof Idea

The key idea of the proof is similar to SPC condition but now before the global deletion starts to freeze, we need to talk about vanishing of local deletion for the stable matched arms (note the sub-optimal arms for each agent can still get locally deleted at this point). So the three important stages are: (1) local deletion vanishes for stable matched arms (from agent A_1 to A_N), (2) freezing of global deletion (from agent 1 to N), (3) vanishing of local deletion of all arms (depending on when the blocking agents freeze global deletion). We next elaborate more on why (1) should precede (2) under α-condition whereas under SPC condition we can directly go to (2).

Under SPC for agent 1 there was no risk of local deletion for it's stable match pair, which is also its best arm, as for this arm agent 1 is also the best agent. This sets up the inductive freezing of the global deletion as agent 1 quickly identifies arm 1 as it's best arm. The vanishing of local deletion is the consequence of the freezing of global deletion of the blocking agents. But under α-condition it is no longer the case as agent 1 is not the most preferred agent for arm 1. Instead we have that the agent A_1 has no risk of local deletion of its stable match pair, a_1, which is (possibly) not the best arm for agent A_1 but for arm a_1 we have A_1 as its best agent. Therefore, agent A_1 will not delete it's stable match pair arm a_1, but unless global deletion eliminates better arms it will not converge to this arm. However, A_1 will stop causing local deletion (which we will prove) for the stable matched arm for agents in the set $\{j : A_1 >_{k^*_j} j, j \in [N]\}$. This will continue inductively. In particular, A_1 stops local deletion of stable matched arm of agent A_2 which in turn stops local deletion caused by agent A_2, so on and so forth.

Where proof of SPC fails for α-condition? Before going into the proof of α-condition we identify why the proof in previous section fails. The key step that breaks when we move from SPC to α-condition is that in Lemma 4 the ineuality (7) does not hold anymore. The issue is we do not have k^*_j to be dominated only by the agents 1 to $(j-1)$, i.e. there may exist agent $j' > j$ such that $j' >_{k^*_j} j$. Similar idea is also exploited in Lemma 5 which also fails to hold for the same reason.

E.2. Notations and Definitions:

We setup the notations required for the regret upper bound proof when the system satisfies α-condition. The right-order in the definition of the α-condition be given as $[N]_r = \{A_1, A_2, \dots, A_N\}$ (a permutation of $[N]$) for the agents, and $\{a_1, a_1, \dots, a_K\}$ (a permutation of $[K]$) for the arms. Whereas, the left-order in the definition is $[N]$ and $[K]$. Also, we recall that k^*_j as the stable matched arm for any agent $j \in [N]$, and j^*_k as the stable matched agent for the arm k, for all $k \in [K], k \le N$.

We now recall that due to α-condition the following statements hold

$$
\begin{aligned}
&(i)\, \forall j \in [N], \forall k > j \in [K], \mu_{jj} > \mu_{jk},\\
&(ii)\, \forall a_k \in [K]_r, k \le N, \forall j > k, A_j \in [N]_r, A_{j^*_{a_k}} >_{a_k} A_j,\\
&(iii)\, \forall A_j \in [N]_r, k^*_{A_j} = a_j \in [K]_r,\\
&(iv)\, \forall j \in [N], k^*_j = j \in [K],
\end{aligned}
$$

Here, (i) and (ii) follows from the definition of α-stability and (iii) and (iv) follows from the Proposition 3. Let us denote by lr the mapping of agents in left order to agents in right order under α-condition, i.e. agent $j = A_{lr(j)}$ for all $j \in [N]$.

Arm Classification: For each agent j, the *dominated arms* ($\mathcal{D}_j$), the *blocking agents* for arm k and agent j ($\mathcal{B}_{jk}$), the set of *hidden arms* ($\mathcal{H}_j$) are defined identically to the SPC scenario. Let $K_W(j)$ be the set of arms each of which is a stable matched arm for some other agent j', is a sub-optimal arm for j, and j is preferred by that arm than its stable pair j', i.e.

$$K_W(j) = \{k : k \in [K], \mu_{jk} < \mu_{jk^*_j}, \exists j' \neq j : (k = k^*_{j'}, j >_k j')\}.$$

We note that $K_W(A_j) \le (K - j)$ as due to α-condition agent $k^*_{j'} \notin K_W(j)$ for any $j \le N$.

Ciritcal Phases: We now define the critical phases when the system satisfies the α-condition

- The phase i for agent j, for some $j \in [N]$, is a *Warmup Phase* if the following are true for each arm $k \in K_W(j)$,:
  1. in phase i arm k is matched with agent j at most $\frac{10\alpha i}{\Delta_{jk}^2}$ times,
  2. in phase i arm k is not agent j's most matched arm
- The phase i for agent j, for some $j \in [N]$, is an α-*Good Phase* if the following are true:
  1. The dominated arms are globally deleted, i.e. $G_j[i] = \mathcal{D}_j$.
  2. The phase i is a *warmup phase* for all agents in $\mathcal{L}_j = \{j' : k_j^* \in K_W(j')\}$.
  3. For each arm $k \notin \mathcal{D}_j \cup k_j^*$, in phase i arm k is matched with agent j at most $\frac{10\alpha i}{\Delta_{jk}^2}$ times.
  4. The stable match pair arm k_j^* is matched the most number of times in phase i.

  The α-good phase is not identical to good phase as condition (2) is additional in this case.
- A phase i for agent j, for some $j \in [N]$ is called α-*Low Collision Phase* if the following are true:
  1. Phase i is a α-good phase for agents 1 to j.
  2. Phase i is a α-good phase for all agent $j' \in \cup_{k \in \mathcal{H}_j} \mathcal{B}_{jk}$.

  The α-low collision phase is identical to low collision phase (in SPC) except the good phase is replaced with α-good phase.

We define for agent j, similar to SPC, $\mathbb{I}_{G_\alpha}[i, j]$ to be the indicator that phase i is a α-good phase, $\mathbb{I}_{LC_\alpha}[i, j]$ to be the indicator that phase i is a α-low collision phase, and $\mathbb{I}_W[i, j]$ to be the indicator that phase i is a warmup phase.

Let $i_1 = \min\{i : (N-1)\frac{10\gamma i}{\Delta_{min}^2} < \beta 2^{(i-1)}\}$. For each agent j, the α-*Freezing* ($F_{\alpha j}$) phase is the phase on or after which the agents 1 to $(j-1)$ are in α-good phase, and all the $j'' \in \mathcal{L}_j$ (henceforth *deadlock agents*) are in warmup phase.

$$F_{\alpha j} = \max\left(i_1, \min\left(\{i : \prod_{i' \geq i}\left(\prod_{j'=1}^{(j-1)} \mathbb{I}_{G_\alpha}[i', j']\right)\left(\prod_{j'' \in \mathcal{L}_j} \mathbb{I}_W[i', j'']\right) = 1\} \cup \{\infty\}\right)\right).$$

Also, we define α-Vanishing phase ($V_{\alpha j}$) similar to SPC

$$V_{\alpha j} = \max\left(i_1, \min\left(\{i : \prod_{i' \geq i} \mathbb{I}_{LC_\alpha}[i', j] = 1\} \cup \{\infty\}\right)\right).$$

Similar to SPC, $V_{\alpha j} = \max\left(F_{\alpha(j+1)}, \cup_{k \in \mathcal{H}_j} \cup_{j' \in \mathcal{B}_{jk}} F_{\alpha j'}\right)$ from the definition of low collision phase.

Finally, for each $j \leq N$, the phase i is the *Unlocked* phase (U_j) if all phases on and after i are warmup phases for all the agents A_1 to A_j.

$$U_j = \max\left(i_1, \min\left(\{i : \prod_{j'=1}^{lr(j)-1} \prod_{i' \geq i} \mathbb{I}_W[i', A'_j] = 1\} \cup \{\infty\}\right)\right).$$

This will be useful in quantifying the α-freezing phase $F_{\alpha j}$ later on.

E.3. Structural results for α-condition

In this section, we collect the important results that hold due to the combinatorial properties of the stable matching system that satisfies the α-condition.

Proposition 4. *If a system satisfies α-condition then we have $k_j^* = j$, $j_{a_k}^* = A_k$ and $k_{A_j}^* = a_j$ for all $1 \leq k, j \leq N$.*

Proof. That under α-condition $k_j^* = j$ for all $1 \leq j \leq N$ follows identically to Proposition 4. For the final relation we note that under α-condition we have for $k = 1$ we have $A_{j_1^*} >_{a_1} A_j$ for all $j > 1$. Thus $j_1^* = A_1$. We can extend the same logic to obtain $j_{a_k}^* = A_k$ for all $1 \leq k \leq N$. □

We now prove that the arm k_j^* can be blocked only by agents in $\mathcal{L}_j$.

Claim 1. *For a stable matching* $\mathbf{k}^*$ *and any agent* j*, we have* $\{j' : j' >_{k_j^*} j\} \subseteq \mathcal{L}_j = \{j' : k_j^* \in K_W(j')\}$.

Proof. We have the stable matching $\mathbf{k}^*$. Let $j >_{k_{j'}^*} j'$ and $\mu_{jk_j^*} < \mu_{jk_{j'}^*}$, then $(j, k_{j'}^*)$ forms a blocking pair as arm $k_{j'}^*$ and agent j will be both happier switching from their respective partners under $\mathbf{k}^*$. Therefore, $\mathbf{k}^*$ is not a stable matching. Thus, for a stable matching $\mathbf{k}^*$ and any two agents $1 \leq j, j' \leq N$, agent j satisfies $\mu_{jk_j^*} > \mu_{jk_{j'}^*}$ if $j >_{k_{j'}^*} j'$. Thus, if $\{j' >_{k_j^*} j\}$ then $\mu_{jk_j^*} < \mu_{jk_{j'}^*}$ so $k_j^* \in K_W(j')$ so $j' \in \mathcal{L}_j$. □

We now characterize the set of deadlock agents for each agent j.

Claim 2. *For each agent* $j \in [N]$*,* $\mathcal{L}_j \subseteq \{A_{j'} : j' = 1, \ldots, lr(j) - 1\}$.

Proof. From α-condition we know that $\forall a_k \in [K]_r, k \leq N, \forall j > k, A_j \in [N]_r, A_{j^*_{a_k}} >_{a_k} A_j$. Further, from Proposition 4 we know that $j^*_{a_k} = A_k$ for all $1 \leq k \leq N$. Therefore, we can observe for any $j, j' \leq N$ and $j < j'$, $A_j >_{k^*_{A_j}} A_{j'}$. In particular, for any $j' > lr(j)$ we have $j = A_{lr(j)} >_{k_j^*} A_{j'}$. Which means for any $j' \geq lr(j)$, we do not have $j' >_{k_j^*} j$ and hence $k_j^* \notin K_W(j')$. This proves that for any $j' \geq lr(j)$ $j' \notin \mathcal{L}_j$, i.e. $\mathcal{L}_j \subseteq \{A_{j'} : j' = 1, \ldots, lr(j) - 1\}$. □

We recall that $lr(j)$ is the index of the agent j in the right-order of α-condition. The above characterization connects the unlock phase with the freezing phase as follows

Claim 3. *For each agent* $j \in [N]$*,* $F_{\alpha j} \leq \max\left(U_{(lr(j)-1)}, \max(F_{\alpha j'} : 1 \leq j' \leq (j-1))\right)$ *w.p.* 1.

Proof. Consider an arbitrary sample path. We know by definition on or after phase $U_{(lr(j)-1)}$, all agents $\{A_{j'} : j' = 1, \ldots, lr(j) - 1\}$ are in warmup phase. We have the set of deadlock agents as $\mathcal{L}_j \subseteq \{A_{j'} : j' = 1, \ldots, lr(j) - 1\}$. Hence, all agents in $\mathcal{L}_j$ are also in warmup phase on or after phase $U_{(lr(j)-1)}$. Further, the agents 1 to $(j-1)$ are in α-good phase from phase $\max(F_{\alpha j'} : 1 \leq j' \leq (j-1))$ onwards. Hence, $F_{\alpha j} \leq \max\left(U_{(lr(j)-1)}, \max(F_{\alpha j'} : 1 \leq j' \leq (j-1))\right)$ with probability 1. □

Next the following lemma captures a few key properties related to the critical phases.

Lemma 7. *For* $i \geq i_1 = \min\{i : (N-1)\frac{10\gamma i}{\Delta^2_{min}} < \beta 2^{(i-1)}\}$*, any* $j \in [N]$*,*

- *if phase* i *and* $(i-1)$ *are warmup phases for all* $j' \in \mathcal{L}_j$ *then* $k_j^* \notin L_j[i] \cup G_j[i]$ *almost surely,*
- *if phase* $i \geq \min(U_{(lr(j)-1)}, F_{\alpha j}) + 1$ *then* $k_j^* \notin L_j[i] \cup G_j[i]$ *almost surely,*
- *if phase* $i \geq V_{\alpha j} + 1$ *collision phase for agent* j *then* $L_j[i] = \emptyset$ *almost surely.*

Proof. The following results hold for an arbitrary sample path giving us almost sure inequalities.

Due to Claim 1 all agents j' which can block arm k_j^* are in $\mathcal{L}_j$. Also $k_j^* \in K_W(j')$ for any agent $j' \in \mathcal{L}_j$ due to the definition of $\mathcal{L}_j$. Therefore, if all agents in $\mathcal{L}_j$ are in warmup phase in phase $(i-1)$ then $k_j^* \notin G_j[i]$ because no agent in $\mathcal{L}_j$ communicates arm k_j^* to agent j, and the other arms can not communicate the arm k_j^* (due to this arm's preference). Furthermore, the total number of times the arm k_j^* can be deleted is at most $(lr(j) - 1)\frac{10\alpha i}{\Delta^2_{jk}} < \beta 2^{(i-1)}$ (the local deletion threshold) for any $i \geq i_1$. Thus k_j^* is not locally deleted, i.e. $k_j^* \notin L_j[i]$. This proves the first part.

We know that the phase $i \geq U_{lr(j)-1} + 1$ and $(i-1) \geq U_{lr(j)-1}$ is a *warmup phase* for all agents in $\mathcal{L}_j = \{j' : k_j^* \in K_W(j')\}$. This is because we know that $\mathcal{L}_j \subseteq \{A_{j'} : j' = 1, \ldots, lr(j) - 1\}$ due to Claim 2. By definition of $F_{\alpha j}$ all agents are in warmup phase for phases $i \geq F_{\alpha j} + 1$ and $(i-1) \geq F_{\alpha j}$. Thus the second result follows due to the first result.

The proof of the third part follows almost identically to the Lemma 1, i.e. by virtue of $i \geq V_{\alpha j} + 1$ being an α-low collision phase. □

E.4. Proof of main results

In this section, we proceed with the regret bound where we leverage the structural properties proven in the previous part. We first state the regret decomposition lemma, which has an identical form to the regret decomposition as in SPC with $F_{\alpha j}$ and $V_{\alpha j}$ in place of F_j and V_j, respectively.

Lemma 8. *The expected regret for agent j can be upper bounded as*

$$\begin{aligned}\mathbb{E}[R_j(T)] \leq \mathbb{E}[S_{F_{\alpha j}}] + \min(\beta|\mathcal{H}_j|,1)\mathbb{E}[S_{V_{\alpha j}}] + \Big((K-1+|\mathcal{B}_{jk_j^*}|)\log_2(T) + NK\mathbb{E}[V_{\alpha j}]\Big) \\ + \sum_{k\notin\mathcal{D}_j}\sum_{j'\in\mathcal{B}_{jk}:k\notin\mathcal{D}_{j'}} \frac{8\gamma\mu_{k_j^*}}{\Delta_{j'k}^2}\left(\log(T)+\sqrt{\tfrac{\pi}{\gamma}\log(T)}\right) + \sum_{k\notin\mathcal{D}_j\cup k_j^*}\frac{8\gamma}{\Delta_{jk}}\left(\log(T)+\sqrt{\tfrac{\pi}{\gamma}\log(T)}\right) \\ + NK\left(1+(\psi(\gamma)+1)\frac{8\gamma}{\Delta_{\min}^2}\right)\end{aligned}$$

Proof Sketch. The proof of the lemma is closely related to the proof of Lemma 3, except for the use of the α-freezing phase $F_{\alpha j}$ instead of the freezing phase F_j, and α-vanishing phase $V_{\alpha j}$ instead of vanishing phaes V_j. The rest of the proof is identical to the proof of Lemma 3 where Lemma 9 is invoked instead of it's identical counterpart (for SPC) Lemma4. □

Lemma 9. *For any $j \in [N]$, $k \notin \mathcal{D}_j \cup k_j^*$, for $\gamma > 1$,*

$$\mathbb{E}\left[(N_{jk}(T) - N_{jk}(S_{F_j}))\right] \leq \psi(\gamma)\tfrac{8}{\Delta_{jk}^2} + 1 + \tfrac{8}{\Delta_{jk}^2}\left(\gamma\log(T) + \sqrt{\pi\gamma\log(T)} + 1\right).$$

Proof Sketch. The proof of the lemma follows the proof of Lemma 4, again with α-freezing phase $F_{\alpha j}$ in place of the freezing phase F_j. Due to Lemma 7 we know that for each phase $i \geq (F_{\alpha j}+1)$ the arm k_j^* is available as it is neither globally deleted, nor locally deleted. Thus once a sub-optimal arm k is played enough times the UCB of arm k_j^* w.h.p. will be higher than the UCB of k at any round after $F_{\alpha j}$. Using the same standard framework as in Lemma 4 this intuition can be formalized as a proof of this lemma. □

We first show that for phases $i \geq U_{j-1}+1$, the probability that phase i is not a warmup phase for agent A_j is low.

Lemma 10. *For any $j \leq N$ and any phase $i \geq i^* = \max(8, i_1, i_2)$ and $\gamma > 1$,*

$$\mathbb{P}[\mathbb{I}_W[i, A_j] = 0 \wedge i \geq U_{j-1}+1] \leq (K-j)2^{-i(\gamma-1)}\left(1+\tfrac{64}{\Delta_{\min}^2}\right),$$

where $i_1 = \min\{i : (N-1)\frac{10\gamma i}{\Delta_{min}^2} < \beta 2^{(i-1)}\}$ and $i_2 = \min\{i : (R-1+C(i-1)) \leq 2^{i+1}\}$.

Proof. For any arbitrary sample path and any $i \geq U_{j-1}+1$, phase i is a warm up phase for all agent A_1 to A_{j-1}. The phase i is not a warmup phase for agent A_j, if there exists an arm $k \in K_W(A_j)$ which is played more than $\frac{10\gamma i}{\Delta_{A_jk}^2}$ times in phase i. Here, by definition for any $k \in K_W(A_j)$ we have $\mu_{A_jk} \leq \mu_{A_ja_j}$ (recall, $k_{A_j}^* = a_j$ due to Proposition 4) which makes sure $\Delta_{A_jk} > 0$.

The set of agents that can block A_j from matching with arm a_j when A_j plays a_j is given by $\mathcal{L}_{A_j} \subseteq \{A_{j'} : 1 \leq j' \leq j-1\}$ due to Claim 2 and $lr(A_j) = j$. But then due to the second point in Lemma 7 we know that $k_{A_j}^* \notin G_{A_j}[i] \cup L_{A_j}[i]$ for any $i \geq U_{j-1}+1$. Therefore, the inequality (i) below holds as to play arm k the UCB of arm $k_j^* = a_j$ can not be less than arm k. The final bound can be obtained identically to the proof of Lemma 5 for $i \geq \max(8, i_1, i_2)$, with the observation $K_W(j) \leq (K-j)$.

Therefore, we obtain the next set of equations

$$\begin{aligned}&\mathbb{P}[\mathbb{I}_W[i, A_j] = 0 \wedge i \geq (U_{j-1}+1)] \\ &\leq \mathbb{P}\left[\cup_{k\in K_W(A_j)}\{(N_{A_jk}[i] - N_{A_jk}[i-1]) > \tfrac{10\gamma i}{\Delta_{A_jk}^2}\} \wedge i \geq (U_{j-1}+1)\right] \\ &\leq \sum_{k\in K_W(A_j)} \mathbb{P}\left[\cup_{t\in S_i}^{(S_{i+1}-1)} N_{A_jk}(t) = \tfrac{10\gamma i}{\Delta_{A_jk}^2} \wedge I_{A_j}(t) = k \wedge i \geq (U_{j-1}+1)\right]\end{aligned}$$

$$\begin{aligned}
&\overset{(i)}{\leq} \sum_{k\in K_W(A_j)} \sum_{t\in S_i}^{(S_{i+1}-1)} \mathbb{P}\left[N_{A_jk}(t) = \tfrac{10\gamma i}{\Delta^2_{A_jk}} \wedge u_{A_jk}(t-1) > u_{A_ja_j}(t-1)\right]\\
&\leq |K_W(A_j)|2^{-i(\gamma-1)}\left(1+\tfrac{64}{\Delta^2_{\min}}\right)\\
&\leq (K-j)2^{-i(\gamma-1)}\left(1+\tfrac{64}{\Delta^2_{\min}}\right)
\end{aligned}$$

This completes the proof. □

The proof of this lemma resembles closely that of Lemma 5 while some arguments are common to Lemma 10.

Lemma 11. *For any agent j and any phase $i \geq i^* = \max\{8, i_1, i_2\}$ and $\gamma > 1$,*

$$\mathbb{P}[\mathbb{I}_{G_\alpha}[i,j] = 0 \wedge i \geq F_{\alpha j}+1] \leq (K-j)2^{-i(\gamma-1)}\left(1+\tfrac{64}{\Delta^2_{\min}}\right),$$

where i_1 and i_2 is as defined in Lemma 10.

Proof. The phase i is a α-good phase for agent j if (1) the dominated arms are deleted $G_j[i] = \mathcal{D}_j$, (2) phase i is a *warmup phase* for all agents in $\mathcal{L}_j = \{j' : k^*_j \in K_W(j')\}$, (3) for each arm $k \notin \mathcal{D}_j \cup k^*_j$, in phase i arm k is matched with agent j at most $\frac{10\alpha i}{\Delta^2_{jk}}$ times, and (4) the stable match pair arm k^*_j is matched the most number of times in phase i. We see that (1) and (2) holds when $i \geq F_{\alpha j}+1$. Also, (4) holds when (1), (2) and (3) holds for any $i \geq i_1$.

Therefore, we will now show (3) holds. In particular, we have the following series of inequalities

$$\begin{aligned}
&\mathbb{P}[\mathbb{I}_{G_\alpha}[i,j] = 0 \wedge i \geq (F_{\alpha j}+1)]\\
&\leq \mathbb{P}\left[\cup_{k\notin\mathcal{D}_j\cup k^*_j}\{(N_{jk}[i] - N_{jk}[i-1]) > \tfrac{10\gamma i}{\Delta^2_{jk}}\} \wedge i \geq (F_{\alpha j}+1)\right]\\
&\leq \sum_{k\notin\mathcal{D}_j\cup k^*_j} \mathbb{P}\left[\cup_{t\in S_i}^{(S_{i+1}-1)} N_{jk}(t) = \tfrac{10\gamma i}{\Delta^2_{jk}} \wedge I_j(t) = k \wedge i \geq (F_{\alpha j}+1)\right]\\
&\overset{(i)}{\leq} \sum_{k\notin\mathcal{D}_j\cup k^*_j} \sum_{t\in S_i}^{(S_{i+1}-1)} \mathbb{P}\left[N_{jk}(t) = \tfrac{10\gamma i}{\Delta^2_{jk}} \wedge u_{jk}(t-1) > u_{jk^*_j}(t-1)\right]\\
&\leq (K-j)2^{-i(\gamma-1)}\left(1+\tfrac{64}{\Delta^2_{\min}}\right).
\end{aligned}$$

We know that for all arms $k \notin \mathcal{D}_j \cup k^*_j$ we have $\Delta_{jk} > 0$ by definition of $\mathcal{D}_j$. Also, inequality (i) holds as due to Lemma 7, we know that after $i \geq (F_{\alpha j}+1)$ the arm k^*_j is not globally or locally deleted. The rest again follows similar to Lemma 5 for $i \geq \max\{8, i_1, i_2\}$. □

Let us define $lr_{\max}(j) = \max(lr(j') : 1 \leq j' \leq j)$, and $\tilde{F}_j = \max\left(U_{(lr_{\max}(j)-1)}, \max(\tilde{F}_{j'} : 1 \leq j' \leq (j-1))\right)$ for each j. It is easy to see that $\tilde{F}_j > F_{\alpha j}$ due to Claim 3 for any j and the fact that $U_{(lr_{\max}(j)-1)} \geq U_{(lr(j)-1)}$ due to the definition of U_j (all agents from A_1 to A_j all are in warmup phase till the end). We now present the the following lemma that bounds the probability that a phase i is not an α-good phase when $i \geq F_j + 1$. We now bound the moments and exponents of $\tilde{F}_j$.

Lemma 12. *For any $j \in [N]$ and $m \geq 1$, the following hold with i^* as defined in Lemma 11*

$$\begin{aligned}
\mathbb{E}[\tilde{F}^m_j] &\leq 2i_1 + (lr_{\max}(j)+j-2)\left((i^*)^m + K\left(1+\tfrac{64}{\Delta^2_{\min}}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}\right)\\
\mathbb{E}[2^{\tilde{F}_j}] &\leq 2i_1 + (lr_{\max}(j)+j-2)\left(2^{i^*} + K\left(1+\tfrac{64}{\Delta^2_{\min}}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}\right)
\end{aligned}$$

Proof. We again inductively bound the expectation of an arbitrary monotonically increasing and continuous (hence invertible) function $g : \mathbb{R} \to \mathbb{R}_+$. We have that $F_0 = i_1$ almost surely by definition (this accounts for the max with i_1 in the definition of F_j).

We calculate the expectation for agent j as

$$\begin{aligned}
\mathbb{E}[g(\tilde{F}_j)] &= \sum_{x\geq 0}\mathbb{P}[g(\tilde{F}_j)\geq x] \leq \sum_{x\geq 0}\mathbb{P}[\tilde{F}_j \geq g^{-1}(x)]\\
&\leq \sum_{x\geq 0}\mathbb{P}[\tilde{F}_j \geq g^{-1}(x), U_{(lr_{\max}(j)-1)} \geq g^{-1}(x)] + \sum_{x\geq 0}\mathbb{P}[\tilde{F}_j \geq g^{-1}(x), U_{(lr_{\max}(j)-1)} < g^{-1}(x)]\\
&\leq \sum_{x\geq 0}\mathbb{P}[U_{(lr_{\max}(j)-1)} \geq g^{-1}(x)] + \sum_{x\geq 0}\mathbb{P}[\tilde{F}_j \geq g^{-1}(x), U_{(lr_{\max}(j)-1)} < g^{-1}(x)]\\
&\leq \mathbb{E}[g(U_{(lr_{\max}(j)-1)})] + \sum_{x\geq 0}\mathbb{P}[\tilde{F}_j \geq g^{-1}(x), U_{(lr_{\max}(j)-1)} < g^{-1}(x)]\\
&\leq i_1 + (lr_{\max}(j)-1)g(i^*) + (lr_{\max}(j)-1)(K - lr_{\max}(j)/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}\\
&+ i_1 + (j-1)g(i^*) + (j-1)(K-j/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}\\
&\leq 2i_1 + (lr_{\max}(j)+j-2)g(i^*)\\
&+ ((lr_{\max}(j)+j-2)K - (lr_{\max}(j)(lr_{\max}(j)-1)+j(j-1))/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}\\
&\leq 2i_1 + (lr_{\max}(j)+j-2)\left(g(i^*) + K\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}\right).
\end{aligned}$$

The last inequality is loose, and we use it for simplicity. For the second last inequality we use the following bounds on $\mathbb{E}[g(U_{(lr_{\max}(j)-1)})]$ and $\mathbb{P}[\tilde{F}_j \geq g^{-1}(x), U_{(lr_{\max}(j)-1)} < g^{-1}(x)]$ which we will prove momentarily.

$$\begin{aligned}
&\mathbb{E}[g(U_j)] \leq i_1 + (j-1)g(i^*) + (j-1)(K-j/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2},\\
&\sum_{x\geq 0}\mathbb{P}[\tilde{F}_j \geq g^{-1}(x), U_{(lr_{\max}(j)-1)} < g^{-1}(x)] \leq i_1 + (j-1)g(i^*) + (j-1)(K-j/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}.
\end{aligned}$$

Case 1: The base case $U_0 = i_1$ holds almost surely by definition. We have

$$\begin{aligned}
\mathbb{E}[g(U_j)] &= \sum_{x\geq 0}\mathbb{P}[g(U_j)\geq x] \leq \mathbb{E}[g(U_{j-1})] + g(i^*) + \sum_{i\geq i^*}\mathbb{P}[U_j \geq i, U_{j-1} < i]\\
&\leq \mathbb{E}[g(F_{j-1})] + g(i^*) + \sum_{i\geq i^*}\mathbb{P}[\{\exists i' \geq i, \mathbb{I}_W[i', j] = 0\}, U_{j-1}+1 \leq i]\\
&\overset{(i)}{\leq} i_1 + (j-1)g(i^*) + (j-1)(K-j/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}.
\end{aligned}$$

Here, for (i) we use the inequality in Lemma 10, and take summations over i (similar to Lemma 6).

Case 2: We again proceed inductively. For any $j \in [N]$, we introduce the notation

$$\mathcal{F}_j := \sum_{x\geq 0}\mathbb{P}[\tilde{F}_j \geq g^{-1}(x), U_{(lr_{\max}(j)-1)} < g^{-1}(x)].$$

In the base case, as $\tilde{F}_0 = i_1$, we have $\mathcal{F}_0 \leq i_1$. Proceeding with the inductive approach

$$\begin{aligned}
&\sum_{x\geq 0}\mathbb{P}[\tilde{F}_j \geq g^{-1}(x), U_{(lr_{\max}(j)-1)} < g^{-1}(x)]\\
&\leq \sum_{x\geq 0}\mathbb{P}[\tilde{F}_j \geq g^{-1}(x), \tilde{F}_{j-1} \geq g^{-1}(x), U_{(lr_{\max}(j-1)-1)} < g^{-1}(x)]\\
&+ \sum_{x\geq 0}\mathbb{P}[\tilde{F}_j \geq g^{-1}(x), \tilde{F}_{j-1} < g^{-1}(x), U_{(lr_{\max}(j-1)-1)} < g^{-1}(x)]\\
&\overset{(i)}{\leq} \sum_{x\geq 0}\mathbb{P}[\tilde{F}_{j-1} \geq g^{-1}(x), U_{(lr_{\max}(j-1)-1)} < g^{-1}(x)]
\end{aligned}$$

$$
\begin{aligned}
&+ \sum_{x\geq 0} \mathbb{P}[\{\exists i' \geq g^{-1}(x), \mathbb{I}_{G_\alpha}[i',j]=0\}, F_{\alpha(j-1)} < g^{-1}(x)] \\
&\leq \mathcal{F}_{j-1} + \sum_{x\geq 0} \mathbb{P}[\{\exists i' \geq g^{-1}(x), \mathbb{I}_{G_\alpha}[i',j]=0\}, F_{\alpha(j-1)} < g^{-1}(x)] \\
&\leq \mathcal{F}_{j-1} + g(i^*) + \sum_{i\geq i^*} \mathbb{P}[\{\exists i' \geq i, \mathbb{I}_{G_\alpha}[i',j]=0\}, F_{\alpha(j-1)} < i] \\
&\leq \mathcal{F}_{j-1} + g(i^*) + \sum_{i\geq i^*}\sum_{i'\geq i} \mathbb{P}[\{\exists i' \geq i, \mathbb{I}_{G_\alpha}[i',j]=0\}, i \geq F_{\alpha(j-1)}+1] \\
&\leq \mathcal{F}_{j-1} + g(i^*) + \sum_{i'\geq i^*} (i'-i^*+1)\mathbb{P}[\{\exists i' \geq i^*, \mathbb{I}_{G_\alpha}[i',j]=0\}, i' \geq F_{\alpha(j-1)}+1] \\
&\leq \mathcal{F}_{j-1} + g(i^*) + \sum_{i'\geq i^*} (i'-i^*+1)\mathbb{P}[\{\exists i' \geq i^*, \mathbb{I}_{G_\alpha}[i',j]=0\}, i' \geq F_{\alpha(j-1)}+1] \\
&\overset{(ii)}{\leq} \mathcal{F}_{j-1} + g(i^*) + (K-j)\left(1+\tfrac{64}{\Delta_{\min}^2}\right) \sum_{i'\geq i^*} (i'-i^*+1)2^{-i'(\gamma-1)} \\
&\leq i_1 + (j-1)g(i^*) + (j-1)(K-j/2)\left(1+\tfrac{64}{\Delta_{\min}^2}\right) \tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}.
\end{aligned}
$$

For the inequality (i) we use the fact that given $U_{(lr_{\max}(j-1)-1)}, \tilde{F}_{j-1} < g^{-1}(x)$ the only way we can have $\tilde{F}_{j-1} \geq g^{-1}(x)$ if for some phase $i' > g^{-1}(x)$ agent j is not in an α-good phase. Then we use Lemma 11 to obtain inequality (ii). □

For the expected rounds upto the end of phase F_j is upper bounded as

$$
\begin{aligned}
\mathbb{E}[S_{F_{\alpha j}}] &= \mathbb{E}[R + C(F_{\alpha j}-1) + 2^{F_{\alpha j}}] \leq \mathbb{E}[R + C(\tilde{F}_j - 1) + 2^{\tilde{F}_j}] \\
&\leq R + C(2i_1 - 1) + C(lr_{\max}(j)+j-2)i^* + (lr_{\max}(j)+j-2)2^{i^*} \\
&+ (C+1)(lr_{\max}(j)+j-2)K\left(1+\tfrac{64}{\Delta_{\min}^2}\right)\tfrac{2^{-(\gamma-1)(i^*-2)}}{(2^{(\gamma-1)}-1)^2}
\end{aligned}
$$

Similar to SPC condition, we define $J_{\max}(j) = \max\left(j+1, \{j' : \exists k \in \mathcal{H}_j, j' \in \mathcal{B}_{jk}\}\right)$. Then as $\tilde{F}_j \geq \tilde{F}_{\alpha j'}$ almost surely for all $j \geq j'$ by definition and $\tilde{F}_j \geq F_{\alpha j}$, we have

$$
V_{\alpha j} = \max\left(F_{\alpha(j+1)}, \cup_{k\in\mathcal{H}_j} \cup_{j'\in\mathcal{B}_{jk}} F_{\alpha j'}\right) \leq \max\left(\tilde{F}_{(j+1)}, \cup_{k\in\mathcal{H}_j} \cup_{j'\in\mathcal{B}_{jk}} \tilde{F}_{j'}\right) = \tilde{F}_{J_{\max}(j)}.
$$

The regret upto the end of the phase when the local deletion vanishes is bounded as

$$
\mathbb{E}[S_{V_{\alpha j}}] \leq \mathbb{E}[V_{\alpha j}^{(\beta+1)}] \leq \mathbb{E}[S_{\tilde{F}_{J_{\max}(j)}}]
$$

The regret bound for the α-condition in Theorem 3 (identically derived as in the SPC case) is obtained by combining the above results as,

$$
\begin{aligned}
&\mathbb{E}[R_j(T)] \\
&\leq \sum_{k\notin\mathcal{D}_j} \sum_{j'\in\mathcal{B}_{jk}: k\notin\mathcal{D}_{j'}} \tfrac{8\gamma\mu_{k_j^*}}{\Delta_{j'k}^2}\left(\log(T) + \sqrt{\tfrac{\pi}{\gamma}\log(T)}\right) + \sum_{k\notin\mathcal{D}_j\cup k_j^*} \tfrac{8\gamma}{\Delta_{jk}}\left(\log(T) + \sqrt{\tfrac{\pi}{\gamma}\log(T)}\right) \\
&+ c_j\log_2(T) + O\left(\tfrac{N^2K^2}{\Delta_{\min}^2} + (\min(1,\beta|\mathcal{H}|_j)f_\alpha(J_{\max}(j)) + f_\alpha(j) - 1)2^{i^*} + N^2Ki^*\right)
\end{aligned}
$$

with the definition that $f_\alpha(j) = j + lr_{\max}(j)$.

This completes the proof of Theorem 3, as the regret bound for the SPC mentioned in the theorem holds due to Theorem 4.

F. Additional Experimental Results

In this section, we present missing details of the dataset generation procedure and additional empirical results.

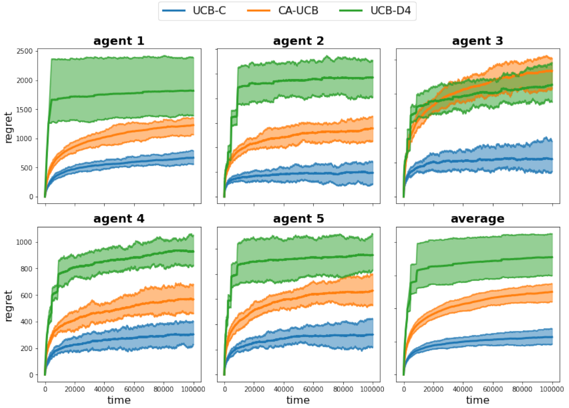


Figure 4. Regret in a general instance (not satisfying α-condition) with 5 agents and 6 arms.

F.1. Synthetic Dataset generation

We use random instances to generate the results in this paper. For each instance the various algorithms are run for 50 times and the average and confidence intervals are constructed using these 50 trials.

For the preference of the agents, we first create a random matrix $\mu \in [0,1]^{N\times K}$ where each entry in the matrix is a i.i.d. $[0,1]$ random variable. The minimum reward gap $\Delta_{min} \approx 0.05$ is enforced through rejection sampling. The agents preferences over the arms is given by the realization of this random matrix. We use different random matrices for different instances.

The preferences of the arms, varies across the three setting – SPC, α-condition, and general instances.

- For a general instance, we simply assign each arm with a random permutation over the agents as its preference list.
- We start with a separate random preference list for each arm. To make this satisfy the SPC condition, we go in the order $1, 2, \dots, K$ of the arms. For an arm i, we find the first position in its preference assigned by the random permutation where an agent $j \geq i$ is present, then swap agent i with agent j to the end (if $j = i$ nothing is done). It is easy to see that this will satisfy the SPC condition.
- We generate the α-condition instance by generating an arbitrary preference list (sample without replacement from possible permutations) for the arms, and then checking whether the instance (along with the agent preference fixed by the arm means) satisfies alpha condition following (Karpov, 2019).

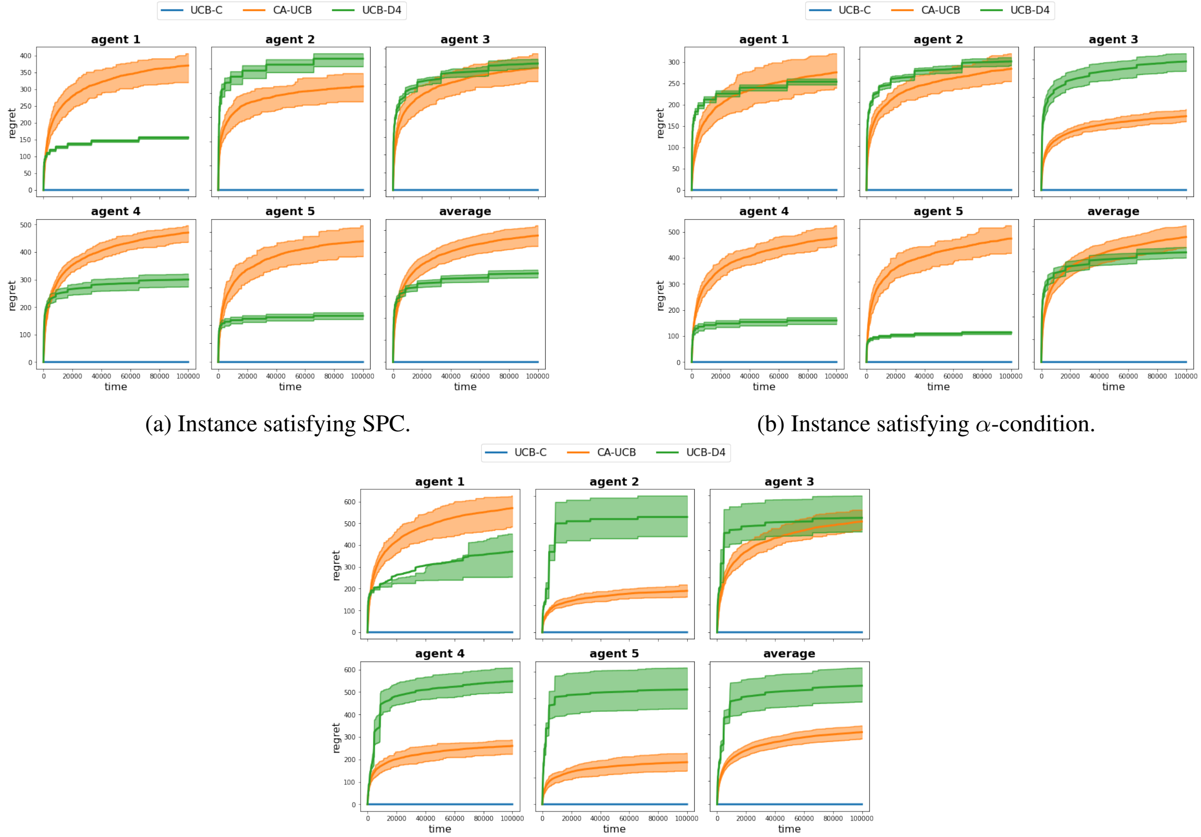
For the UCB-D4 algorithm we use $\beta = 1/2K$, for the CA-UCB we use $\lambda = 0.2$ and for Phased ETC we use $\epsilon = 0.2$ for the $N = 5$ and $K = 6$ case.

F.2. Performance of UCB-C, CA-UCB and UCB-D4 on general instances

In this sub-section, we describe the results of the three algorithms with $N = 5$ agents and $K = 6$ arms on instances that go beyond the uniqueness consistency assumption. Note that in theory, UCB-C provides the optimal $\log(T)$ guarantee, CA-UCB provides a (possibly sub-optimal) guarantee of $\log^2(T)$ while we have no theoretical upper-bound on the regret of UCB-D4. Nonetheless, the results in Figure 4 seem to indicate that CA-UCB has a potentially stronger theoretical upper-bound since its performance is very close to that of UCB-C which has $\log(T)$ upper-bound in the worst-case. Surprisingly, we also see that UCB-D4 *converges* with all the agents eventually obtaining a sub-linear regret indicating that this algorithm may indeed have theoretical upper-bounds even in the more general setup.

F.3. Collision Regret for UCB-C, CA-UCB and UCB-D4

In this sub-section, we show the collision regret incurred by each of the three algorithms in the three settings under which we study their overall regret. As expected, UCB-C has no collision regret because of centralized communication. Surprisingly, CA-UCB has high regret due to collision despite having additional feedback in the SPC setting. This seems to indicate that



(a) Instance satisfying SPC.

(b) Instance satisfying α-condition.

(c) General instances.

Figure 5. Collision regret comparison with 5 agents and 6 arms.

most of the regret contribution for CA-UCB comes from collisions and once they are resolved the dynamics of should settle to a state which incurs no further regret.

F.4. Performance of the algorithms on larger instances

In this sub-section we run the algorithms for larger instances. In particular, we have $N = 11$ agents and $K = 15$ agents.[2]

Tuned Phase Length: We tune the phase length for larger instances. The tuning mainly balances some boundary conditions arising due to large communication blocks (which is only there in the fully decentralized setting) for large instances. Specifically, with large instances in the initial phases communication creates large regret if the phase lengths are small where not many samples can be explored. For tuning Phased ETC (Algorithm 1) we use exponent c_0, and multiplier c_1, where the i-th phase now has length $c_1 \times c_0^i$. We have $c_1 = 1$ and $c_0 = 2$ for Algorithm 1. For tuning UCB-D4 (Algorithm 2) we introduce exponent c_0, and multiplier c_1, where the i-th phase now has length $\big((N-1)K + c_1 \times c_0^i\big)$. The UCB-D4(Algorithm 2) presented in the main paper we have $c_0 = 2$, and $c_1 = 1$.

The hyper-parameters for these plots are as follows. We use
1. phase exponent $c_0 = 1.5$, phase multiplier $c_1 = 1$, and exploration degree $\epsilon = 0.2$ for Phased-UCB,
2. phase exponent $c_0 = 1.2$, phase multiplier $c_1 = 3$, and the local collision threshold $\beta = 1/2K$ for UCB-D4, and
3. $\lambda = 0.2$ for CA-UCB.

The results that were previously observed also hold similarly for this larger instance. We note that the negative regret in the centralized UCB is natural, as during the initial phases an agent can match with an arm which has higher mean than its stable matched arm.

[2] The number 11 was chosen to obtain a rectangular 3×4 grid plot

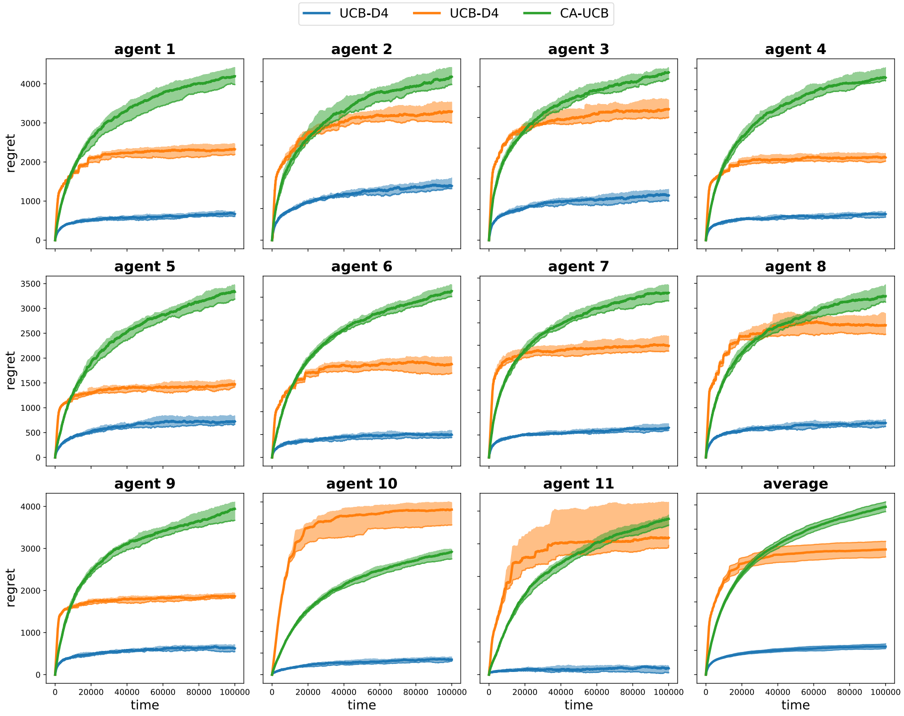


Figure 6. Instance satisfying SPC with 11 agents, and 15 arms.

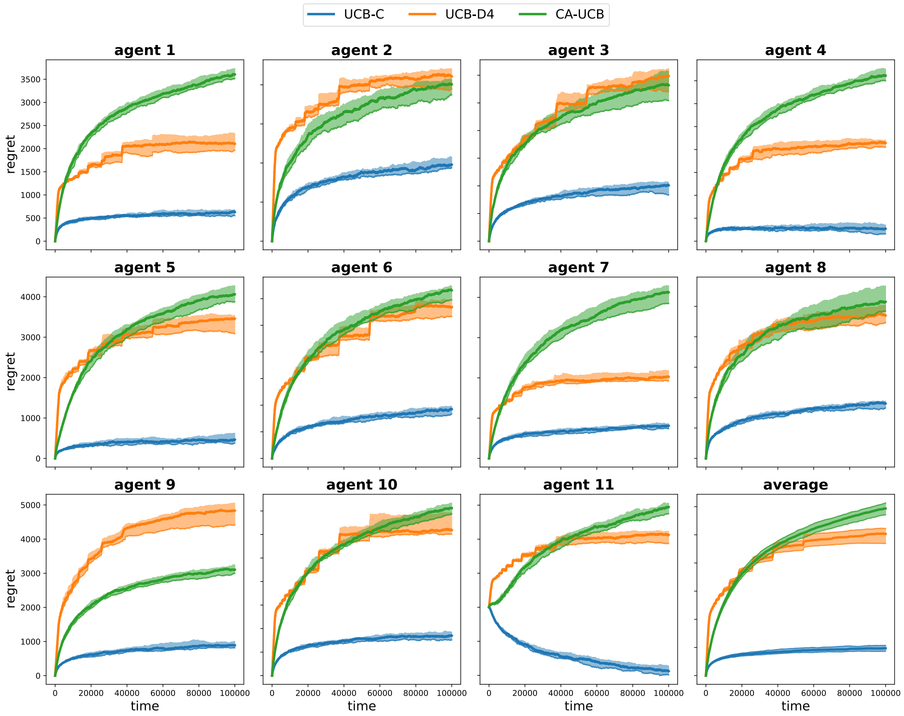


Figure 7. Instance satisfying α-condition with 11 agents, and 15 arms.

Regret Guarantees: The regret bounds remain mostly unchanged due to the above tuning. The regret of the modified Phased ETC is given by replacing the $\log_2(T)$ by $\log_{c_0}(T/c_1)$, and changing the constant to $\Theta\left(c_0^{1/\Delta^{2/\varepsilon}}\right)$. For

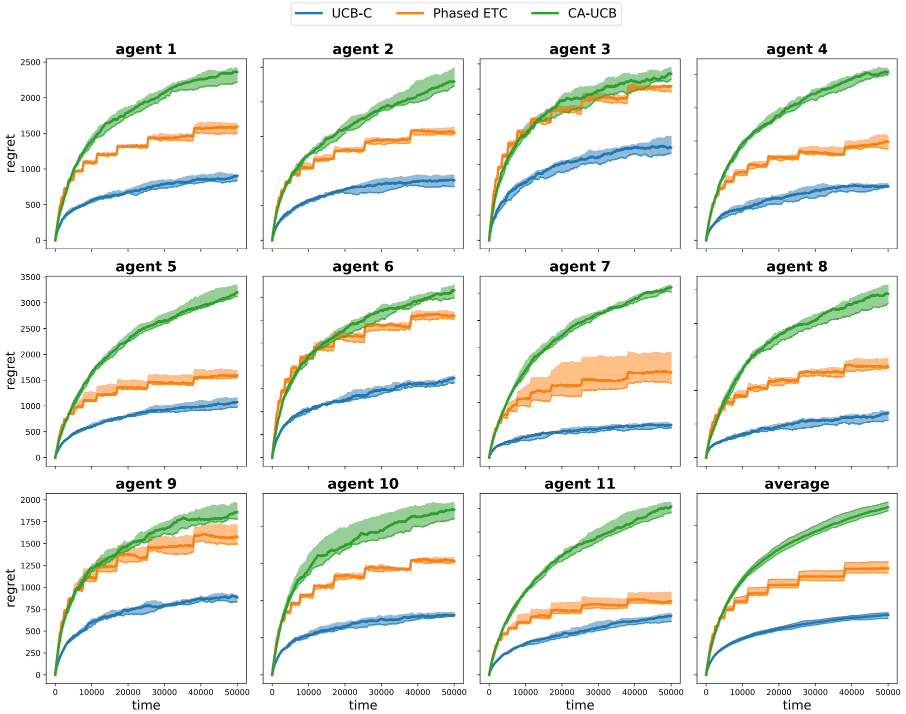


Figure 8. General instances with 11 agents, and 15 arms.

the modified UCB-D4 algorithm the $\log(T)$ regret due to collision and sub-optimal play does not change. The communication regret changes to $(K - 1 + |\mathcal{B}_{jk_j^*}|)\log_{c_0}(T/c_1)$. Finally, the constant part of the regret still remains $O\left(\max\left\{\frac{N}{\Delta_{min}^2}\log(\frac{N}{\Delta_{min}^2}), NK\log(NK)\right\}\right)$.