NO-REGRET LEARNING FOR CAPACITY-CONSTRAINED α-FAIR MULTI-AGENT BANDITS

Anonymous authors
Paper under double-blind review

ABSTRACT

We study stochastic multi-agent bandits with finite arm capacities and a weighted α-fair objective. The capacity-coupled matching action space and the non-Lipschitz objective near zero make standard upper confidence bound (UCB) analysis inapplicable. We propose a centralized two-phase cover-elimination-UCB algorithm that adaptively identifies a safe scale and uses a global elimination rule to remove certifiably poor base-arms while preserving optimal ones, thereby entering a Lipschitz regime. We derive a two-term regret upper bound, consisting of a capacity-coupled screening term for near-zero bad base-arms and a learning term in the safe Lipschitz regime, and establish a lower bound matching the T- and $\mu^*_{\min}$-dependence up to logarithmic factors. We then develop a low auction-call complexity decentralized algorithm based on geometric shrinking, utilizing dual prices from a decentralized auction to compute forced-insertion gaps through augmented alternating cycles. Experiments demonstrate the effectiveness of our algorithms.

1 INTRODUCTION

Multi-agent multi-armed bandits (MAMAB) have emerged as a powerful framework for modeling sequential decision-making in distributed systems, where multiple agents concurrently learn to interact with an unknown environment. Applications for MAMAB include dynamic spectrum access, recommender systems, and edge computing Darak & Hanawal (2019); Wang et al. (2024). In typical MAMAB problems, agents aim to maximize the summation of their expected rewards over T rounds, but this utilitarian objective can be unfair. Specifically, some agents have to sacrifice for the global objective by selecting arms with low rewards. This concern has motivated a growing line of fair bandit research, aiming to ensure equitable treatment across arms, agents, or time rounds Wu & Li (2024); Bistritz et al. (2020); Barman et al. (2023). Fairness among agents is especially important in resource allocation applications, such as edge networks and electric-vehicle charging, where leaving certain agents underserved is unacceptable.

A few recent works have studied fairness among agents in MAMAB, primarily adopting the Nash product of rewards or Nash social welfare (NSW) as the objective Hossain et al. (2021); Jones et al. (2023); Zhang et al. (2024). However, these studies have two critical limitations that hinder their applicability to many real-world scenarios. *First*, they ignore **limited arm capacities**, whereas real systems such as edge servers or charging stations can serve only a bounded number of agents simultaneously. *Second*, they rely on a **centralized decision maker** that assigns arms to all agents, while practical systems often require decentralized decisions under limited coordination.

In this paper, we bridge this gap by studying *capacity-constrained multi-agent bandits with a weighted α-fair objective*, designing both centralized and decentralized algorithms. We consider a system with M agents and N arms, where each arm has a limited capacity. At each round, agents form a many-to-one matching with the arms subject to limited arm capacities. Each agent m receives a stochastic reward whose mean $\mu_{mn} \in (0, 1]$ depends on the chosen arm n. Our objective is to maximize the long-term weighted α-fair utility: $F_\alpha(p, \mu) = \sum_m w_m U_\alpha(\sum_n \mu_{mn} p_{mn})$, where $U_\alpha(x) = \frac{x^{1-\alpha}}{1-\alpha}$ for $\alpha \neq 1$ and $U_1(x) = \log x$, p is a feasible capacity-constrained matching, and w_m's are positive weights summing to 1. The parameter $\alpha \geq 0$ controls the fairness-efficiency trade-off. Specifically, $\alpha = 0$ recovers the utilitarian sum, $\alpha = 1$ corresponds to proportional fairness, and $\alpha \to \infty$ approaches max-min fairness. We mainly focus on the regime $\alpha > 1$, where $U_\alpha(\cdot)$ becomes **non-Lipschitz** near zero, which is a significant technical challenge we address in this paper.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Our Novelty and Core Challenge: Non-Lipschitz nature of the α-fair utility function. Most prior studies of α-fairness in bandits restrict attention to the regime $0 \leq \alpha < 1$ Sinha et al. (2023); Chaudhary & Sinha (2024) or explicitly avoid the non-Lipschitz region by adding a reward floor Cayci et al. (2020). The case $\alpha = 1$ is closely related to weighted geometric-mean welfare Zhang et al. (2024). However, in our formulation, the non-Lipschitz nature of $U_\alpha(\cdot)$ for $\alpha > 1$ is the most fundamental obstacle. For any $x \in (0, 1]$, the derivative $U'_\alpha(x) = x^{-\alpha}$ diverges as $x \to 0$. This means that standard upper confidence bound (UCB) regret decomposition methods, which rely on a global Lipschitz constant to relate estimation error to utility loss, break down completely. Worse still, even in the offline setting, capacity constraints can force some agents to receive near-zero utilities in the optimal matching. Therefore, the non-Lipschitz behavior is not a corner case but a *structural property* of the problem that any algorithm must handle. While Zhang et al. (2024) also addresses non-Lipschitz for fair bandits, it exploits the boundedness of the objective (NSW lies in $[0, 1]$) to upper bound the regret in the small-utility regime. However, the α-fair utility for $\alpha > 1$ is *negative and unbounded* near zero, making such approaches inapplicable. Thus, a fundamentally new method is required to address the non-Lipschitz nature of the α-fair utility function.

Contributions. Our main contributions are as follows:

- **Centralized algorithm with regret upper and lower bounds.** We design a novel adaptive two-phase cover-elimination-UCB algorithm for the centralized setting. It employs an adaptive stopping rule and a global elimination rule to safely enter a Lipschitz regime. The regret consists of a capacity-coupled screening term for near-zero bad base-arms and a learning term in the Lipschitz regime. Informally, with $\mu^*_{\min}$ denoting the minimum optimal utility, $G_{\max}$ the worst exploratory loss, and Δ_{cov} the cover length capturing dependence on agents and capacities, the regret is $\widetilde{O}\left(\Delta_{\text{cov}} G_{\max}/\mu^*_{\min} + (\mu^*_{\min})^{\frac{1}{2}-\alpha}\sqrt{NT}\right)$. We prove a lower bound matching the T- and $\mu^*_{\min}$-dependence up to logarithmic factors.
- **Decentralized algorithm with low auction complexity.** We then propose a decentralized geometric shrinking algorithm with low auction complexity. At each epoch, the algorithm runs only one decentralized auction on the current active graph, and uses the resulting dual prices to compute forced-insertion gaps for all non-selected edges through shared shortest-path computations. This avoids solving one constrained auction per edge. This framework yields exact freezing on identifiable instances and admits a low-auction variant for general instances, leading to $O(\log \log T)$ auction calls with an instance-dependent regret guarantee.
- **Numerical experiments.** We conduct extensive numerical experiments. We compare our algorithms with previous works and demonstrate the effectiveness of our algorithms.

Related Works. Several works study MAMAB under different interaction models, including collision-based rewards Wang et al. (2020b;a), independent agent rewards Wang et al. (2023); Shao et al. (2025), and shared rewards when agents pull the same arm Wang et al. (2022a;b). However, all these works fail to consider the fairness issue.

Fairness has been studied across arms Wu & Li (2024); Wu et al. (2025); Sinha (2024), rounds Barman et al. (2023); Sawarni et al. (2023); Krishna et al. (2025), and agents Hossain et al. (2021); Jones et al. (2023); Zhang et al. (2024). For fair MAMAB, Hossain et al. (2021) proposes a centralized Nash-product framework, Jones et al. (2023) improves the regret upon Hossain et al. (2021), and Zhang et al. (2024) studies NSW (i.e., geometric mean), whose non-Lipschitz structure introduces additional challenges. However, these works assume centralized decisions, ignore arm-capacity constraints, and focus on Nash product or NSW, which are only specific instances of the α-fair utility family.

A few works have explored α-fair objectives in bandit problems, though not in the multi-agent capacity-constrained setting we consider. Sinha et al. (2023) studies no-regret resource allocation with an aggregate α-fair objective under adversarial arrivals and non-separable utilities, but restrict attention to the regime $0 \leq \alpha < 1$. Chaudhary & Sinha (2024) considers α-fair contextual bandits, also for $0 \leq \alpha < 1$, and explicitly introduces a minimum reward to keep utilities away from zero. Cayci et al. (2020) studies continuous-time scheduling with α-fair utilities under delayed feedback, again avoiding the non-Lipschitz regime.

Compared with all existing works above, in our problem, the combination of capacity constraints and the $\alpha > 1$ regime makes near-zero utilities *structurally unavoidable* even in the optimal matching.

Thus, the non-Lipschitz nature of $U_\alpha(\cdot)$ is an intrinsic property of the problem, requiring a fundamentally new algorithmic and analytic method. To tackle these challenges, we propose a novel adaptive two-phase cover-elimination UCB algorithm for the centralized setting. Furthermore, we design a decentralized geometric shrinking algorithm with low auction-call complexity.

2 Problem Formulation

We consider a stochastic MAMAB framework with M agents and N arms. Time proceeds in discrete rounds indexed by $t = 1, 2, \ldots, T$. Each agent $m \in \mathcal{M} \triangleq \{1, \ldots, M\}$ and each arm $n \in \mathcal{N} \triangleq \{1, \ldots, N\}$ are associated with an unknown mean reward $\mu_{mn} \in (0, 1]$. In our setting, each pair of agent and arm can be regarded as a *base-arm*. When agent m is matched to arm n at round t, it receives an i.i.d. reward $X_{mn}^t \in [0, 1]$ drawn from a distribution with $\mathbb{E}[X_{mn}^t] = \mu_{mn}$. The reward means are heterogeneous: different agents may have different means for the same arm.

Capacity-Constrained Matchings. Each arm $n \in \mathcal{N}$ has a limited capacity $C_n \in \mathbb{N}$, which is the maximum number of agents it can serve at a single round. A matching at round t is represented by a binary matrix $p^t \in \{0, 1\}^{M \times N}$, where $p_{mn}^t = 1$ indicates that agent m is assigned to arm n. The matching must satisfy the following feasibility constraints: i) $\sum_{n=1}^{N} p_{mn}^t = 1, \forall m \in \mathcal{M}$ and ii) $\sum_{m=1}^{M} p_{mn}^t \leq C_n, \forall n \in \mathcal{N}$. The first constraint ensures each agent is matched to exactly one arm, and the second constraint indicates limited arm capacities. To ensure feasibility, we assume $\sum_{n \in \mathcal{N}} C_n \geq M$ throughout. We denote $\mathcal{P}(\mathcal{M} \times \mathcal{N})$ as the set of all feasible matchings on the complete graph $\mathcal{M} \times \mathcal{N}$.

Weighted α-Fair Utility. Under a feasible matching p, the expected reward (utility) obtained by agent m is $u_m(p, \mu) \triangleq \sum_{n=1}^{N} \mu_{mn} p_{mn}$, where $\mu = (\mu_{mn})_{m \in \mathcal{M}, n \in \mathcal{N}}$ is the matrix of true means. Let each agent m be associated with a positive weight $w_m > 0$, normalized such that $\sum_{m=1}^{M} w_m = 1$. We denote $w_{\min} \triangleq \min_{m \in \mathcal{M}} w_m$. For a given $\alpha \geq 0$, the *weighted α-fair utility* of the system is $F_\alpha(p, \mu) \triangleq \sum_{m=1}^{M} w_m \, U_\alpha\big(u_m(p, \mu)\big)$, where $U_\alpha(\cdot)$ is the α-fair utility function given by $U_\alpha(x) = \frac{x^{1-\alpha}}{1-\alpha}$ for $\alpha \neq 1$. We define the extended-value convention $U_\alpha(0) = -\infty$ for $\alpha > 1$, and $U_\alpha^{-1}(-\infty) = 0$. We focus primarily on the regime $\alpha > 1$, where the utility function is *non-Lipschitz* near zero, which poses the central technical challenge.

Regret. If the mean reward μ were known a priori, the optimal strategy $\mathrm{Oracle}(\mathcal{M} \times \mathcal{N}, \mu)$ would repeatedly play a fixed feasible matching that maximizes the weighted α-fair utility: $p^* \in \arg\max_{p \in \mathcal{P}(\mathcal{M} \times \mathcal{N})} F_\alpha(p, \mu)$. We denote the minimum per-agent utility under this optimal matching by $\mu_{\min}^* \triangleq \min_{m \in \mathcal{M}} u_m(p^*, \mu)$. This quantity plays a key role in our setting, since it captures the "hardness" of the instance: when $\mu_{\min}^*$ is small, some agents receive near-zero utilities even under the optimal matching, forcing the algorithm to operate in the non-Lipschitz regime.

The objective is to minimize the expected cumulative regret over T rounds, which is the difference between the total weighted α-fair utility of the optimal strategy and that of the algorithm. Mathematically, $R(T) \triangleq \sum_{t=1}^{T} \Big(F_\alpha(p^*, \mu) - F_\alpha(p^t, \mu) \Big)$.

3 Centralized Algorithm: Adaptive Two-Phase Cover-Elimination-UCB

In this section, we present our algorithm for the centralized setting, where a single learner controls the matching at each round and observes all realized rewards. The centralized setting allows us to focus on and address the core technical challenge, i.e., the non-Lipschitz nature of α-fair utility.

The non-Lipschitz nature of $U_\alpha(\cdot)$ for $\alpha > 1$ poses a fundamental obstacle. As mentioned earlier, the non-Lipschitz nature arises not only from online learning but also from the offline problem structure (i.e., capacity coupling). On the one hand, due to capacity coupling, classic UCB algorithms can unavoidably assign some agents to arms with very small means and drive these agents' utilities close to zero. Moreover, from an analytic perspective, standard UCB-type regret decomposition

methods, which rely on a global Lipschitz constant, are completely inapplicable. Specifically, for standard bandits with Lipschitz objectives, the regret decomposes as $F_\alpha(p^*, \mu) - F_\alpha(p^t, \mu) \leq \ell \cdot \sum_{m,n} |\mathrm{UCB}^t_{mn} - \mu_{mn}|$, where ℓ is the Lipschitz constant. However, for $\alpha > 1$, the α-fair utility $U_\alpha(x)$ has derivative $x^{-\alpha}$, which diverges as $x \to 0$. Even worse, the offline optimal matching p^* itself may assign some agents extremely small utilities due to capacity constraints. When $\mu^*_{\min}$ is small, any near-optimal algorithm must operate in the non-Lipschitz regime, making standard UCB inapplicable. On the other hand, classic arm-wise elimination methods via confidence comparisons are unsafe. The reason is that low-mean base-arms can be in the offline optimal matching under capacity constraints, and such a direct elimination may remove the offline optimal base-arms.

To address the above challenges, we propose a novel adaptive two-phase cover-elimination-UCB algorithm, as shown in Algorithm 1. The key ideas are to first adaptively conduct **uniform exploration** over all base-arms until the confidence intervals of all base-arms are at a comparable rate (Lines 3-15). At the end of Phase I, Algorithm 1 implements **safe elimination via a global certificate** (Lines 16-18). This elimination removes very poor base-arms while preserving optimal base-arms. This yields that all surviving base-arms have means away from zero, and thus $U_\alpha(\cdot)$ becomes Lipschitz. Phase II then runs the standard UCB algorithm on surviving base-arms (Lines 19-23).

3.1 Phase I: Adaptive Cover Exploration

Cover Schedule via Edge Decomposition. To explore uniformly under capacity constraints, we first construct a set of feasible matchings that cover every base-arm. Given the complete bipartite graph $\mathcal{M} \times \mathcal{N}$ and arm capacities $\{C_n\}$, for each arm n, we partition its incident edges into C_n groups of size at most $\lceil M/C_n \rceil$ and regard these groups as incident to C_n unit-capacity copies of arm n. By Kőnig's theorem, the edges of this expanded graph can be decomposed into $\Delta(\mathcal{M} \times \mathcal{N})$ disjoint matchings, where $\Delta(\mathcal{M} \times \mathcal{N}) = \max\left\{N, \max_n \lceil M/C_n \rceil\right\}$. Playing these $\Delta(\mathcal{M} \times \mathcal{N})$ matchings in sequence guarantees that every base-arm (m, n) is selected exactly once per cycle. By repeating this schedule for multiple cycles, we achieve uniform exploration while satisfying arm capacity constraints. Algorithm 3 in Appendix B describes this construction.

The cover exploration may specify only a partial matching $\widetilde{p}$ to force certain agent–arm pairs to be sampled. We use $\mathrm{Complete}(\widetilde{p}, \{C_n\})$ to augment $\widetilde{p}$ into a full feasible matching. The details are in Algorithm 4 in Appendix C.

Data-Driven Stopping Rule. To obtain tight concentration when means are small, we develop an *estimate-dependent* confidence radius. After N^t_{mn} pulls of base-arm (m, n), let $\bar{\mu}^t_{mn}$ be the empirical average. The confidence radius and lower and upper confidence bounds are

$$r^t_{mn} = 3\sqrt{\frac{\bar{\mu}^t_{mn} \log(MNT^2)}{N^t_{mn}}}, \mathrm{LCB}^t_{mn} = \max\{\bar{\mu}^t_{mn} - r^t_{mn}, 0\}, \mathrm{UCB}^t_{mn} = \min\{\bar{\mu}^t_{mn} + r^t_{mn}, 1\}. \tag{1}$$

Note that this radius scales with $\sqrt{\bar{\mu}^t_{mn}}$ and enables a stronger concentration inequality that is essential for controlling regret when means are small. After each cover cycle (i.e., every $\Delta(\mathcal{M} \times \mathcal{N})$ rounds), we calculate a lower bound $\mathrm{LB}(t)$ on the optimal α-fair utility by solving the matching problem with the lower confidence bounds LCB^t_{mn} as the estimated means. We denote $\mathrm{LCB}^t \triangleq (\mathrm{LCB}^t_{mn})_{m \in \mathcal{M}, n \in \mathcal{N}}$ and $\mathrm{UCB}^t \triangleq (\mathrm{UCB}^t_{mn})_{m \in \mathcal{M}, n \in \mathcal{N}}$. Since $\mathrm{LCB}^t \leq \mu$ holds with high probability, $\mathrm{LB}(t)$ is a valid lower bound on the optimal utility. From $\mathrm{LB}(t)$, we derive $\underline{\mu}(t)$, a data-driven lower bound on the minimum per-agent utility, and set $\theta(t) = \underline{\mu}(t)/15$. The algorithm stops Phase I when every base-arm has been pulled at least $5\log(MNT^2)/\theta(t)$ times. At this round, denoted t_{se}, all confidence intervals are at a comparable rate, providing a basis for safe elimination. If t_{se} is not reached, terminate after the exploration phase.

3.2 Safe Elimination and UCB

Due to clear drawbacks of classic UCB and elimination methods in our problem, we propose a global safe elimination rule. Specifically, at round t_{se}, for each base-arm (m, n), we calculate an upper bound $\mathrm{UB}_{mn}(t_{\mathrm{se}})$ on the α-fair objective. Intuitively, this is an optimistic upper bound on the objective of any feasible matching containing (m, n). If $\mathrm{UB}_{mn}(t_{\mathrm{se}}) < \mathrm{LB}(t_{\mathrm{se}})$, then even under the

Algorithm 1 Adaptive Two-Phase Cover-Elimination-UCB Algorithm

Initialize $\{N_{mn}^0 = 0, \bar{\mu}_{mn}^0 = 0\}$ for all $(m,n) \in \mathcal{M} \times \mathcal{N}$, and set $t \leftarrow 0$;
Build Sched $\leftarrow$ BuildCover$(\mathcal{M} \times \mathcal{N})$ via Algorithm 3;
while $t \leq T$ **do**
  for $h = 1, \ldots, \Delta(\mathcal{M} \times \mathcal{N})$ **do**
    $t \leftarrow t+1$; $\widetilde{p}^t \leftarrow$ Sched$[h]$; $p^t \leftarrow$ Complete$(\widetilde{p}^t, \{C_n\})$ via Algorithm 4;
    Play p^t, observe rewards on matched pairs;
    Update $N_{mn}^t = N_{mn}^{t-1} + p_{mn}^t$ and $\bar{\mu}_{mn}^t = \frac{N_{mn}^{t-1}\bar{\mu}_{mn}^{t-1} + p_{mn}^t X_{mn}^t}{N_{mn}^t}$;
  end for
  Compute LCB_{mn}^t for all $(m,n) \in \mathcal{M} \times \mathcal{N}$ according to (1);
  $p^{\mathrm{LB}}(t) \leftarrow \mathrm{Oracle}(\mathcal{M} \times \mathcal{N}, \mathrm{LCB}^t)$, $\mathrm{LB}(t) \leftarrow F_\alpha(p^{\mathrm{LB}}(t), \mathrm{LCB}^t)$;
  $\underline{\mu}(t) \leftarrow U_\alpha^{-1}\left(\frac{\mathrm{LB}(t) - (1-w_{\min})U_\alpha(1)}{w_{\min}}\right)$, $\theta(t) \leftarrow \underline{\mu}(t)/15$, with $\underline{\mu}(t), \theta(t) = 0$ if $\mathrm{LB}(t) = -\infty$;
  if $\theta(t) > 0$ and $\min_{(m,n)\in\mathcal{M}\times\mathcal{N}} N_{mn}^t \geq \frac{5\log(MNT^2)}{\theta(t)}$ **then**
    $t_{\mathrm{se}} \leftarrow t$; **break**;
  end if
end while
Compute $\mathrm{UCB}_{mn}^{t_{\mathrm{se}}}$ for all $(m,n) \in \mathcal{M} \times \mathcal{N}$ according to (1) and set $\overline{\mathrm{UCB}}_m(t_{\mathrm{se}}) \leftarrow \max_{n\in\mathcal{N}} \mathrm{UCB}_{mn}^{t_{\mathrm{se}}}$ for all $m \in \mathcal{M}$;
Compute $\mathrm{UB}_{mn}(t_{\mathrm{se}}) \leftarrow w_m U_\alpha(\mathrm{UCB}_{mn}^{t_{\mathrm{se}}}) + \sum_{i\neq m} w_i U_\alpha(\overline{\mathrm{UCB}}_i(t_{\mathrm{se}}))$ for all $(m,n) \in \mathcal{M}\times\mathcal{N}$;
Eliminate: $\mathcal{A}_{\mathrm{se}} \leftarrow \{(m,n) \in \mathcal{M} \times \mathcal{N} : \mathrm{UB}_{mn}(t_{\mathrm{se}}) \geq \mathrm{LB}(t_{\mathrm{se}})\}$;
while $t \leq T$ **do**
  Update UCB_{mn}^t for all $(m,n) \in \mathcal{A}_{\mathrm{se}}$ according to (1);
  $p^t \leftarrow \mathrm{Oracle}(\mathcal{A}_{\mathrm{se}}, \mathrm{UCB}^t)$, play p^t, and observe rewards on matched pairs;
  Update $N_{mn}^{t+1} = N_{mn}^t + p_{mn}^t$ and $\bar{\mu}_{mn}^{t+1} = \frac{N_{mn}^t \bar{\mu}_{mn}^t + p_{mn}^t X_{mn}^t}{N_{mn}^{t+1}}$, then set $t \leftarrow t+1$;
end while

most optimistic situation, no feasible matching containing (m,n) can outperform the current lower bound. Hence, (m,n) is eliminated. We later show that the purpose of this rule is twofold. First, it removes extremely poor base-arms, which makes $U_\alpha(\cdot)$ Lipschitz. Second, it preserves offline optimal base-arms despite capacity coupling. After elimination, we derive a retained set $\mathcal{A}_{\mathrm{se}}$, in which all remaining base-arms have mean rewards bounded away from zero. Phase II then runs the standard UCB algorithm on $\mathcal{A}_{\mathrm{se}}$.

3.3 Regret Upper Bound and Lower Bound

We define the worst one-round exploratory loss $G_{\max}(\mu) \triangleq \max_{p\in\mathcal{P}(\mathcal{M}\times\mathcal{N})}\Big(F_\alpha(p^*, \mu) - F_\alpha(p, \mu)\Big)$. We show our main theoretical result for the centralized algorithm in the following theorem. We provide a proof sketch here, and defer the full proof to Appendix A.1.

Theorem 1. *For any $\alpha > 1$ and weight $w = (w_m)_{m\in\mathcal{M}}$, Algorithm 1 achieves expected regret $\mathbb{E}[R(T)] = \widetilde{O}\Big(\Delta(\mathcal{M}\times\mathcal{N})\, G_{\max}(\mu)/\mu_{\min}^* + (\mu_{\min}^*)^{\frac{1}{2}-\alpha}\sqrt{NT}\Big)$, where $\widetilde{O}(\cdot)$ hides logarithmic dependence on M, N, T and constants depending on α and w.*

Step 1: High-probability events. We first establish that the estimate-dependent confidence bounds yield three events with probability at least $1 - O(1/T)$:

- $\mathcal{E}_1$ **(LCB validity):** $\mathrm{LCB}_{mn}^t \leq \mu_{mn}$ for all (m,n,t).
- $\mathcal{E}_2$ **(Multiplicative concentration):** When $\mu_{mn} N_{mn}^t \geq 9\log(MNT^2)$, $|\bar{\mu}_{mn}^t - \mu_{mn}| \leq \sqrt{3\mu_{mn}\log(MNT^2)/N_{mn}^t}$.
- $\mathcal{E}_3$ **(Uniform small-mean control):** For all base-arms (m,n), all rounds t, and all $\beta \in (0,1]$, if $\mu_{mn} \leq \frac{9}{5}\beta$ and $N_{mn}^t \geq 5\log(MNT^2)/\beta$, then $\bar{\mu}_{mn}^t \leq 3\beta$.

These high-probability events are consequences of the multiplicative Chernoff bound and its extension. These events indicate two crucial properties. First, when a base-arm has been pulled enough times relative to its true mean, its empirical average is close to its true mean. Second, base-arms with very small means cannot have excessively large empirical averages once sufficiently sampled.

Let $\mathcal{E} \triangleq \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$. Conditioned on $\mathcal{E}$, we have the following results for the LCB and UCB values.

- When $\mu_{mn} N_{mn}^t \geq 75 \log(MNT^2)$, $\mathrm{LCB}_{mn}^t \geq c_{\mathrm{L}} \mu_{mn}$, where $c_{\mathrm{L}} = \frac{4}{5} - \sqrt{\frac{54}{375}} \geq 0.42$.
- If $\mu_{mn} \leq \frac{9}{5}\beta$ and $N_{mn}^t \geq \frac{5\log(MNT^2)}{\beta}$, then $\mathrm{UCB}_{mn}^t \leq c_{\mathrm{U}} \beta$, where $c_{\mathrm{U}} = 3 + 3\sqrt{\frac{3}{5}}$.

These two results are consequences of $\mathcal{E}_2$ and $\mathcal{E}_3$, respectively.

Step 2: Bounding Phase I stopping time. Under $\mathcal{E}$, we show that the stopping condition is satisfied within $t_{\mathrm{se}} \leq \Delta(\mathcal{M} \times \mathcal{N}) \cdot \lceil 5\log(MNT^2)/\beta^* \rceil$ rounds, where $\beta^* = \Theta(\mu_{\min}^*)$. The proof works by showing that once every base-arm has been pulled $\lceil 5\log(MNT^2)/\beta^* \rceil$ times, the optimal base-arms satisfy $\mathrm{LCB}_{mn}^t \geq c_{\mathrm{L}} \mu_{mn}$, which yields a sufficiently large $\theta(t)$ to trigger the stopping rule.

Step 3: Safe elimination and Lipschitz regime. At t_{se}, we prove:

- For any $(m,n) \in p^*$, we have $\mathrm{UB}_{mn}(t_{\mathrm{se}}) \geq \mathrm{LB}(t_{\mathrm{se}})$, so no optimal base-arm is eliminated.
- Any base-arm (m,n) with $\mu_{mn} \leq \frac{9}{5}\beta^*$ satisfies $\mathrm{UB}_{mn}(t_{\mathrm{se}}) < \mathrm{LB}(t_{\mathrm{se}})$ and is eliminated.

Therefore, all surviving base-arms satisfy $\mu_{mn} \geq \frac{9}{5}\beta^*$, and we denote $\underline{u} \triangleq \frac{9}{5}\beta^*$. For all surviving base-arms $\mu_{mn} \geq \underline{u}$, we prove that

$$\mu_{mn} \leq \mathrm{UCB}_{mn}^t \leq \mu_{mn} + \left(\sqrt{3} + \sqrt{9 + 3\sqrt{3}}\right) \sqrt{\frac{\mu_{mn} \log(MNT^2)}{N_{mn}^t}}. \tag{2}$$

Step 4: Phase II regret via Lipschitz decomposition. On the retained set $\mathcal{A}_{\mathrm{se}}$, every matching $p \in \mathcal{P}(\mathcal{A}_{\mathrm{se}})$ satisfies $u_m(p,\mu) \geq \underline{u}$ for all m. Thus, for any $t > t_{\mathrm{se}}$, using (2), we have

$$F_\alpha(p^*,\mu) - F_\alpha(p^t,\mu) \leq F_\alpha(p^t, \mathrm{UCB}^t) - F_\alpha(p^t,\mu) \leq \sum_{m=1}^{M} w_m \mu_{ma_m^t}^{-\alpha} \left(u_m(p^t, \mathrm{UCB}^t) - u_m(p^t,\mu)\right)$$

$$\leq O\left(\sum_{m=1}^{M} w_m \mu_{ma_m^t}^{-\alpha} \sqrt{\frac{\mu_{ma_m^t} \log(MNT^2)}{N_{ma_m}^t}}\right) \leq \widetilde{O}\left(\sum_{m=1}^{M} w_m \mu_{ma_m^t}^{\frac{1}{2}-\alpha} \sqrt{\frac{1}{N_{ma_m}^t}}\right),$$

where a_m^t is the arm assigned to agent m at round t. Plugging $\mu_{ma_m^t} \geq \underline{u}$, summing over $t > t_{\mathrm{se}}$ and using the standard inequality $\sum_t 1/\sqrt{N_{mn}^t} \leq 2\sqrt{N_{mn}^T}$ yields Phase II regret $R_{\mathrm{II}}(T) = \widetilde{O}\left(\underline{u}^{\frac{1}{2}-\alpha}\sqrt{NT}\right) = \widetilde{O}\left((\mu_{\min}^*)^{\frac{1}{2}-\alpha}\sqrt{NT}\right)$.

Step 5: Adding Phase I. Since Phase I lasts t_{se} rounds, its regret is at most $G_{\max}(\mu) t_{\mathrm{se}} = \widetilde{O}(\Delta(\mathcal{M} \times \mathcal{N}) G_{\max}(\mu)/\mu_{\min}^*)$. Combining two phases and the complementary event $\mathcal{E}^c$ gives Theorem 1.

Remark 1. *Compared with standard UCB and elimination methods, our approach has two key differences. First, the estimate-dependent confidence radius enables Chernoff-based concentration, which establishes $\mathcal{E}$ for bounding t_{se} and proving safe elimination. Replacing it with a Hoeffding radius $\sqrt{\log T / N_{mn}^t}$ would require a different analysis and degrade the Phase I condition from order $1/\theta(t)$ to $1/\theta(t)^2$, and the Phase II learning term from $(\mu_{\min}^*)^{\frac{1}{2}-\alpha}\sqrt{NT}$ to $(\mu_{\min}^*)^{-\alpha}\sqrt{NT}$. Second, the global elimination rule removes poor base-arms while preserving optimal ones, thereby ensuring a Lipschitz regime in Phase II.*

Compared with Zhang et al. (2024), which also addresses non-Lipschitz, our algorithmic and analytic techniques are fundamentally different. Specifically, the NSW objective considered in Zhang et al. (2024) is bounded within $[0,1]$ when individual utilities lie in $[0,1]$. After introducing a threshold σ, all rounds with some agent utility below σ can be controlled by boundedness, while the remaining rounds admit a Lipschitz-style analysis. In contrast, our weighted α-fair utility for $\alpha > 1$ is negative and unbounded when some agent receives utility near zero. Hence, one cannot simply keep all small-utility rounds in the regret decomposition and control them by a per-round trivial bound. This necessitates our two-phase design with data-driven safe-regime detection and global elimination.

Next, we establish a matching lower bound. We consider uniformly good policies, as defined in Lai & Robbins (1985). A policy π is uniformly good if, for every fixed Bernoulli instance ν with a unique optimal matching and every $a > 0$, $\mathbb{E}_\nu^\pi[R(T)] = o(T^a)$. The following theorem shows that the lower bound matches the dependence of the upper bound on T and $\mu_{\min}^*$ up to logarithmic factors. The detailed proof is in Appendix A.2.

Theorem 2. *Fix $\alpha > 1$ and $N \geq 4$. Let $B \triangleq \lfloor N/2 \rfloor$, $M = B$, and use uniform weights $w_m = 1/M$ and capacities $C_1 = M$, $C_n = 1$ for $2 \leq n \leq B+1$, $C_n = M$ for $B+2 \leq n \leq N$. For any $u \in (0, 1/8]$ and $\epsilon \in (0, u/4]$, there exists a family $\mathcal{V}(u, \epsilon)$ of Bernoulli instances with unique optimal matchings such that every $\nu \in \mathcal{V}(u, \epsilon)$ satisfies $\Delta(\mathcal{M} \times \mathcal{N}) = N$, $\mu^*_{\min} = u$, and $G_{\max}(\nu) = U_\alpha(u) - U_\alpha(\epsilon)$. For every uniformly good policy π and all sufficiently large T, $\sup_{\nu \in \mathcal{V}(u,\epsilon)} \mathbb{E}^\pi_\nu[R(T)] = \widetilde{\Omega}\Big(\Delta(\mathcal{M} \times \mathcal{N}) G_{\max}/\mu^*_{\min} + (\mu^*_{\min})^{\frac{1}{2}-\alpha}\sqrt{NT}\Big)$.*

4 Decentralized Algorithm: Geometric Shrinking via Insertion Gaps

In this section, we design a decentralized algorithm for the capacity-constrained α-fair bandit. Since the centralized algorithm already achieves a desired regret bound, the decentralized goal is not to redesign the learning rule or simply replace the oracle by an auction. Instead, we preserve the same safe-regime logic while reducing the number of auction calls from T (i.e., querying the auction at every round) to $O(\log \log T)$. The algorithm has two phases: a rare-auction warm-up phase that reaches the safe regime, and a Phase-II geometric shrinking procedure driven by insertion gaps.

Phase I Overview. Phase I serves to enter the same safe regime as the centralized algorithm. It uses decentralized exploration to obtain uniform sample counts, and calls the auction only at geometrically spaced checkpoints. At the first checkpoint satisfying the stopping rule, it returns t_{se}, the safe scale $u_{\text{se}} \triangleq \frac{9}{5}\theta_{\text{se}}$, and the retained edge set $\mathcal{A}_{\text{se}}$ after a global elimination. Since the checkpoints are geometric, Phase I uses only $O(\log \log T)$ auction calls. We defer the details to Appendix D.2.

Phase II Overview. Phase II is the main part of the decentralized design. Hereinafter, we refer to a base arm (m, n) as an edge e for simplicity. As shown in Algorithm 2, starting from $\mathcal{A}_0 \triangleq \mathcal{A}_{\text{se}}$, Phase II proceeds in epochs with geometrically growing sample targets $n_q = n_0 g^q$, where $g \geq 2$ and $n_0 = \min_{e \in \mathcal{A}_0} N_e(t_{\text{se}})$. Let t_q denote the time at the beginning of epoch q. At epoch q, the algorithm runs the decentralized auction (Algorithm 8) on the current active graph $\mathcal{A}_q$, uses the resulting dual prices to compute insertion gaps for all non-selected edges via a decentralized shortest-path algorithm (Algorithm 5) (Lines 5-6), and removes the edges whose observable gaps are sufficiently large (Line 11). The surviving edges are then refreshed (Lines 12-16). If, at some epoch, all non-selected edges have observable insertion gaps larger than the confidence threshold, the current matching is certified as the unique optimal matching on the current active graph, and the algorithm freezes (Lines 8-10). Otherwise, the shrinking-refresh loop continues.

Insertion Gaps and Augmented Alternating Cycles. We now explain how to compute insertion gaps in Algorithm 2 **without solving one constrained auction per edge**. At epoch q, define $\widehat{v}^{(q)}_{mn} = w_m U_\alpha(\text{UCB}_{mn}(t_q))$ and the estimated objective $\widehat{W}_q(p) \triangleq \sum_{(m,n) \in \mathcal{A}_q} \widehat{v}^{(q)}_{mn} p_{mn}, p \in \mathcal{P}(\mathcal{A}_q)$. The auction returns a matching $\widehat{p}_q$, arm prices $\lambda^{(q)}$, and agent net values $a^{(q)}_m \triangleq \max_{k \in \mathcal{N}_m(\mathcal{A}_q)} \big(\widehat{v}^{(q)}_{mk} - \lambda^{(q)}_k\big)$, where $\mathcal{N}_m(\mathcal{A}_q) \triangleq \{n : (m, n) \in \mathcal{A}_q\}$. Hence, $a^{(q)}_m$ is the best priced value available to agent m, while $\lambda^{(q)}_n$ is the scarcity price of arm n. The key quantity in Phase II is the insertion gap of a non-selected edge e, namely, the loss in objective value if e is forced into the current matching problem. Naively computing these gaps would require solving one constrained matching problem, e.g., calling one auction, for each non-selected edge. Instead, we compute them using the dual prices returned by the main auction. For an active edge (i, k), define its priced-value loss by $\widehat{r}^{(q)}_{ik} \triangleq a^{(q)}_i + \lambda^{(q)}_k - \widehat{v}^{(q)}_{ik} = \max_{\ell \in \mathcal{N}_i(\mathcal{A}_q)} \big(\widehat{v}^{(q)}_{i\ell} - \lambda^{(q)}_\ell\big) - \big(\widehat{v}^{(q)}_{ik} - \lambda^{(q)}_k\big) \geq 0$. It measures the loss in priced value if agent i is reassigned to arm k instead of its priced-best choice.

Fix a non-selected edge $e = (m, n)$. Forcing e moves agent m to arm n and frees its original slot at $n_m(\widehat{p}_q)$. If n is full, one of its currently matched agents must move to another arm, possibly displacing another agent. To describe both occupied and unused capacities within the same exchange, replace each arm k by slots $(k, 1), \ldots, (k, C_k)$ and fill unused slots with zero-value dummy agents. A displaced real agent can then replace either a real agent or a dummy. For example, a displaced dummy can occupy the slot vacated by m and close the exchange, or move to another occupied slot and continue the reallocation. Under this representation, $\widehat{p}_q$ and any feasible matching containing e

Algorithm 2 Phase-II: Geometric Shrinking via Insertion Gaps

Require: Horizon T, growth factor $g \geq 2$, and $(t_{\mathrm{se}}, u_{\mathrm{se}}, \mathcal{A}_{\mathrm{se}})$ returned by Algorithm 9.

Set $q \leftarrow 0$, $t \leftarrow t_{\mathrm{se}} + 1$, $\mathcal{A}_q \leftarrow \mathcal{A}_0 \leftarrow \mathcal{A}_{\mathrm{se}}$, and $n_0 \leftarrow \min_{e \in \mathcal{A}_0} N_e(t_{\mathrm{se}})$;
while $t \leq T$ **do**
  Set $t_q \leftarrow t - 1$ and $\varepsilon_q \leftarrow \left(\sqrt{3} + \sqrt{9 + 3\sqrt{3}}\right) u_{\mathrm{se}}^{\frac{1}{2}-\alpha} \sqrt{\log(MNT^2)/n_q}$;
  Compute $\widehat{v}_{mn}^{(q)} \leftarrow w_m U_\alpha(\mathrm{UCB}_{mn}(t_q))$ for all $(m,n) \in \mathcal{A}_q$;
  Run Algorithm 8 on $\mathcal{A}_q$ with $\widehat{v}^{(q)}$ and precision $\epsilon_{\mathrm{auc},q} \leq \varepsilon_q / M$, obtaining $(\widehat{p}_q, a^{(q)}, \lambda^{(q)})$;
  Run Algorithm 5 on $(\widehat{p}_q, a^{(q)}, \lambda^{(q)}, \widehat{v}^{(q)}, \mathcal{A}_q)$, obtaining $\widehat{\Delta}_{e,q}^{\mathrm{ins}}$ and $\widehat{p}_{q,e}^{\mathrm{ins}}$ for all $e \in \mathcal{A}_q \setminus \mathcal{X}(\widehat{p}_q)$;
  Set $\widehat{\Gamma}_q \leftarrow \min\{\widehat{\Delta}_{e,q}^{\mathrm{ins}} : e \in \mathcal{A}_q \setminus \mathcal{X}(\widehat{p}_q)\}$, with $\widehat{\Gamma}_q = +\infty$ if $\mathcal{A}_q = \mathcal{X}(\widehat{p}_q)$;
  if $\widehat{\Gamma}_q > 3\varepsilon_q$ **then**
    Certified freezing: set $p_{\mathrm{fr}} \leftarrow \widehat{p}_q$, replay p_{fr} in all remaining rounds, and **break**;
  end if
  $\mathcal{J}_q \leftarrow \{e \in \mathcal{A}_q \setminus \mathcal{X}(\widehat{p}_q) : \widehat{\Delta}_{e,q}^{\mathrm{ins}} \leq 3\varepsilon_q\}$, $\mathcal{A}_{q+1} \leftarrow \mathcal{X}(\widehat{p}_q) \cup \mathcal{J}_q$, $\mathcal{R}_q \leftarrow \{\widehat{p}_q\} \cup \{\widehat{p}_{q,e}^{\mathrm{ins}} : e \in \mathcal{J}_q\}$;
  for $c = 1, \ldots, (g-1)n_q$ **do**
    for each $p \in \mathcal{R}_q$ **do**
      Play p, update $N_{mn}^t, \bar{\mu}_{mn}^t$, and set $t \leftarrow t + 1$;
    end for
  end for
  Set $\mathcal{A}_q \leftarrow \mathcal{A}_{q+1}$, $n_{q+1} \leftarrow g n_q$, and $q \leftarrow q + 1$;
end while

extend to perfect one-to-one matchings. Their symmetric difference decomposes into vertex-disjoint alternating cycles. Applying only the cycle containing the slot edge corresponding to e preserves one assignment per agent and slot. Removing dummy agents and merging the slots yields a feasible matching containing e. We compute this exchange implicitly through arm and slack states.

For each agent i, let $n_i(\widehat{p}_q)$ denote its current arm. Define the matched agent set $H_n(\widehat{p}_q) \triangleq \{i \in \mathcal{M} : n_i(\widehat{p}_q) = n\}$ and the slack-arm set $S(\widehat{p}_q) \triangleq \{n \in \mathcal{N} : |H_n(\widehat{p}_q)| < C_n\}$. Fix the root m. After its forced move, let $D_{n|m}^{(d)}$ denote the minimum additional cost of completing the exchange from arm n, using at most d real-agent or dummy moves. Reaching $n_m(\widehat{p}_q)$ completes the exchange because m's original slot is available. With no further moves allowed, we therefore initialize

$$D_{n|m}^{(0)} = \begin{cases} 0, & n = n_m(\widehat{p}_q), \\ +\infty, & \text{otherwise.} \end{cases}$$

If the next displaced agent is a real agent $i \neq m$, assigning it to an admissible arm $k \neq n_i(\widehat{p}_q)$ incurs the local priced-value loss $\widehat{r}_{ik}^{(q)}$, followed by the remaining cost $D_{k|m}^{(d)}$. Minimizing over its possible destinations gives $B_{i|m}^{(d+1)} = \min_{k \in \mathcal{N}_i(\mathcal{A}_q) \setminus \{n_i(\widehat{p}_q)\}} \{\widehat{r}_{ik}^{(q)} + D_{k|m}^{(d)}\}$. We set $B_{m|m}^{(d+1)} = +\infty$, since the forced assignment of the root must remain unchanged.

At a slack arm, the exchange can instead displace a dummy occupying an unused slot. Moving this dummy to arm k has priced-value loss $\lambda_k^{(q)}$, because its value and its original slot's price are both zero. Its best continuation is $Z_m^{(d+1)} = \min_{k \in \mathcal{N}(\mathcal{A}_q)} \{\lambda_k^{(q)} + D_{k|m}^{(d)}\}$. Choosing $k = n_m(\widehat{p}_q)$ closes the exchange, whereas choosing another arm allows further reassignments. This message is shared across all slack arms because their dummy agents have identical values and available destinations.

Each arm now chooses whether to retain its previous continuation, displace a currently matched real agent, or, when slack, displace a dummy. This gives

$$D_{n|m}^{(d+1)} = \begin{cases} 0, & n = n_m(\widehat{p}_q), \\ \min\left\{ D_{n|m}^{(d)}, \min\limits_{i \in H_n(\widehat{p}_q) \setminus \{m\}} B_{i|m}^{(d+1)}, Z_m^{(d+1)} \right\}, & n \in S(\widehat{p}_q) \setminus \{n_m(\widehat{p}_q)\}, \\ \min\left\{ D_{n|m}^{(d)}, \min\limits_{i \in H_n(\widehat{p}_q) \setminus \{m\}} B_{i|m}^{(d+1)} \right\}, & \text{otherwise.} \end{cases}$$

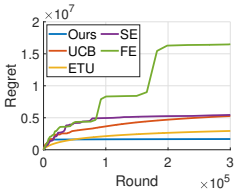


Figure 1: Regret with $M = 6, N = 6, \alpha = 8$

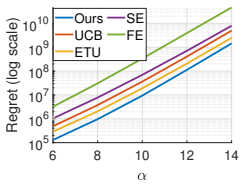


Figure 2: Regret for different α

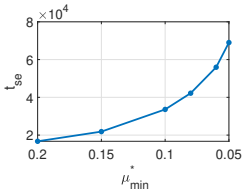


Figure 3: t_{se} for different $\mu^*_{\min}$

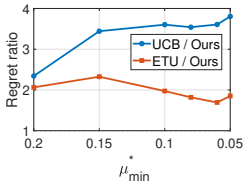


Figure 4: Regret ratio for different $\mu^*_{\min}$

Here, $\min \varnothing = +\infty$. The real-agent minimum selects which current occupant to reassign, while $Z_m^{(d+1)}$ supplies the additional option available at a slack arm. Retaining $D_{n|m}^{(d)}$ ensures that allowing more moves does not discard a less costly continuation found earlier. Algorithm 5 in Appendix D.1.1 describes the complete procedure.

The minimum-length cycle induces a feasible forced-insertion matching $\widehat{p}^{\mathrm{ins}}_{q,e}$, with observable insertion gap $\widehat{\Delta}^{\mathrm{ins}}_{e,q} = \widehat{W}_q(\widehat{p}_q) - \widehat{W}_q(\widehat{p}^{\mathrm{ins}}_{q,e})$. Let $\widetilde{\Delta}^{\mathrm{ins}}_{e,q} = \widehat{W}_q(\widehat{p}_q) - \max_{p \in \mathcal{P}(\mathcal{A}_q),\, p_e = 1} \widehat{W}_q(p)$ denote the exact estimated insertion loss. Using the alternating-cycle decomposition and the auction's $\epsilon_{\mathrm{auc},q}$-complementary slackness, we prove in Appendix D.1.1 that $\widehat{\Delta}^{\mathrm{ins}}_{e,q}$ is within $O(\varepsilon_q)$ of $\widetilde{\Delta}^{\mathrm{ins}}_{e,q}$. Together with the matching confidence bound, any edge satisfying $\widehat{\Delta}^{\mathrm{ins}}_{e,q} > 3\varepsilon_q$ can be safely eliminated. Hence, all insertion gaps are obtained through one shared decentralized algorithm, rather than one constrained auction per non-selected edge. This is the key complexity advantage of Phase II (see Proposition 6 in Appendix D.1.1).

We show the main result of the decentralized algorithm as follows. The proof is in Appendix D.1.3.

Theorem 3 (Decentralized regret and auction-call complexity)**.** *For each edge $e \in (\mathcal{M} \times \mathcal{N}) \setminus \mathcal{X}(p^*)$, let $\Delta^{\mathrm{ins}}_e(\mathcal{M} \times \mathcal{N})$ denote its true insertion gap on the complete graph. If the original problem admits a unique optimal matching p^*, the decentralized algorithm satisfies $\mathbb{E}[R(T)] = \widetilde{O}\left(\frac{\Delta(\mathcal{M}\times\mathcal{N}) G_{\max}(\mu)}{\mu^*_{\min}} + (\mu^*_{\min})^{1-2\alpha} \sum_{e \in (\mathcal{M}\times\mathcal{N})\setminus\mathcal{X}(p^*)} \frac{1}{\Delta^{\mathrm{ins}}_e(\mathcal{M}\times\mathcal{N})}\right)$. For every fixed instance, on $\mathcal{E}$, the algorithm triggers certified freezing after $O(\log\log T)$ epochs, and the number of auction calls is $O(\log\log T)$ on $\mathcal{E}$ and in expectation. After freezing, no regret or auction calls are incurred.*

For $O(\log\log T)$ auction calls on all instances, one can truncate Algorithm 2 at $Q_T = \min\{q : \varepsilon_q \le H_T^{-1}\}$, where $H_T = (\log T)^b, b > 0$, after the shrinking update, and then replay the current auction output. This yields $O(\log\log T)$ auction calls and an additional $O(T/H_T)$ tail-regret term.

5 Experiments

We compare our centralized algorithm with classic UCB, successive elimination (SE) Even-Dar et al. (2006), the fixed-length explore-then-UCB (ETU) Zhang et al. (2024), and fair elimination (FE) Wang et al. (2025), with $T = 3 \times 10^5$. Figure 1 plots the cumulative regret. Our algorithm achieves the lowest regret, since it explicitly detects a data-dependent safe scale and performs safe elimination. In contrast, UCB and ETU can leave near-zero bad arms insufficiently identified, SE may eliminate optimal base arms, and FE's max-min objective differs from weighted α-fairness. Figure 2 shows that the regret increases with α, reflecting the increasing non-Lipschitz difficulty, but our algorithm remains consistently better. Figures 3 and 4 show that, as $\mu^*_{\min}$ decreases, the safe-entry time t_{se} grows, while UCB and ETU can incur several times our regret. Additional experiments of centralized and decentralized algorithms are in Appendix E.

6 Conclusion

In this paper, we study capacity-constrained multi-agent bandits with a weighted α-fair objective in the non-Lipschitz regime $\alpha > 1$. We develop a centralized two-phase algorithm whose regret dependence on T and $\mu^*_{\min}$ is matched by the lower bound up to logarithmic factors. We further design a decentralized insertion-gap shrinking algorithm with $O(\log\log T)$ auction calls on identifiable instances and a truncation for general instances. Experiments demonstrate the effectiveness of both algorithms. Future work can extend the framework to time-varying settings.

AI USE STATEMENT

(This section is **required** and does not count toward the page limit.)

In this work, we used generative AI tools to polish the writing. The authors take full responsibility for the final manuscript.

ETHICS STATEMENT

(This section is **recommended** and does not count toward the page limit.)

We adhere to the ICLR Code of Ethics. Our work does not involve human subjects, raise discrimination concerns, or pose privacy and security risks.

REPRODUCIBILITY STATEMENT

(This section is **recommended** and does not count toward the page limit.)

The problem assumptions and algorithms are described in the main text, with detailed proofs and additional algorithmic procedures provided in the appendix. Additional experimental results are reported in Appendix E. We provide the code in the supplementary material.

REFERENCES

Siddharth Barman, Arindam Khan, Arnab Maiti, and Ayush Sawarni. Fairness and welfare quantification for regret in multi-armed bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 6762–6769, 2023.

Dimitri P Bertsekas and David A Castanon. The auction algorithm for the transportation problem. *Annals of Operations Research*, 20(1):67–96, 1989.

Ilai Bistritz, Tavor Baharav, Amir Leshem, and Nicholas Bambos. My fair bandit: Distributed learning of max-min fairness with multi-player bandits. In *International Conference on Machine Learning*, pp. 930–940, 2020.

Semih Cayci, Swati Gupta, and Atilla Eryilmaz. Group-fair online allocation in continuous time. In *Advances in Neural Information Processing Systems*, pp. 13750–13761, 2020.

Siddhant Chaudhary and Abhishek Sinha. alpha-fair contextual bandits. In *ICML 2024 Workshop: Foundations of Reinforcement Learning and Control – Connections and Perspectives*, 2024.

Sumit J. Darak and Manjesh K. Hanawal. Multi-player multi-armed bandits for stable allocation in heterogeneous ad-hoc networks. *IEEE Journal on Selected Areas in Communications*, 37(10): 2350–2363, 2019.

Devdatt P Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.

Eyal Even-Dar, Shie Mannor, Yishay Mansour, and Sridhar Mahadevan. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of Machine Learning Research*, 7(6), 2006.

Safwan Hossain, Evi Micha, and Nisarg Shah. Fair algorithms for multi-agent multi-armed bandits. In *Advances in Neural Information Processing Systems*, pp. 24005–24017, 2021.

Matthew Jones, Huy Nguyen, and Thy Nguyen. An efficient algorithm for fair multi-agent multi-armed bandit with low regret. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 8159–8167, 2023.

Anand Krishna, Philips George John, Adarsh Barik, and Vincent YF Tan. p-mean regret for stochastic bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 17966–17973, 2025.

Tze Leung Lai and Herbert Robbins. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, 6(1):4–22, 1985.

Ayush Sawarni, Soumyabrata Pal, and Siddharth Barman. Nash regret guarantees for linear bandits. In *Advances in Neural Information Processing Systems*, pp. 33288–33318, 2023.

Junning Shao, Siwei Wang, and Zhixuan Fang. Learning with limited shared information in multi-agent multi-armed bandit. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*, pp. 1914–1922, 2025.

Abhishek Sinha. Banditq: Fair bandits with guaranteed rewards. In *Proceedings of the Fortieth Conference on Uncertainty in Artificial Intelligence*, pp. 3227–3244, 2024.

Abhishek Sinha, Ativ Joshi, Rajarshi Bhattacharjee, Cameron Musco, and Mohammad Hajiesmaili. No-regret algorithms for fair resource allocation. In *Advances in Neural Information Processing Systems*, pp. 48083–48109, 2023.

Po-An Wang, Alexandre Proutiere, Kaito Ariu, Yassir Jedra, and Alessio Russo. Optimal algorithms for multiplayer multi-armed bandits. In *International Conference on Artificial Intelligence and Statistics*, pp. 4120–4129, 2020a.

Xiong Wang, Jiancheng Ye, and John C.S. Lui. Online learning aided decentralized multi-user task offloading for mobile edge computing. *IEEE Transactions on Mobile Computing*, 23(4): 3328–3342, 2024.

Xuchuang Wang, Hong Xie, and John C.S. Lui. Multiple-play stochastic bandits with shareable finite-capacity arms. In *International Conference on Machine Learning*, pp. 23181–23212, 2022a.

Xuchuang Wang, Hong Xie, and John C.S. Lui. Multi-player multi-armed bandits with finite shareable resources arms: Learning algorithms & applications. In *The 31st International Joint Conference on Artificial Intelligence*, 2022b.

Xuchuang Wang, Lin Yang, Yu-Zhen Janice Chen, Xutong Liu, Mohammad Hajiesmaili, Don Towsley, and John C.S. Lui. Achieving near-optimal individual regret & low communications in multi-agent bandits. In *International Conference on Learning Representations*, 2023.

Yuanhao Wang, Jiachen Hu, Xiaoyu Chen, and Liwei Wang. Distributed bandit learning: Near-optimal regret with efficient communication. In *International Conference on Learning Representations*, 2020b.

Zilong Wang, Zhiyao Zhang, and Shuai Li. Optimal algorithm for max-min fair bandit. In *International Conference on Machine Learning*, pp. 65480–65495, 2025.

Xiaoyi Wu and Bin Li. Achieving regular and fair learning in combinatorial multi-armed bandit. In *IEEE Conference on Computer Communications*, pp. 361–370, 2024.

Xiaoyi Wu, Bo Ji, and Bin Li. On the low-complexity of fair learning for combinatorial multi-armed bandit. In *IEEE Conference on Computer Communications*, pp. 1–10, 2025.

Mengxiao Zhang, Ramiro Deo-Campo Vuong, and Haipeng Luo. No-regret learning for fair multi-agent social welfare optimization. In *Advances in Neural Information Processing Systems*, pp. 57671–57700, 2024.

A Omitted Proofs in Section 3

In this section, we provide the omitted proofs for the results in Section 3. We provide the proof for Theorem 1 in Appendix A.1, and the proof for Theorem 2 in Appendix A.2.

A.1 Proof of Theorem 1

We begin by establishing the foundational concentration inequalities.

A.1.1 Concentration Inequalities

Lemma 1 (Chernoff bound Dubhashi & Panconesi (2009)). *Let $Y_1, \ldots, Y_n$ be independent random variables taking values in $[0,1]$, and define $\bar{Y} \triangleq \frac{1}{n}\sum_{i=1}^{n} Y_i$. Let $v = \mathbb{E}[\bar{Y}]$ denote the mean of $\bar{Y}$. Then, for any $0 \le \delta \le 1$, it holds that*

$$\mathbb{P}(|\bar{Y} - v| \ge \delta v) \le 2\exp\left(-\frac{\delta^2}{3} nv\right).$$

We next introduce a variant of the Chernoff bound.

Lemma 2 (Chernoff extension). *Let $Y_1, \ldots, Y_n$ be independent random variables taking values in $[0,1]$, and define $\bar{Y} \triangleq \frac{1}{n}\sum_{i=1}^{n} Y_i$. Assume that the mean of $\bar{Y}$ is upper bounded by $v_H > 0$, i.e., $\mathbb{E}[\bar{Y}] \le v_H$. Then, for any $0 \le \delta \le 1$, it holds that*

$$\mathbb{P}\big(\bar{Y} \ge (1+\delta)v_H\big) \;\le\; \exp\Big(-\frac{\delta^2}{3} nv_H\Big).$$

Proof. For any $\lambda > 0$, by Markov's inequality, we have

$$\begin{aligned}
\mathbb{P}\big(\bar{Y} \ge (1+\delta)v_H\big) &= \mathbb{P}\left(\sum_{i=1}^{n} Y_i \ge n(1+\delta)v_H\right) \\
&= \mathbb{P}\Big(\mathrm{e}^{\lambda\sum_{i=1}^{n} Y_i} \ge \mathrm{e}^{\lambda n(1+\delta)v_H}\Big) \\
&\le \mathrm{e}^{-\lambda n(1+\delta)v_H}\,\mathbb{E}\Big[\mathrm{e}^{\lambda\sum_{i=1}^{n} Y_i}\Big].
\end{aligned}$$

By independence of Y_i's for all i,

$$\mathbb{E}\Big[\mathrm{e}^{\lambda\sum_{i=1}^{n} Y_i}\Big] = \prod_{i=1}^{n}\mathbb{E}\big[\mathrm{e}^{\lambda Y_i}\big].$$

Since $Y_i \in [0,1]$, for any $y \in [0,1]$, the convexity of the exponential function implies

$$\mathrm{e}^{\lambda y} \le (1-y)\mathrm{e}^0 + y\mathrm{e}^{\lambda} = 1 + y(\mathrm{e}^{\lambda} - 1).$$

Taking expectations yields

$$\mathbb{E}\big[\mathrm{e}^{\lambda Y_i}\big] \le 1 + (\mathrm{e}^{\lambda} - 1)\mathbb{E}[Y_i] \le \exp\big((\mathrm{e}^{\lambda} - 1)\mathbb{E}[Y_i]\big),$$

which is due to $1 + x \le \mathrm{e}^x$ for all $x \in \mathbb{R}$. Hence,

$$\prod_{i=1}^{n}\mathbb{E}\big[\mathrm{e}^{\lambda Y_i}\big] \le \exp\Big((\mathrm{e}^{\lambda} - 1)\sum_{i=1}^{n}\mathbb{E}[Y_i]\Big).$$

Let $\mu \triangleq \mathbb{E}[\bar{Y}] = \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[Y_i]$. Since $\mu \le v_H$, we have

$$\sum_{i=1}^{n}\mathbb{E}[Y_i] = n\mu \le nv_H.$$

Combining the above bounds, we obtain for any $\lambda > 0$:

$$\begin{aligned}
\mathbb{P}\big(\bar{Y} \ge (1+\delta)v_H\big) &\le \exp\Big(-\lambda n(1+\delta)v_H\Big)\,\exp\Big((\mathrm{e}^{\lambda} - 1)nv_H\Big) \\
&= \exp\Big(-nv_H\, g(\lambda, \delta)\Big),
\end{aligned}$$

where $g(\lambda, \delta) \triangleq (1+\delta)\lambda - (\mathrm{e}^{\lambda} - 1)$.

To make this bound as tight as possible, we maximize $g(\lambda, \delta)$ over $\lambda > 0$ for a fixed δ. Differentiating with respect to λ gives

$$\frac{\partial g}{\partial \lambda} = (1+\delta) - \mathrm{e}^{\lambda}.$$

Setting this derivative to zero yields

$$\mathrm{e}^{\lambda} = 1 + \delta \quad \Rightarrow \quad \lambda^* = \ln(1+\delta).$$

Substituting λ^* into $g(\lambda, \delta)$, we obtain

$$\begin{aligned} g(\lambda^*, \delta) &= (1+\delta)\ln(1+\delta) - \big((1+\delta) - 1\big) \\ &= (1+\delta)\ln(1+\delta) - \delta. \end{aligned}$$

Therefore,

$$\mathbb{P}\big(\bar{Y} \geq (1+\delta)v_H\big) \leq \exp\Big(- nv_H\big[(1+\delta)\ln(1+\delta) - \delta\big]\Big).$$

For $x \in [0,1]$, the function $\log(1+x) - x + x^2/2$ is nondecreasing, because its derivative is $x^2/(1+x) \geq 0$. Thus, $\log(1+x) \geq x - x^2/2$. Consequently, for $\delta \in [0,1]$,

$$(1+\delta)\log(1+\delta) - \delta = \int_0^{\delta} \log(1+x)\,\mathrm{d}x \geq \frac{\delta^2}{2} - \frac{\delta^3}{6} \geq \frac{\delta^2}{3}.$$

Substituting this inequality into the preceding bound gives

$$\mathbb{P}\big(\bar{Y} \geq (1+\delta)v_H\big) \leq \exp\left(-\frac{\delta^2}{3} n v_H\right).$$

The case $\delta = 0$ also follows directly from the trivial probability upper bound of one. This completes the proof. □

A.1.2 High-Probability Events

We establish concentration events that hold uniformly over sample counts, and then evaluate them at the random counts generated by the learning algorithm.

For each base-arm $e = (m, n)$, realize its rewards as an i.i.d. sequence

$$X_{e,1}, X_{e,2}, \ldots,$$

with values in $[0,1]$ and mean $\mu_e = \mu_{mn}$. The variable $X_{e,s}$ is revealed when e is selected for the s-th time. For every deterministic $s \geq 1$, define

$$\widehat{\mu}_{e,s} \triangleq \frac{1}{s}\sum_{\ell=1}^{s} X_{e,\ell},$$

$$\mathrm{LCB}_e(s) \triangleq \max\left\{\widehat{\mu}_{e,s} - 3\sqrt{\frac{\widehat{\mu}_{e,s}\log(MNT^2)}{s}}, 0\right\}.$$

Set $\mathrm{LCB}_e(0) = 0$. Pathwise, whenever $N_{mn}^t > 0$,

$$\bar{\mu}_{mn}^t = \widehat{\mu}_{e,N_{mn}^t}, \qquad \mathrm{LCB}_{mn}^t = \mathrm{LCB}_e(N_{mn}^t).$$

We first apply concentration inequalities to deterministic sample counts. The random counts are substituted only after constructing events that hold simultaneously for all $s \leq T$.

Lemma 3 (Event 1: LCB validity)**.** *Define*

$$\mathcal{E}_1 \triangleq \left\{\forall (m,n) \in \mathcal{M} \times \mathcal{N},\ \forall t \leq T : \mathrm{LCB}_{mn}^t \leq \mu_{mn}\right\}.$$

Then

$$\mathbb{P}(\mathcal{E}_1) \geq 1 - \frac{2}{T}.$$

Proof. Fix a base-arm e and a deterministic sample count $s \in \{1, \ldots, T\}$. Write

$$\mu \triangleq \mu_e, \qquad \widehat{\mu} \triangleq \widehat{\mu}_{e,s}, \qquad \ell_T \triangleq \log(MNT^2).$$

Since $\mu > 0$,

$$\mathrm{LCB}_e(s) > \mu \quad \Longleftrightarrow \quad \widehat{\mu} - 3\sqrt{\frac{\widehat{\mu}\ell_T}{s}} > \mu.$$

For bounded i.i.d. rewards, Bernstein's inequality gives

$$\mathbb{P}(\widehat{\mu} - \mu > x) \le \exp\left(-\frac{sx^2}{2(\mu + x/3)}\right), \qquad x > 0.$$

Case 1: $\mu s \ge \ell_T$**.** If $\mathrm{LCB}_e(s) > \mu$, then $\widehat{\mu} > \mu$ and

$$\widehat{\mu} - \mu > 3\sqrt{\frac{\widehat{\mu}\ell_T}{s}} \ge 3\sqrt{\frac{\mu\ell_T}{s}}.$$

Set $x = 3\sqrt{\mu\ell_T/s}$. The condition $\mu s \ge \ell_T$ implies $\mu + x/3 \le 2\mu$. Therefore,

$$\mathbb{P}(\mathrm{LCB}_e(s) > \mu) \le \exp\left(-\frac{9}{4}\ell_T\right) \le \frac{1}{MNT^2}.$$

Case 2: $\mu s < \ell_T$**.** If $\mathrm{LCB}_e(s) > \mu$, then $\mathrm{LCB}_e(s) > 0$, which implies

$$\widehat{\mu} > \frac{9\ell_T}{s}.$$

Consequently,

$$\widehat{\mu} - \mu > \frac{8\ell_T}{s}.$$

Applying Bernstein's inequality with $x = 8\ell_T/s$ and using $\mu < \ell_T/s$ gives

$$\mathbb{P}(\mathrm{LCB}_e(s) > \mu) \le \exp\left(-\frac{s(8\ell_T/s)^2}{2(\ell_T/s + 8\ell_T/(3s))}\right) = \exp\left(-\frac{96}{11}\ell_T\right) \le \frac{1}{MNT^2}.$$

Thus, for every deterministic pair (e, s),

$$\mathbb{P}(\mathrm{LCB}_e(s) > \mu_e) \le \frac{2}{MNT^2}.$$

A union bound over all MN base-arms and all $s \in \{1, \dots, T\}$ yields

$$\mathbb{P}\left(\exists e, \exists s \le T : \mathrm{LCB}_e(s) > \mu_e\right) \le \frac{2}{T}.$$

On the complementary event, substituting $s = N_{mn}^t$ is valid simultaneously for every $t \le T$. The case $N_{mn}^t = 0$ is covered by $\mathrm{LCB}_e(0) = 0$. This proves the claim. □

Lemma 4 (Event 2: Multiplicative concentration)**.** *Define*

$$\mathcal{E}_2 \triangleq \left\{ \forall (m,n) \in \mathcal{M} \times \mathcal{N}, \forall t \le T \text{ with } \mu_{mn} N_{mn}^t \ge 9\log(MNT^2) : |\bar{\mu}_{mn}^t - \mu_{mn}| \le \sqrt{\tfrac{3\mu_{mn}\log(MNT^2)}{N_{mn}^t}} \right\}.$$

Then

$$\mathbb{P}(\mathcal{E}_2) \ge 1 - \frac{2}{T}.$$

Proof. Fix a base-arm e and a deterministic sample count $s \le T$ satisfying $\mu_e s \ge 9\log(MNT^2)$. Set

$$\delta \triangleq \sqrt{\frac{3\log(MNT^2)}{\mu_e s}} \le \frac{1}{\sqrt{3}} < 1.$$

Lemma 1 gives

$$\mathbb{P}\left(|\widehat{\mu}_{e,s} - \mu_e| > \sqrt{\frac{3\mu_e\log(MNT^2)}{s}}\right) \le 2\exp\left(-\frac{\delta^2}{3}s\mu_e\right) = \frac{2}{MNT^2}.$$

A union bound over all base-arms and all deterministic sample counts satisfying the stated condition gives failure probability at most $2/T$. On the resulting simultaneous event, substitute $s = N_{mn}^t$ and use $\bar{\mu}_{mn}^t = \widehat{\mu}_{e,N_{mn}^t}$. This yields $\mathcal{E}_2$ for every $t \le T$, including adaptive sampling times and stopping times bounded by T. □

Lemma 5 (Event 3: Uniform small-mean control)**.** *Use the pull-indexed rewards and empirical means defined above. For $e = (m, n)$ and $s \geq 1$,*

$$\widehat{\mu}_{(m,n),s} = \frac{1}{s} \sum_{\ell=1}^{s} X_{(m,n),\ell},$$

and $\bar{\mu}^t_{mn} = \widehat{\mu}_{(m,n),N^t_{mn}}$ whenever $N^t_{mn} > 0$.

Define the event

$$\mathcal{E}_3 \triangleq \left\{ \begin{array}{l} \forall (m,n) \in \mathcal{M} \times \mathcal{N},\ \forall s \in \{1, \dots, T\},\ \forall \beta \in (0,1] : \\ \mu_{mn} \leq \frac{9}{5}\beta \ \text{ and } \ s \geq \frac{5\log(MNT^2)}{\beta} \Longrightarrow \widehat{\mu}_{(m,n),s} \leq 3\beta \end{array} \right\}.$$

Then $\mathbb{P}(\mathcal{E}_3) \geq 1 - \frac{1}{T}$.

Proof. Fix a base-arm (m, n) and a pull count $s \in \{1, \dots, T\}$. Define $b_{(m,n),s} \triangleq \max\left\{\frac{5}{9}\mu_{mn}, \frac{5\log(MNT^2)}{s}\right\}$. Then,

$$\mu_{mn} \leq \frac{9}{5} b_{(m,n),s}, \quad s \cdot \frac{9}{5} b_{(m,n),s} \geq 9\log(MNT^2).$$

Apply Lemma 2 with $v_H = \frac{9}{5} b_{(m,n),s}, \delta = \frac{2}{3}$. Since $(1+\delta)v_H = \frac{5}{3} \cdot \frac{9}{5} b_{(m,n),s} = 3b_{(m,n),s}$, we obtain

$$\mathbb{P}\big(\widehat{\mu}_{(m,n),s} \geq 3b_{(m,n),s}\big) \leq \exp\left(-\frac{\delta^2}{3} s v_H\right) \leq \exp\left(-\frac{4}{27} \cdot 9\log(MNT^2)\right) = \exp\left(-\frac{4}{3}\log(MNT^2)\right).$$

Taking a union bound over all MN base-arms and all $s \in \{1, \dots, T\}$ gives

$$\mathbb{P}\big(\exists (m,n), s : \widehat{\mu}_{(m,n),s} \geq 3b_{(m,n),s}\big) \leq MNT \exp\left(-\frac{4}{3}\log(MNT^2)\right) \leq \frac{1}{T}.$$

Therefore, with probability at least $1 - 1/T$, for every base-arm e and every pull count s, $\widehat{\mu}_{(m,n),s} < 3b_{(m,n),s}$.

Now take any $\beta \in (0, 1]$ such that $\mu_{mn} \leq \frac{9}{5}\beta, s \geq \frac{5\log(MNT^2)}{\beta}$. These two inequalities imply $\frac{5}{9}\mu_{mn} \leq \beta, \frac{5\log(MNT^2)}{s} \leq \beta$. Hence,

$$b_{e,s} = \max\left\{\frac{5}{9}\mu_{mn}, \frac{5\log(MNT^2)}{s}\right\} \leq \beta.$$

Thus,

$$\widehat{\mu}_{(m,n),s} < 3b_{(m,n),s} \leq 3\beta.$$

This proves the claim. □

In the remainder of the analysis, we condition on the intersection event

$$\mathcal{E} \triangleq \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3,$$

which holds with probability at least $1 - \frac{5}{T}$.

A.1.3 Confidence Bound Properties

Define the numerical constants

$$c_{\mathrm{L}} \triangleq \frac{4}{5} - \sqrt{\frac{54}{375}} \geq 0.42, \qquad c_{\mathrm{U}} \triangleq 3 + 3\sqrt{\frac{3}{5}}.$$

Lemma 6 (LCB lower bound)**.** *Conditioned on $\mathcal{E}$, for any (m, n, t) such that $\mu_{mn} N^t_{mn} \geq 75\log(MNT^2)$,*

$$\mathrm{LCB}^t_{mn} \geq c_{\mathrm{L}} \mu_{mn}.$$

Proof. Event $\mathcal{E}_2$ implies $\bar{\mu}^t_{mn} \geq \mu_{mn} - \sqrt{\frac{3\mu_{mn}\log(MNT^2)}{N^t_{mn}}}$. Since $\mu_{mn} N^t_{mn} \geq 75\log(MNT^2)$, we have $\sqrt{\frac{3\mu_{mn}\log(MNT^2)}{N^t_{mn}}} \leq \frac{1}{5}\mu_{mn}$, hence $\bar{\mu}^t_{mn} \geq \frac{4}{5}\mu_{mn}$. Also, Event $\mathcal{E}_2$ implies $\bar{\mu}^t_{mn} \leq$

$\mu_{mn} + \frac{1}{5}\mu_{mn} = \frac{6}{5}\mu_{mn}$. Therefore,

$$\mathrm{LCB}^t_{mn} = \bar{\mu}^t_{mn} - 3\sqrt{\frac{\bar{\mu}^t_{mn}\log(MNT^2)}{N^t_{mn}}} \geq \frac{4}{5}\mu_{mn} - 3\sqrt{\frac{\frac{6}{5}\mu_{mn}\log(MNT^2)}{N^t_{mn}}} \geq \left(\frac{4}{5} - \sqrt{\frac{54}{375}}\right)\mu_{mn},$$

which proves the claim. □

Lemma 7 (UCB upper bound for small means)**.** *Conditioned on $\mathcal{E}$, for any base-arm (m,n), any time t, and any possibly data-dependent random variable $\beta_t \in (0,1]$, if*

$$\mu_{mn} \leq \frac{9}{5}\beta_t \qquad \textit{and} \qquad N^t_{mn} \geq \frac{5\log(MNT^2)}{\beta_t},$$

then

$$\mathrm{UCB}^t_{mn} \leq c_{\mathrm{U}}\beta_t, \qquad c_{\mathrm{U}} = 3 + 3\sqrt{\frac{3}{5}}.$$

Proof. Let $e = (m,n)$ and $s = N^t_{mn}$. Since $\mathcal{E}_3$ holds simultaneously for all $\beta \in (0,1]$, it can be applied pathwise with the random choice $\beta = \beta_t$. Therefore, $\bar{\mu}^t_{mn} = \widehat{\mu}_{e,s} \leq 3\beta_t$. Using the definition of the UCB index and the condition $N^t_{mn} \geq 5\log(MNT^2)/\beta_t$, we obtain

$$\mathrm{UCB}^t_{mn} \leq \bar{\mu}^t_{mn} + 3\sqrt{\frac{\bar{\mu}^t_{mn}\log(MNT^2)}{N^t_{mn}}} \leq 3\beta_t + 3\sqrt{\frac{3\beta_t\log(MNT^2)}{5\log(MNT^2)/\beta_t}} = \left(3 + 3\sqrt{\frac{3}{5}}\right)\beta_t.$$

□

Lemma 8 (Confidence interval coverage)**.** *Conditioned on $\mathcal{E}$, for any (m,n,t) such that $\mu_{mn}N^t_{mn} \geq 9\log(MNT^2)$,*

$$\mu_{mn} \leq \mathrm{UCB}^t_{mn} \leq \mu_{mn} + \left(\sqrt{3} + \sqrt{9 + 3\sqrt{3}}\right)\sqrt{\frac{\mu_{mn}\log(MNT^2)}{N^t_{mn}}}.$$

Proof. Event $\mathcal{E}_2$ yields $\bar{\mu}^t_{mn} \geq \mu_{mn} - \sqrt{\frac{3\mu_{mn}\log(MNT^2)}{N^t_{mn}}}$ and

$$\bar{\mu}^t_{mn} \geq \mu_{mn} - \sqrt{\frac{3\mu_{mn}\log(MNT^2)}{N^t_{mn}}} \geq \left(1 - \sqrt{\frac{1}{3}}\right)\mu_{mn},$$

where the last inequality uses $\mu_{mn}N^t_{mn} \geq 9\log(MNT^2)$. Thus,

$$\begin{aligned}\mathrm{UCB}^t_{mn} = \bar{\mu}^t_{mn} + 3\sqrt{\frac{\bar{\mu}^t_{mn}\log(MNT^2)}{N^t_{mn}}} &\geq \mu_{mn} - \sqrt{\frac{3\mu_{mn}\log(MNT^2)}{N^t_{mn}}} + 3\sqrt{\frac{\left(1 - \sqrt{\frac{1}{3}}\right)\mu_{mn}\log(MNT^2)}{N^t_{mn}}} \\ &\geq \mu_{mn}.\end{aligned}$$

On the other hand, event $\mathcal{E}_2$ also gives $\bar{\mu}^t_{mn} \leq \mu_{mn} + \sqrt{\frac{3\mu_{mn}\log(MNT^2)}{N^t_{mn}}}$ and $\bar{\mu}^t_{mn} \leq \left(1 + \sqrt{\frac{1}{3}}\right)\mu_{mn}$. Therefore,

$$\begin{aligned}\mathrm{UCB}^t_{mn} = \bar{\mu}^t_{mn} + 3\sqrt{\frac{\bar{\mu}^t_{mn}\log(MNT^2)}{N^t_{mn}}} &\leq \mu_{mn} + \sqrt{\frac{3\mu_{mn}\log(MNT^2)}{N^t_{mn}}} + 3\sqrt{\frac{\left(1 + \sqrt{\frac{1}{3}}\right)\mu_{mn}\log(MNT^2)}{N^t_{mn}}} \\ &\leq \mu_{mn} + \left(\sqrt{3} + \sqrt{9 + 3\sqrt{3}}\right)\sqrt{\frac{\mu_{mn}\log(MNT^2)}{N^t_{mn}}}.\end{aligned}$$

This completes the proof. □

A.1.4 Bounding the Phase I Stopping Time

We now bound the phase I stopping time t_{se}.

Proposition 1 (Upper bound on t_{se})**.** *Conditioned on $\mathcal{E}$, the stopping time t_{se} satisfies*

$$t_{\mathrm{se}} \le t^\dagger \triangleq \Delta(\mathcal{M}\times\mathcal{N}) \cdot \left\lceil \frac{5\log(MNT^2)}{\beta^*} \right\rceil,$$

where

$$\beta^* \triangleq \frac{1}{15}\left(\frac{(c_{\mathrm{L}}\mu^*_{\min})^{1-\alpha} - (1-w_{\min})}{w_{\min}}\right)^{\frac{1}{1-\alpha}} = \Theta(\mu^*_{\min}).$$

Here, $\Theta(\cdot)$ hides constants depending on α and weight $w = (w_m)_{m\in\mathcal{M}}$.

Proof. Let $N_B \triangleq \left\lceil \frac{5\log(MNT^2)}{\beta^*} \right\rceil$ and $t^\dagger = \Delta(\mathcal{M}\times\mathcal{N})\cdot N_B$. By the cover property, at time $t^\dagger$, every base-arm has been pulled at least N_B times, i.e., $\min_{(m,n)} N^{t^\dagger}_{mn} \ge N_B$.

We first note that $\beta^* \le c_{\mathrm{L}}\mu^*_{\min}/15$. Let $n^*(m)$ denote the arm matched to agent m in the optimal matching p^*. For each optimal base-arm $(m, n^*(m))$, we have $\mu_{mn^*(m)} \ge \mu^*_{\min}$. Moreover,

$$\mu_{mn^*(m)} N^{t^\dagger}_{mn^*(m)} \ge \mu^*_{\min} N_B \ge \mu^*_{\min}\cdot\frac{5\log(MNT^2)}{\beta^*} \ge \mu^*_{\min}\cdot\frac{5\log(MNT^2)}{c_{\mathrm{L}}\mu^*_{\min}/15} \ge 75\log(MNT^2).$$

By Lemma 6, we obtain $\mathrm{LCB}^{t^\dagger}_{mn^*(m)} \ge c_{\mathrm{L}}\mu_{mn^*(m)} \ge c_{\mathrm{L}}\mu^*_{\min}$ for all m. Consequently,

$$\mathrm{LB}(t^\dagger) = \max_{p\in\mathcal{P}} F_\alpha(p, \mathrm{LCB}^{t^\dagger}) \ge F_\alpha(p^*, \mathrm{LCB}^{t^\dagger}) \ge \sum_{m=1}^{M} w_m U_\alpha(c_{\mathrm{L}}\mu^*_{\min}) = U_\alpha(c_{\mathrm{L}}\mu^*_{\min}).$$

By definition of $\underline{\mu}(\cdot)$,

$$\underline{\mu}(t^\dagger) = U_\alpha^{-1}\left(\frac{\mathrm{LB}(t^\dagger) - (1-w_{\min})U_\alpha(1)}{w_{\min}}\right) \ge \left(\frac{(c_{\mathrm{L}}\mu^*_{\min})^{1-\alpha} - (1-w_{\min})}{w_{\min}}\right)^{\frac{1}{1-\alpha}} = 15\beta^*.$$

Thus, $\theta(t^\dagger) = \underline{\mu}(t^\dagger)/15 \ge \beta^*$. Therefore,

$$\frac{5\log(MNT^2)}{\theta(t^\dagger)} \le \frac{5\log(MNT^2)}{\beta^*} \le N_B \le \min_{(m,n)} N^{t^\dagger}_{mn},$$

so the stopping condition is satisfied by $t^\dagger$. Hence, $t_{\mathrm{se}} \le t^\dagger$. □

A.1.5 Safe Elimination Properties

In this part, we first prove that no optimal base-arm is eliminated by the global elimination rule in Algorithm 1. We then show that all base-arms with means below than a threshold are eliminated, so surviving base-arms have means bounded away from zero.

Proposition 2 (No optimal base-arm eliminated)**.** *Conditioned on $\mathcal{E}$, for every base-arm (m,n) belonging to the true optimal matching p^*, we have $\mathrm{UB}_{mn}(t_{\mathrm{se}}) \ge \mathrm{LB}(t_{\mathrm{se}})$, so no optimal base-arm is eliminated.*

Proof. Let $\mathrm{OPT} \triangleq F_\alpha(p^*, \mu)$. Because $U_\alpha(\cdot)$ is increasing and $\mathrm{LCB}^t \le \mu$ entrywise by $\mathcal{E}_1$, we have

$$\mathrm{LB}(t_{\mathrm{se}}) = \max_{p\in\mathcal{P}(\mathcal{M}\times\mathcal{N})} F_\alpha(p, \mathrm{LCB}^{t_{\mathrm{se}}}) \le \max_{p\in\mathcal{P}(\mathcal{M}\times\mathcal{N})} F_\alpha(p,\mu) = \mathrm{OPT}.$$

On the other hand, let $m_0 \in \arg\min_{i\in\mathcal{M}} u_i(p^*,\mu)$, so that $u_{m_0}(p^*,\mu) = \mu^*_{\min}$. Since $U_\alpha(\cdot)$ is increasing and $u_i(p^*,\mu) \le 1$ for all i, we have

$$\begin{aligned}\mathrm{OPT} = w_{m_0}U_\alpha(\mu^*_{\min}) + \sum_{i\ne m_0} w_i U_\alpha(u_i(p^*,\mu)) &\le w_{m_0}U_\alpha(\mu^*_{\min}) + (1-w_{m_0})U_\alpha(1)\\ &\le w_{\min}U_\alpha(\mu^*_{\min}) + (1-w_{\min})U_\alpha(1),\end{aligned}$$

where the last inequality follows from $w_{m_0} \ge w_{\min}$ and $U_\alpha(\mu^*_{\min}) \le U_\alpha(1)$. Together with $\mathrm{LB}(t_{\mathrm{se}}) \le \mathrm{OPT}$, this gives

$$\mathrm{LB}(t_{\mathrm{se}}) \le (1-w_{\min})U_\alpha(1) + w_{\min}U_\alpha(\mu^*_{\min}).$$

By the definition of $\underline{\mu}(\cdot)$,

$$U_\alpha(\underline{\mu}(t_{\text{se}})) = \frac{\text{LB}(t_{\text{se}}) - (1 - w_{\min})U_\alpha(1)}{w_{\min}} \le U_\alpha(\mu^*_{\min}).$$

Since $U_\alpha(\cdot)$ is increasing, $\underline{\mu}(t_{\text{se}}) \le \mu^*_{\min}$, and hence $\mu^*_{\min} \ge 15\theta(t_{\text{se}})$.

By the stopping rule, $N^{t_{\text{se}}}_{mn} \ge 5\log(MNT^2)/\theta(t_{\text{se}})$ for all (m,n). Let $n^*(m)$ denote the arm matched to agent m in the optimal matching p^*. Therefore, for any optimal base-arm $(m, n^*(m))$, we have $\mu_{mn^*(m)} \ge \mu^*_{\min} \ge 15\theta(t_{\text{se}})$, so

$$\mu_{mn^*(m)} N^{t_{\text{se}}}_{mn^*(m)} \ge 15\theta(t_{\text{se}}) \cdot \frac{5\log(MNT^2)}{\theta(t_{\text{se}})} = 75\log(MNT^2) \ge 9\log(MNT^2).$$

By Lemma 8, $\text{UCB}^{t_{\text{se}}}_{mn^*(m)} \ge \mu_{mn^*(m)}$.

Now fix any $m \in \mathcal{M}$ and consider the upper bound:

$$\text{UB}_{mn^*(m)}(t_{\text{se}}) = w_m U_\alpha(\text{UCB}^{t_{\text{se}}}_{mn^*(m)}) + \sum_{i \ne m} w_i U_\alpha(\overline{\text{UCB}}_i(t_{\text{se}})),$$

where $\overline{\text{UCB}}_i(t_{\text{se}}) = \max_{k \in \mathcal{N}} \text{UCB}^{t_{\text{se}}}_{ik}$. Since $\overline{\text{UCB}}_i(t_{\text{se}}) \ge \text{UCB}^{t_{\text{se}}}_{in^*(i)}$ for all $i \ne m$, we obtain

$$\text{UB}_{mn^*(m)}(t_{\text{se}}) \ge \sum_{i=1}^{M} w_i U_\alpha(\text{UCB}^{t_{\text{se}}}_{in^*(i)}) = F_\alpha(p^*, \text{UCB}^{t_{\text{se}}}).$$

By Lemma 8 and monotonicity of U_α, $F_\alpha(p^*, \text{UCB}^{t_{\text{se}}}) \ge F_\alpha(p^*, \mu) = \text{OPT} \ge \text{LB}(t_{\text{se}})$. Thus, $\text{UB}_{mn^*(m)}(t_{\text{se}}) \ge \text{LB}(t_{\text{se}})$, and the optimal base-arm $(m, n^*(m))$ is not eliminated. □

Proposition 3 (Bad arms elimination)**.** *For each $m \in \mathcal{M}$, define*

$$\underline{\mu}^{\text{row}}_m(t_{\text{se}}) \triangleq U_\alpha^{-1}\left(\frac{\text{LB}(t_{\text{se}}) - \sum_{i \ne m} w_i U_\alpha(\overline{\text{UCB}}_i(t_{\text{se}}))}{w_m}\right), \qquad \theta^{\text{row}}_m(t_{\text{se}}) \triangleq \frac{\underline{\mu}^{\text{row}}_m(t_{\text{se}})}{15}.$$

Conditioned on $\mathcal{E}$, at time t_{se}, every base-arm (m,n) satisfying $\mu_{mn} \le \frac{9}{5}\,\theta^{\text{row}}_m(t_{\text{se}})$ is eliminated. In particular, every base-arm with $\mu_{mn} \le \frac{9}{5}\,\beta^*$ is eliminated.

Proof. Denote $\theta_{\text{se}} \triangleq \theta(t_{\text{se}})$. By the stopping rule, for all (m,n),

$$N^{t_{\text{se}}}_{mn} \ge \frac{5\log(MNT^2)}{\theta_{\text{se}}}. \tag{3}$$

We first record two basic facts about the random scales. Since the stopping rule is triggered at t_{se}, we have $\theta_{\text{se}} > 0$. Moreover, since every entry of $\text{LCB}^{t_{\text{se}}}$ is at most one and $U_\alpha(\cdot)$ is increasing, $\text{LB}(t_{\text{se}}) \le U_\alpha(1)$. By the definition of $\underline{\mu}(t_{\text{se}}) = 15\theta_{\text{se}}$,

$$U_\alpha(15\theta_{\text{se}}) = \frac{\text{LB}(t_{\text{se}}) - (1 - w_{\min})U_\alpha(1)}{w_{\min}} \le U_\alpha(1).$$

Since $U_\alpha(\cdot)$ is strictly increasing, this implies $15\theta_{\text{se}} \le 1$.

We next show that for every agent m, $\theta^{\text{row}}_m(t_{\text{se}}) \ge \theta_{\text{se}}$. Since $\overline{\text{UCB}}_i(t_{\text{se}}) \le 1$ and $U_\alpha(\cdot)$ is increasing,

$$\sum_{i \ne m} w_i U_\alpha(\overline{\text{UCB}}_i(t_{\text{se}})) \le \sum_{i \ne m} w_i U_\alpha(1) = (1 - w_m)U_\alpha(1).$$

Moreover, by the definition of $\underline{\mu}(t)$ and $\theta(t) = \underline{\mu}(t)/15$,

$$\text{LB}(t_{\text{se}}) = (1 - w_{\min})U_\alpha(1) + w_{\min}U_\alpha(15\theta_{\text{se}}).$$

Therefore,

$$\begin{aligned}
\text{LB}(t_{\text{se}}) - \sum_{i \ne m} w_i U_\alpha(\overline{\text{UCB}}_i(t_{\text{se}})) &\ge (1 - w_{\min})U_\alpha(1) + w_{\min}U_\alpha(15\theta_{\text{se}}) - (1 - w_m)U_\alpha(1) \\
&= (w_m - w_{\min})U_\alpha(1) + w_{\min}U_\alpha(15\theta_{\text{se}}) \\
&\ge (w_m - w_{\min})U_\alpha(15\theta_{\text{se}}) + w_{\min}U_\alpha(15\theta_{\text{se}}) \\
&= w_m U_\alpha(15\theta_{\text{se}}),
\end{aligned}$$

where the second inequality uses $15\theta_{\mathrm{se}} \le 1$ and hence $U_\alpha(1) \ge U_\alpha(15\theta_{\mathrm{se}})$. Dividing by w_m and applying the monotonicity of U_α^{-1} yields $\underline{\mu}_m^{\mathrm{row}}(t_{\mathrm{se}}) \ge 15\theta_{\mathrm{se}}$, and hence $\theta_m^{\mathrm{row}}(t_{\mathrm{se}}) \ge \theta_{\mathrm{se}} > 0$.

We also show that $\theta_m^{\mathrm{row}}(t_{\mathrm{se}}) \le 1/15$. Let p^{LB} be a maximizer in the definition of $\mathrm{LB}(t_{\mathrm{se}})$, and let $n_{p^{\mathrm{LB}}}(i)$ be the arm assigned to agent i under p^{LB}. Then

$$\mathrm{LB}(t_{\mathrm{se}}) = \sum_{i=1}^{M} w_i U_\alpha\Big(\mathrm{LCB}_{i,n_{p^{\mathrm{LB}}}(i)}^{t_{\mathrm{se}}}\Big).$$

Since

$$\mathrm{LCB}_{i,n_{p^{\mathrm{LB}}}(i)}^{t_{\mathrm{se}}} \le \mathrm{UCB}_{i,n_{p^{\mathrm{LB}}}(i)}^{t_{\mathrm{se}}} \le \overline{\mathrm{UCB}}_i(t_{\mathrm{se}}),$$

and $U_\alpha(\cdot)$ is increasing,

$$\mathrm{LB}(t_{\mathrm{se}}) \le \sum_{i=1}^{M} w_i U_\alpha(\overline{\mathrm{UCB}}_i(t_{\mathrm{se}})).$$

Therefore,

$$\frac{\mathrm{LB}(t_{\mathrm{se}}) - \sum_{i\neq m} w_i U_\alpha(\overline{\mathrm{UCB}}_i(t_{\mathrm{se}}))}{w_m} \le U_\alpha(\overline{\mathrm{UCB}}_m(t_{\mathrm{se}})) \le U_\alpha(1).$$

Applying U_α^{-1} gives

$$\underline{\mu}_m^{\mathrm{row}}(t_{\mathrm{se}}) \le \overline{\mathrm{UCB}}_m(t_{\mathrm{se}}) \le 1,$$

and hence

$$0 < \theta_m^{\mathrm{row}}(t_{\mathrm{se}}) \le \frac{1}{15} < 1.$$

Now fix an agent m and a base-arm (m,n) such that $\mu_{mn} \le \frac{9}{5}\,\theta_m^{\mathrm{row}}(t_{\mathrm{se}})$. Since

$$\theta_m^{\mathrm{row}}(t_{\mathrm{se}}) \ge \theta_{\mathrm{se}},$$

(3) implies

$$N_{mn}^{t_{\mathrm{se}}} \ge \frac{5\log(MNT^2)}{\theta_{\mathrm{se}}} \ge \frac{5\log(MNT^2)}{\theta_m^{\mathrm{row}}(t_{\mathrm{se}})}.$$

Since Lemma 7 is uniform in the threshold parameter, we apply it with the random choice

$$\beta_t = \theta_m^{\mathrm{row}}(t_{\mathrm{se}}).$$

Thus,

$$\mathrm{UCB}_{mn}^{t_{\mathrm{se}}} \le c_{\mathrm{U}}\theta_m^{\mathrm{row}}(t_{\mathrm{se}}).$$

Therefore,

$$\mathrm{UB}_{mn}(t_{\mathrm{se}}) = w_m U_\alpha(\mathrm{UCB}_{mn}^{t_{\mathrm{se}}}) + \sum_{i\neq m} w_i U_\alpha(\overline{\mathrm{UCB}}_i(t_{\mathrm{se}})) \le w_m U_\alpha(c_{\mathrm{U}}\theta_m^{\mathrm{row}}(t_{\mathrm{se}})) + \sum_{i\neq m} w_i U_\alpha(\overline{\mathrm{UCB}}_i(t_{\mathrm{se}})).$$

On the other hand, by the definition of $\underline{\mu}_m^{\mathrm{row}}(t_{\mathrm{se}})$ and $\theta_m^{\mathrm{row}}(t_{\mathrm{se}}) = \underline{\mu}_m^{\mathrm{row}}(t_{\mathrm{se}})/15$,

$$\mathrm{LB}(t_{\mathrm{se}}) = \sum_{i\neq m} w_i U_\alpha(\overline{\mathrm{UCB}}_i(t_{\mathrm{se}})) + w_m U_\alpha\big(15\theta_m^{\mathrm{row}}(t_{\mathrm{se}})\big).$$

Since $U_\alpha(\cdot)$ is strictly increasing and $c_{\mathrm{U}} = 3+3\sqrt{\frac{3}{5}} < 15$, we obtain $\mathrm{UB}_{mn}(t_{\mathrm{se}}) < \mathrm{LB}(t_{\mathrm{se}})$. Hence, (m,n) is eliminated.

It remains to show that $\theta_{\mathrm{se}} \ge \beta^*$. Suppose, for contradiction, that $\theta_{\mathrm{se}} < \beta^*$. By the stopping rule, for every base-arm (m,n),

$$N_{mn}^{t_{\mathrm{se}}} \ge \frac{5\log(MNT^2)}{\theta_{\mathrm{se}}} > \frac{5\log(MNT^2)}{\beta^*}.$$

In particular, for every optimal base-arm $(m, n^*(m))$,

$$\mu_{mn^*(m)} N_{mn^*(m)}^{t_{\mathrm{se}}} \ge \mu_{\min}^* \frac{5\log(MNT^2)}{\beta^*}.$$

By the definition of β^*, we have $\beta^* \le \frac{c_{\mathrm{L}}\mu_{\min}^*}{15}$. Hence,

$$\mu_{mn^*(m)} N_{mn^*(m)}^{t_{\mathrm{se}}} \ge \frac{75}{c_{\mathrm{L}}}\log(MNT^2) \ge 75\log(MNT^2).$$

Therefore, Lemma 6 gives $\mathrm{LCB}^{t_{\mathrm{se}}}_{mn^*(m)} \geq c_{\mathrm{L}}\mu_{mn^*(m)} \geq c_{\mathrm{L}}\mu^*_{\min}$. Thus,

$$\mathrm{LB}(t_{\mathrm{se}}) \geq F_\alpha(p^*, \mathrm{LCB}^{t_{\mathrm{se}}}) \geq U_\alpha(c_{\mathrm{L}}\mu^*_{\min}).$$

By the definition of β^*,

$$U_\alpha(c_{\mathrm{L}}\mu^*_{\min}) = (1 - w_{\min})U_\alpha(1) + w_{\min}U_\alpha(15\beta^*).$$

Hence,

$$\mathrm{LB}(t_{\mathrm{se}}) \geq (1 - w_{\min})U_\alpha(1) + w_{\min}U_\alpha(15\beta^*).$$

Using the definition of $\underline{\mu}(\cdot)$,

$$U_\alpha(\underline{\mu}(t_{\mathrm{se}})) = \frac{\mathrm{LB}(t_{\mathrm{se}}) - (1 - w_{\min})U_\alpha(1)}{w_{\min}} \geq U_\alpha(15\beta^*).$$

Since U_α is increasing, we obtain $\underline{\mu}(t_{\mathrm{se}}) \geq 15\beta^*, \theta_{\mathrm{se}} = \underline{\mu}(t_{\mathrm{se}})/15 \geq \beta^*$, which contradicts $\theta_{\mathrm{se}} < \beta^*$. Therefore, $\theta_{\mathrm{se}} \geq \beta^*$.

Since we have shown $\theta^{\mathrm{row}}_m(t_{\mathrm{se}}) \geq \theta_{\mathrm{se}}$, it follows that

$$\theta^{\mathrm{row}}_m(t_{\mathrm{se}}) \geq \beta^*, \qquad \forall m \in \mathcal{M}.$$

Therefore, every base-arm (m, n) with $\mu_{mn} \leq \frac{9}{5}\beta^*$ also satisfies $\mu_{mn} \leq \frac{9}{5}\,\theta^{\mathrm{row}}_m(t_{\mathrm{se}})$, and is thus eliminated. □

A.1.6 Regret Analysis

We first bound the per-round regret in Phase II. We set $\underline{u} \triangleq \frac{9}{5}\beta^*$, and thus $\underline{u} = \Theta(\mu^*_{\min})$.

Lemma 9 (Phase II per-round regret)**.** *Conditioned on $\mathcal{E}$, for any $t > t_{\mathrm{se}}$,*

$$F_\alpha(p^*, \mu) - F_\alpha(p^t, \mu) \leq \sum_{m=1}^{M} \big(u_m(p^t, \mu)\big)^{-\alpha} w_m \Big(u_m(p^t, \mathrm{UCB}^t) - u_m(p^t, \mu)\Big).$$

Proof. By Proposition 2, no optimal base-arm is eliminated. By Proposition 3, all surviving base-arms have

$$\mu_{mn} \geq \frac{9}{5}\theta^{\mathrm{row}}_m(t_{\mathrm{se}}) \geq \frac{9}{5}\theta_{\mathrm{se}}.$$

Together with the stopping rule

$$N^{t_{\mathrm{se}}}_{mn} \geq \frac{5\log(MNT^2)}{\theta_{\mathrm{se}}},$$

we have $\mu_{mn}N^t_{mn} \geq 9\log(MNT^2)$ for all surviving base-arms. Thus, Lemma 8 gives $\mathrm{UCB}^t_{mn} \geq \mu_{mn}$. Thus,

$$F_\alpha(p^*, \mu) \leq F_\alpha(p^*, \mathrm{UCB}^t) \leq F_\alpha(p^t, \mathrm{UCB}^t).$$

Therefore,

$$\begin{aligned}
F_\alpha(p^*, \mu) - F_\alpha(p^t, \mu) &\leq F_\alpha(p^*, \mathrm{UCB}^t) - F_\alpha(p^t, \mu) \\
&\leq F_\alpha(p^t, \mathrm{UCB}^t) - F_\alpha(p^t, \mu) \\
&= \sum_{m=1}^{M} w_m \Big(U_\alpha\big(u_m(p^t, \mathrm{UCB}^t)\big) - U_\alpha\big(u_m(p^t, \mu)\big)\Big) \\
&\leq \sum_{m=1}^{M} \big(u_m(p^t, \mu)\big)^{-\alpha} w_m \Big(u_m(p^t, \mathrm{UCB}^t) - u_m(p^t, \mu)\Big).
\end{aligned}$$

□

We now restate and prove Theorem 1 as follows.

Theorem 4. *For $\alpha > 1$, Algorithm 1 achieves expected regret* $\mathbb{E}[R(T)] = \widetilde{O}\Big(\Delta(\mathcal{M} \times \mathcal{N})\, G_{\max}(\mu)/\mu^*_{\min} + (\mu^*_{\min})^{\frac{1}{2}-\alpha}\sqrt{NT}\Big)$.

Proof. Decompose $R(T) = R_{\mathrm{I}}(T) + R_{\mathrm{II}}(T)$.

Phase I: The per-round regret is bounded by $G_{\max}(\mu) = \max_{p\in\mathcal{P}(\mathcal{M}\times\mathcal{N})}(F_\alpha(p^*,\mu) - F_\alpha(p,\mu))$. By Proposition 1, $t_{\text{se}} \le \Delta(\mathcal{M}\times\mathcal{N})\cdot O(\log T/\beta^*) = \widetilde{O}(1/(\mu^*_{\min}))$. Thus

$$R_{\text{I}}(T) \le G_{\max}(\mu) t_{\text{se}} = \widetilde{O}\left(\Delta(\mathcal{M}\times\mathcal{N}) G_{\max}(\mu)/(\mu^*_{\min})\right).$$

Phase II: By Lemma 9,

$$\begin{aligned}
R_{\text{II}}(T) &\le \sum_{t=t_{\text{se}}+1}^{T}\sum_{m=1}^{M}\left(u_m(p^t,\mu)\right)^{-\alpha} w_m\left(u_m(p^t,\text{UCB}^t) - u_m(p^t,\mu)\right) \\
&\overset{(a)}{\le} O\left(\sum_{t=t_{\text{se}}+1}^{T}\sum_{m=1}^{M} w_m \mu_{ma_m^t}^{-\alpha}\sqrt{\frac{\mu_{ma_m^t}\log(MNT^2)}{N^t_{ma_m}}}\right) \\
&\le \widetilde{O}\left(\sum_{t=t_{\text{se}}+1}^{T}\sum_{m=1}^{M} w_m \mu_{ma_m^t}^{\frac{1}{2}-\alpha}\sqrt{\frac{1}{N^t_{ma_m}}}\right) \\
&= \widetilde{O}\left(\underline{u}^{\frac{1}{2}-\alpha}\sum_{m=1}^{M} w_m \sum_{n=1}^{N}\sum_{s=1}^{N^T_{mn}}\sqrt{\frac{1}{s}}\right) \\
&\le \widetilde{O}\left(\underline{u}^{\frac{1}{2}-\alpha}\sum_{m=1}^{M} w_m \sum_{n=1}^{N}\int_0^{N^T_{mn}}\sqrt{\frac{1}{s}}\mathrm{d}s\right) \\
&= \widetilde{O}\left(\underline{u}^{\frac{1}{2}-\alpha}\sum_{m=1}^{M} w_m \sum_{n=1}^{N}\sqrt{N^T_{mn}}\right) \\
&\overset{(b)}{\le} \widetilde{O}\left(\underline{u}^{\frac{1}{2}-\alpha}\sum_{m=1}^{M} w_m \sqrt{N\sum_{n=1}^{N} N^T_{mn}}\right) \\
&\le \widetilde{O}\left(\underline{u}^{\frac{1}{2}-\alpha}\sum_{m=1}^{M} w_m \sqrt{NT}\right) \\
&= \widetilde{O}\left(\underline{u}^{\frac{1}{2}-\alpha}\sqrt{NT}\right),
\end{aligned}$$

where step (a) is by Lemma 8 and a_m^t is the arm matched to agent m at round t, and step (b) is due to Cauchy-Schwarz inequality.

Plugging $\underline{u} = \Theta(\mu^*_{\min})$ and combining both phases and the complementary event $\mathcal{E}^c$ complete the proof. □

A.2 Proof of Theorem 2

We restate and prove Theorem 2 as follows.

Theorem 5. *Fix $\alpha > 1$, $N \ge 4$, $u \in (0, 1/8]$, and $\epsilon \in (0, u/4]$. Let $B \triangleq \lfloor N/2 \rfloor$, $M = B$, and $w_m = 1/M$. Consider the capacities*

$$C_n = \begin{cases} 1, & 2 \le n \le B+1, \\ M, & n = 1 \text{ or } B+2 \le n \le N. \end{cases}$$

There exists a family $\mathcal{V}(u,\epsilon)$ of Bernoulli instances such that every $\nu \in \mathcal{V}(u,\epsilon)$ has a unique optimal matching and satisfies

$$\Delta(\mathcal{M}\times\mathcal{N}) = N, \qquad \mu^*_{\min} = u, \qquad G_{\max}(\nu) = U_\alpha(u) - U_\alpha(\epsilon).$$

For every uniformly good policy π, there exists $T_0 = T_0(\pi, N, \alpha, u, \epsilon)$ such that, for every $T \ge \max\{T_0, N/u\}$,

$$\sup_{\nu\in\mathcal{V}(u,\epsilon)} \mathbb{E}^\pi_\nu[R(T)] = \widetilde{\Omega}\left(\frac{\Delta(\mathcal{M}\times\mathcal{N})G_{\max}}{\mu^*_{\min}} + (\mu^*_{\min})^{\frac{1}{2}-\alpha}\sqrt{NT}\right).$$

Proof. Define

$$\mathcal{H} \triangleq \{1, B+2, \dots, N\}, \qquad K \triangleq |\mathcal{H}| = N - B.$$

For every $\boldsymbol{\theta} = (\theta_1, \dots, \theta_M) \in \mathcal{H}^M$ and $\delta \in (0, u/4]$, define $\nu_{\boldsymbol{\theta},\delta}$ by

$$\mu_{mn} = \begin{cases} u, & n = \theta_m, \\ u - \delta, & n \in \mathcal{H} \setminus \{\theta_m\}, \\ \epsilon, & 2 \le n \le B+1. \end{cases}$$

Rewards are independent Bernoulli random variables across base-arms and pulls. Let

$$\mathcal{V}(u, \epsilon) \triangleq \{\nu_{\boldsymbol{\theta},\delta} : \boldsymbol{\theta} \in \mathcal{H}^M,\ 0 < \delta \le u/4\}.$$

Since every arm in $\mathcal{H}$ has capacity M, the matching $m \mapsto \theta_m$ is feasible and uniquely optimal. Thus, every instance in this family satisfies $\mu^*_{\min} = u$. Moreover,

$$\Delta(\mathcal{M} \times \mathcal{N}) = \max\left\{N, \max_n \left\lceil \frac{M}{C_n} \right\rceil\right\} = N.$$

There are $B = M$ unit-capacity arms with mean ϵ, so a feasible matching can assign every agent to a distinct such arm. Consequently, with

$$G \triangleq U_\alpha(u) - U_\alpha(\epsilon),$$

we have $G_{\max}(\nu) = G$ for every $\nu \in \mathcal{V}(u, \epsilon)$.

Step 1: Fixed-instance screening lower bound. Fix $\delta \in (0, u/4]$ independently of T, and let

$$\nu_{0,\delta} \triangleq \nu_{(1,\dots,1),\delta}.$$

Its unique optimal matching assigns every agent to arm 1.

For a bad edge (m, n) with $2 \le n \le B+1$, change only its mean from ϵ to $2u$. In the resulting alternative instance, the unique optimal matching assigns m to n and every other agent to arm 1. Lemma 10 therefore gives

$$\liminf_{T\to\infty} \frac{\mathbb{E}^\pi_{\nu_{0,\delta}}[N_{mn}(T)]}{\log T} \ge \frac{1}{\mathrm{kl}(\epsilon, 2u)}.$$

For Bernoulli relative entropy, $\log x \le x - 1$ implies

$$\mathrm{kl}(p, q) \le \frac{(p-q)^2}{q(1-q)}.$$

Hence,

$$\mathrm{kl}(\epsilon, 2u) \le \frac{(2u-\epsilon)^2}{2u(1-2u)} \le \frac{2u}{1-2u} \le 4u.$$

For all sufficiently large T,

$$\mathbb{E}^\pi_{\nu_{0,\delta}}[N_{mn}(T)] \ge \frac{\log T}{8u}.$$

There are finitely many bad edges, so one threshold for T can be chosen to make this inequality hold simultaneously for all of them.

Each pull of a bad edge contributes at least G/M to the instantaneous regret. Therefore,

$$\mathbb{E}^\pi_{\nu_{0,\delta}}[R(T)] \ge \sum_{m=1}^{M} \sum_{n=2}^{B+1} \frac{G}{M}\, \mathbb{E}^\pi_{\nu_{0,\delta}}[N_{mn}(T)] \ge \frac{BG}{8u} \log T.$$

Since $B \ge N/3$ for $N \ge 4$,

$$\mathbb{E}^\pi_{\nu_{0,\delta}}[R(T)] \ge \frac{1}{24} \frac{NG}{u} \log T$$

for every fixed δ and all sufficiently large T.

Fix $\delta = u/4$ and denote the resulting instance by

$$\nu^{\mathrm{scr}} \triangleq \nu_{0,u/4}.$$

This instance is independent of T. Therefore, for every uniformly good policy π, there exists $T_0 = T_0(\pi, \alpha, N, u, \epsilon) < \infty$ such that, for all $T \ge T_0$,

$$\mathbb{E}^\pi_{\nu^{\mathrm{scr}}}[R(T)] \ge \frac{1}{24} \frac{NG}{u} \log T. \tag{4}$$

Step 2: Finite-time learning lower bound. We now give an argument that holds for every policy, every finite horizon, and every $\delta \in (0, u/4]$.

Fix an agent m and a vector $\boldsymbol{\theta}_{-m} \in \mathcal{H}^{M-1}$. Construct a reference instance $\bar{\nu}_{m,\boldsymbol{\theta}_{-m},\delta}$ as follows. For every $i \neq m$, retain the means specified by θ_i. For agent m, give every edge to $\mathcal{H}$ mean $u - \delta$. All edges to the unit-capacity arms retain mean ϵ.

Let P_0 denote the trajectory law of policy π under this reference instance. For every $j \in \mathcal{H}$, let P_j denote its trajectory law under $\nu_{(\boldsymbol{\theta}_{-m},j),\delta}$. Write $\mathbb{E}_0$ and $\mathbb{E}_j$ for expectations under these laws, and define

$$x_j \triangleq \mathbb{E}_0[N_{mj}(T)], \qquad d_\delta \triangleq \mathrm{kl}(u - \delta, u).$$

The two instances differ only on edge (m, j). The KL chain rule under semi-bandit feedback gives

$$\mathrm{KL}(P_0, P_j) = x_j d_\delta.$$

Because $0 \leq N_{mj}(T) \leq T$, Pinsker's inequality implies

$$\mathbb{E}_j[N_{mj}(T)] \leq x_j + T\sqrt{\frac{x_j d_\delta}{2}}.$$

Since agent m selects one arm per round, $\sum_{j\in\mathcal{H}} x_j \leq T$. Thus, Cauchy-Schwarz gives

$$\frac{1}{K}\sum_{j\in\mathcal{H}} \mathbb{E}_j[N_{mj}(T)] \leq \frac{T}{K} + T\sqrt{\frac{T d_\delta}{2K}}.$$

Under $\nu_{\boldsymbol{\theta},\delta}$, every selection by agent m other than θ_m incurs regret at least

$$\frac{1}{M}\big(U_\alpha(u) - U_\alpha(u - \delta)\big).$$

Averaging over $\boldsymbol{\theta}_{-m}$ and then summing over all agents yields

$$\frac{1}{K^M}\sum_{\boldsymbol{\theta}\in\mathcal{H}^M} \mathbb{E}^\pi_{\nu_{\boldsymbol{\theta},\delta}}[R(T)] \geq T\big(U_\alpha(u) - U_\alpha(u-\delta)\big)\left(1 - \frac{1}{K} - \sqrt{\frac{T\mathrm{kl}(u-\delta,u)}{2K}}\right). \tag{5}$$

This inequality is finite-time and does not invoke uniform goodness.

For $T \geq N/u$, choose

$$\delta_T \triangleq \frac{1}{4}\sqrt{\frac{Nu}{T}} \leq \frac{u}{4}.$$

Then

$$\mathrm{kl}(u - \delta_T, u) \leq \frac{\delta_T^2}{u(1-u)} = \frac{N}{16T(1-u)}.$$

Using $K \geq N/2$, $K \geq 2$, and $u \leq 1/8$, we obtain

$$1 - \frac{1}{K} - \sqrt{\frac{T\mathrm{kl}(u-\delta_T,u)}{2K}} \geq \frac{1}{2} - \frac{1}{\sqrt{14}} \geq \frac{1}{6}.$$

Combining this inequality with $U_\alpha(u) - U_\alpha(u - \delta_T) \geq \delta_T u^{-\alpha}$ in (5) gives

$$\frac{1}{K^M}\sum_{\boldsymbol{\theta}\in\mathcal{H}^M} \mathbb{E}^\pi_{\nu_{\boldsymbol{\theta},\delta_T}}[R(T)] \geq \frac{1}{24}u^{\frac{1}{2}-\alpha}\sqrt{NT}.$$

Therefore, there exists $\boldsymbol{\theta}_T = \boldsymbol{\theta}_T(\pi, T) \in \mathcal{H}^M$ such that

$$\mathbb{E}^\pi_{\nu_{\boldsymbol{\theta}_T,\delta_T}}[R(T)] \geq \frac{1}{24}u^{\frac{1}{2}-\alpha}\sqrt{NT}. \tag{6}$$

Step 3: Combining the two lower bounds. Both ν^{scr} and $\nu_{\boldsymbol{\theta}_T,\delta_T}$ belong to $\mathcal{V}(u, \epsilon)$, and all instances in this family have the same capacities and identical values of $\Delta(\mathcal{M}\times\mathcal{N})$, $\mu^*_{\min}$, and $G_{\max}$.

Fix a uniformly good policy π and $T \geq \max\{T_0, N/u\}$. Since both hard instances belong to $\mathcal{V}(u,\epsilon)$,

$$\begin{aligned}\sup_{\nu\in\mathcal{V}(u,\epsilon)} \mathbb{E}_\nu^\pi[R(T)] &\geq \max\left\{\mathbb{E}_{\nu^{\mathrm{scr}}}^\pi[R(T)], \mathbb{E}_{\nu_{\boldsymbol{\theta}_T,\delta_T}}^\pi[R(T)]\right\} \\ &\geq \max\left\{\frac{1}{24}\frac{NG}{u}\log T, \frac{1}{24}u^{\frac{1}{2}-\alpha}\sqrt{NT}\right\} \\ &\geq \frac{1}{48}\left[\frac{NG}{u}\log T + u^{\frac{1}{2}-\alpha}\sqrt{NT}\right].\end{aligned}$$

Finally, using $N = \Delta(\mathcal{M}\times\mathcal{N})$, $u = \mu_{\min}^*$, and $G = G_{\max}(\nu)$ for every $\nu \in \mathcal{V}(u,\epsilon)$, we obtain

$$\sup_{\nu\in\mathcal{V}(u,\epsilon)} \mathbb{E}_\nu^\pi[R(T)] = \widetilde{\Omega}\left(\frac{\Delta(\mathcal{M}\times\mathcal{N})G_{\max}}{\mu_{\min}^*} + (\mu_{\min}^*)^{\frac{1}{2}-\alpha}\sqrt{NT}\right),$$

which proves the theorem. □

Lemma 10 (Change-of-measure lower bound for a mandatory base-arm)**.** *Consider two fixed Bernoulli instances ν and ν' with independent base-arm rewards and unique optimal matchings, which differ only on one base-arm e. Let the mean of e be μ_e under ν and μ_e' under ν', with $\mu_e \neq \mu_e'$. Suppose that no optimal matching under ν contains e, while every optimal matching under ν' contains e. Then every uniformly good policy π satisfies*

$$\liminf_{T\to\infty} \frac{\mathbb{E}_\nu^\pi[N_e(T)]}{\log T} \geq \frac{1}{\mathrm{kl}(\mu_e,\mu_e')}.$$

Consequently,

$$\mathbb{E}_\nu^\pi[N_e(T)] = \widetilde{\Omega}\left(\frac{1}{\mathrm{kl}(\mu_e,\mu_e')}\right).$$

Proof. Let $\mathbb{P}_\nu^\pi$ and $\mathbb{P}_{\nu'}^\pi$ denote the trajectory laws induced by policy π under ν and ν', respectively. Since the two instances differ only on base-arm e, the chain rule for KL divergence under semi-bandit feedback gives

$$\mathrm{KL}(\mathbb{P}_\nu^\pi, \mathbb{P}_{\nu'}^\pi) = \mathbb{E}_\nu^\pi[N_e(T)]\,\mathrm{kl}(\mu_e,\mu_e').$$

Since no optimal matching under ν contains e, and the feasible matching set is finite, there exists

$$\Delta_e \triangleq \max_{p\in\mathcal{P}} F_\alpha(p,\mu^\nu) - \max_{p\in\mathcal{P}:\, e\in X(p)} F_\alpha(p,\mu^\nu) > 0.$$

Hence, every round selecting e incurs at least Δ_e regret under ν. By uniform goodness,

$$\Delta_e\,\mathbb{E}_\nu^\pi[N_e(T)] \leq \mathbb{E}_\nu^\pi[R(T)] = o(T^a), \qquad \forall a > 0,$$

and therefore,

$$\mathbb{E}_\nu^\pi[N_e(T)] = o(T^a), \qquad \forall a > 0.$$

Define the event

$$A_T \triangleq \{N_e(T) \leq T/2\}.$$

By Markov's inequality,

$$\mathbb{P}_\nu^\pi(A_T^c) = \mathbb{P}_\nu^\pi(N_e(T) > T/2) \leq \frac{2\mathbb{E}_\nu^\pi[N_e(T)]}{T} = o(T^{a-1}), \qquad \forall a > 0.$$

In particular, $\mathbb{P}_\nu^\pi(A_T) \to 1$.

On the other hand, since every optimal matching under ν' contains e, there exists

$$\Delta_e' \triangleq \max_{p\in\mathcal{P}} F_\alpha(p,\mu^{\nu'}) - \max_{p\in\mathcal{P}:\, e\notin X(p)} F_\alpha(p,\mu^{\nu'}) > 0.$$

Thus, every round not selecting e incurs at least Δ_e' regret under ν'. By uniform goodness,

$$\Delta_e'\,\mathbb{E}_{\nu'}^\pi[T - N_e(T)] \leq \mathbb{E}_{\nu'}^\pi[R(T)] = o(T^a), \qquad \forall a > 0,$$

and hence,

$$\mathbb{E}_{\nu'}^\pi[T - N_e(T)] = o(T^a), \qquad \forall a > 0.$$

If A_T occurs, then $T - N_e(T) \geq T/2$. Therefore,

$$\mathbb{P}_{\nu'}^\pi(A_T) \leq \frac{2\mathbb{E}_{\nu'}^\pi[T - N_e(T)]}{T} = o(T^{a-1}), \qquad \forall a > 0.$$

Equivalently, for every $b < 1$,

$$\mathbb{P}^{\pi}_{\nu'}(A_T) = o(T^{-b}).$$

By the binary data-processing inequality for KL divergence,

$$\mathrm{KL}(\mathbb{P}^{\pi}_{\nu}, \mathbb{P}^{\pi}_{\nu'}) \geq d(\mathbb{P}^{\pi}_{\nu}(A_T), \mathbb{P}^{\pi}_{\nu'}(A_T))\,,$$

where $d(\cdot, \cdot)$ denotes the binary relative entropy.

Fix any $\varepsilon \in (0, 1)$. Since

$$\mathbb{P}^{\pi}_{\nu}(A_T) = 1 - o(1), \qquad \mathbb{P}^{\pi}_{\nu'}(A_T) = o(T^{-1+\varepsilon}),$$

we have

$$d(\mathbb{P}^{\pi}_{\nu}(A_T), \mathbb{P}^{\pi}_{\nu'}(A_T)) \geq (1 - \varepsilon - o(1)) \log T.$$

Combining this with the trajectory-KL identity yields

$$\mathbb{E}^{\pi}_{\nu}[N_e(T)]\, \mathrm{kl}(\mu_e, \mu'_e) \geq (1 - \varepsilon - o(1)) \log T.$$

Since $\varepsilon > 0$ is arbitrary,

$$\liminf_{T\to\infty} \frac{\mathbb{E}^{\pi}_{\nu}[N_e(T)]}{\log T} \geq \frac{1}{\mathrm{kl}(\mu_e, \mu'_e)}.$$

This completes the proof. □

B Cover Exploration Algorithm

The cover exploration algorithm is shown in Algorithm 3.

Algorithm 3 BuildCover($\mathcal{A}$)

Require: Base-arm set $\mathcal{A} = \mathcal{M} \times \mathcal{N}$ and capacities $\{C_n\}$.
1: Construct arm-copy nodes $\mathcal{N}' = \{(n, c) : c = 1, \ldots, C_n\}$;
2: For each arm n, partition $\mathcal{A}_n = \{(m, n) \in \mathcal{A}\}$ into $\{\mathcal{A}_{n,c}\}_{c=1}^{C_n}$ such that $|\mathcal{A}_{n,c}| \leq \lceil |\mathcal{A}_n|/C_n \rceil$;
3: Build expanded edge set $\mathcal{A}' \leftarrow \{(m, (n, c)) : (m, n) \in \mathcal{A}_{n,c}\}$;
4: Decompose $\mathcal{A}'$ into $\Delta(\mathcal{M} \times \mathcal{N})$ edge-disjoint matchings $\{M_h\}_{h=1}^{\Delta(\mathcal{M}\times\mathcal{N})}$ via Kőnig's theorem;
5: Map each M_h back to a partial capacity-feasible matching $\widetilde{p}^{(h)}$ by collapsing copies $(n, c) \mapsto n$;
6: **return** The cover schedule $(\widetilde{p}^{(1)}, \ldots, \widetilde{p}^{(\Delta(\mathcal{M}\times\mathcal{N}))})$.

C Completion Algorithm

The completion algorithm is shown in Algorithm 4.

Algorithm 4 Complete($\widetilde{p}, \{C_n\}$)

Require: $\widetilde{p} \in \{0, 1\}^{M\times N}$ and capacities $\{C_n\}$.
Ensure: A full feasible matching $p \in \{0, 1\}^{\mathcal{M}\times\mathcal{N}}$.
1: $p \leftarrow \widetilde{p}$.
2: $\mathrm{cap}[n] \leftarrow C_n - \sum_{m=1}^{M} p_{mn}$ for all n.
3: $U \leftarrow \{m : \sum_{n=1}^{N} p_{mn} = 0\}$
4: **for** each $m \in U$ **do**
5: pick any n with $\mathrm{cap}[n] > 0$;
6: set $p_{mn} \leftarrow 1$ and $\mathrm{cap}[n] \leftarrow \mathrm{cap}[n] - 1$;
7: **end for**
8: **return** p.

D Omitted Details in Section 4

In this appendix, we provide the full decentralized details omitted from Section 4. Since the main technical contribution of the decentralized algorithm is the insertion-gap shrinking mechanism, we first present the decentralized insertion-gap computation and the Phase-II proofs in Appendix D.1. We then present the auxiliary Phase-I routines, including decentralized exploration, decentralized completion, and rare-auction exploration, in Appendix D.2. These Phase-I routines are used only to reproduce the safe retained set required by the centralized analysis and are not the main part of the decentralized design.

D.1 Phase II Details: Decentralized Insertion-gap Computation and Proof

D.1.1 Decentralized Insertion-gap Computation

We first establish the exchange structure for capacitated matchings, then give an implicit shortest-path recursion and prove its forced-insertion accuracy.

For an active graph $\mathcal{A}$, define

$$\mathcal{P}_e(\mathcal{A}) \triangleq \{p \in \mathcal{P}(\mathcal{A}) : p_e = 1\}.$$

At epoch q, let

$$\widehat{p}_q^\star \in \arg\max_{p \in \mathcal{P}(\mathcal{A}_q)} \widehat{W}_q(p).$$

Whenever $\mathcal{P}_e(\mathcal{A}_q) \neq \varnothing$, define

$$p_{q,e}^{\star,\mathrm{ins,est}} \in \arg\max_{p \in \mathcal{P}_e(\mathcal{A}_q)} \widehat{W}_q(p),$$

$$\Delta_{e,q}^{\star,\mathrm{ins,est}} \triangleq \widehat{W}_q(\widehat{p}_q^\star) - \widehat{W}_q(p_{q,e}^{\star,\mathrm{ins,est}}).$$

If $\mathcal{P}_e(\mathcal{A}_q) = \varnothing$, define this insertion gap to be $+\infty$.

We suppress the epoch index below. Let $(\widehat{p}, a, \lambda)$ be the auction output on $\mathcal{A}$, and define

$$\mathcal{X}(\widehat{p}) \triangleq \{(m,n) \in \mathcal{A} : \widehat{p}_{mn} = 1\}.$$

Write $n_i \triangleq n_i(\widehat{p})$ for the arm currently assigned to agent i, and let

$$H_n \triangleq \{i : n_i = n\}, \qquad S \triangleq \{n : |H_n| < C_n\}, \qquad \mathcal{N}(\mathcal{A}) \triangleq \{n : \exists i,\ (i,n) \in \mathcal{A}\}.$$

Arms without incident active edges are omitted from the computation. The auction prices satisfy

$$\lambda_n \geq 0, \qquad \lambda_n = 0 \quad \text{for } n \in S,$$

by Lemma 14. The priced-value losses are

$$\widehat{r}_{ik} \triangleq a_i + \lambda_k - \widehat{v}_{ik} \geq 0, \qquad (i,k) \in \mathcal{A}.$$

Slot duplication and alternating cycles. Replace arm n by C_n distinct slots. Every real agent incident to n is connected to all these slots, with the same value $\widehat{v}_{in}$. Add $\sum_n C_n - M$ dummy agents, each connected to every slot with value zero. A full capacitated matching extends to a perfect matching in this expanded graph by assigning the dummy agents to unused slots. Conversely, removing dummy assignments and merging the slots of each arm recovers a full capacitated matching.

Fix a non-selected edge $e = (m,n)$ such that $\mathcal{P}_e(\mathcal{A}) \neq \varnothing$, and let $p_e^\star$ maximize $\widehat{W}$ over $\mathcal{P}_e(\mathcal{A})$. Extend $\widehat{p}$ and $p_e^\star$ to perfect matchings on the same expanded graph. After their common edges are removed, every remaining vertex has one incident edge from each matching. Their symmetric difference therefore decomposes into vertex-disjoint alternating cycles.

Let C_e be the cycle containing the slot edge that assigns m to arm n in $p_e^\star$. Replace the current-matching edges of C_e by its alternative edges. Every vertex on the cycle loses one assignment and receives one assignment. The resulting perfect matching contains the forced edge, and its projection is a feasible member of $\mathcal{P}_e(\mathcal{A})$. Applying all cycles recovers $p_e^\star$.

Implicit representation of the augmented exchange. The slot-expanded graph need not be constructed explicitly. Instead, use the active arms as states and introduce a shared slack state 0. For

every active edge (i, k) with $k \neq n_i$, add the agent-labeled transition

$$n_i \xrightarrow{i} k \quad \text{with length } \widehat{r}_{ik}.$$

For every slack arm $s \in S$, add a transition $s \to 0$ of length zero, and for every $k \in \mathcal{N}(\mathcal{A})$, add a transition $0 \to k$ of length λ_k. Hence, the two-step transition $s \to 0 \to k$ represents moving a dummy agent from an unused slot of s to a slot of k.

An alternating cycle in the slot-expanded graph projects to a closed walk in the implicit state graph. Since all transition lengths are nonnegative, a minimum-length continuation can be chosen simple. A simple cycle containing the forced transition $n_m \xrightarrow{m} n$ visits each arm state and the slack state at most once, and thus moves each real agent at most once. If the cycle contains $s \to 0 \to k$, a dummy leaves an unused slot of the slack arm s and occupies a slot of k. After projection, this allows one additional real assignment at s and releases one real assignment at k. All other visited arms have one real assignment entering and one leaving. Therefore, the projected exchange preserves all capacities and yields a feasible matching containing the forced edge.

For a slot-expanded alternating cycle C, let $L_{\widehat{p}}(C)$ denote the sum of the priced-value losses of its new real-agent edges and the costs λ_k of its new dummy-agent edges. We write $\widehat{p} \oplus C$ for the capacitated matching obtained by applying C and projecting the resulting slot matching back to the original arms.

Shared root-distance recursion. Fix a root agent m. After the forced transition $n_m \to n$, we seek a continuation from n back to n_m. Agent m cannot be moved again during this continuation.

Let $D^{(d)}_{n|m}$ be the minimum continuation length using at most d transitions, where a real-agent move counts as one transition and a two-arc dummy move $s \to 0 \to k$ also counts as one. Initialize

$$D^{(0)}_{n|m} = \begin{cases} 0, & n = n_m, \\ +\infty, & n \neq n_m. \end{cases} \tag{7}$$

For $i \neq m$, define

$$B^{(d+1)}_{i|m} \triangleq \min_{k \in \mathcal{N}_i(\mathcal{A}) \setminus \{n_i\}} \left\{ \widehat{r}_{ik} + D^{(d)}_{k|m} \right\}, \tag{8}$$

and set $B^{(d+1)}_{m|m} = +\infty$. The shared slack message is

$$Z^{(d+1)}_m \triangleq \min_{k \in \mathcal{N}(\mathcal{A})} \left\{ \lambda_k + D^{(d)}_{k|m} \right\}. \tag{9}$$

The arm update is

$$D^{(d+1)}_{n|m} = \begin{cases} 0, & n = n_m, \\ \min\left\{ D^{(d)}_{n|m}, \min\limits_{i \in H_n \setminus \{m\}} B^{(d+1)}_{i|m}, Z^{(d+1)}_m \right\}, & n \in S \setminus \{n_m\}, \\ \min\left\{ D^{(d)}_{n|m}, \min\limits_{i \in H_n \setminus \{m\}} B^{(d+1)}_{i|m} \right\}, & \text{otherwise.} \end{cases} \tag{10}$$

We use the convention $\min \varnothing = +\infty$.

The message $Z^{(d+1)}_m$ is obtained by a minimum reduction of the arm messages $\lambda_k + D^{(d)}_{k|m}$ and is shared by all slack arms. This requires no additional auction.

For path reconstruction, each state stores its minimizing decision and the corresponding previous iteration. Among equal-length candidates, choose the one with fewer transitions, followed by a fixed deterministic tie-breaking rule. The transition count is updated by adding one for a real or dummy move and adding zero when retaining $D^{(d)}_{n|m}$. This secondary criterion excludes unnecessary zero-length cycles.

Proposition 4 (Exact augmented root-distance recursion)**.** *Fix a root agent m. For every $d \geq 0$, $D^{(d)}_{n|m}$ is the minimum length of a continuation from n to n_m using at most d real or dummy transitions and no further move of m. The recursion is non-increasing in d and stabilizes after at most M iterations.*

Algorithm 5 Decentralized Insertion-Gap Algorithm

Require: Auction matching $\widehat{p} \in \mathcal{P}(\mathcal{A})$, auction outputs (a, λ), finite values $\widehat{v}$, and capacities $\{C_n\}$.
Ensure: Observable gaps $\widehat{\Delta}_e^{\mathrm{ins}}$ and feasible forced matchings when they exist.

Each active arm n forms H_n, broadcasts λ_n, and reports whether $n \in S$;
Each agent i computes $\widehat{r}_{ik}$ on its active incident edges;
For every root m, initialize $D_{n|m}^{(0)}$ by (7);
for $d = 0, \ldots, M-1$ **do**
  Each arm broadcasts its current root-distance vector to its incident active agents;
  for each agent i in parallel **do**
    Compute $B_{i|m}^{(d+1)}$ for all $m \neq i$ by (8), and set $B_{i|i}^{(d+1)} \leftarrow +\infty$;
    Send the minimizing values and decisions to arm n_i;
  end for
  Compute the shared messages $Z_m^{(d+1)}$ for all roots by (9);
  for each active arm n in parallel **do**
    Update $D_{n|m}^{(d+1)}$ for every root using (10);
    Store the selected decision $\chi_{n|m}^{(d+1)}$, using length and then transition count for tie-breaking;
  end for
end for
Set $D_{n|m}^{\star} \leftarrow D_{n|m}^{(M)}$;
for each non-selected edge $e = (m, n)$ **do**
  if $D_{n|m}^{\star} = +\infty$ **then**
    Set $\widehat{\Delta}_e^{\mathrm{ins}} \leftarrow +\infty$ and $\widehat{p}_e^{\mathrm{ins}} \leftarrow \perp$;
  else
    Initialize the new real assignments $\mathcal{E}_e^{\mathrm{new}} \leftarrow \{(m, n)\}$, $k \leftarrow n$, and $h \leftarrow M$;
    while $k \neq n_m$ **do**
      Read the stored decision $\chi_{k|m}^{(h)}$;
      if the decision retains the previous value **then**
        Keep k unchanged;
      else if the decision moves a real agent i to arm ℓ **then**
        Add (i, ℓ) to $\mathcal{E}_e^{\mathrm{new}}$ and set $k \leftarrow \ell$;
      else
        Follow the stored dummy transition $k \to 0 \to \ell$ and set $k \leftarrow \ell$;
      end if
      Set $h \leftarrow h - 1$;
    end while
    Construct $\widehat{p}_e^{\mathrm{ins}}$ from $\widehat{p}$ by replacing each real assignment listed in $\mathcal{E}_e^{\mathrm{new}}$;
    Set $\widehat{\Delta}_e^{\mathrm{ins}} \leftarrow \widehat{W}(\widehat{p}) - \widehat{W}(\widehat{p}_e^{\mathrm{ins}})$;
  end if
end for

For every non-selected edge $e = (m, n)$ with $\mathcal{P}_e(\mathcal{A}) \neq \varnothing$,

$$\Lambda_e^{\star} \triangleq \widehat{r}_{mn} + D_{n|m}^{\star}$$

equals the minimum length of a slot-expanded alternating cycle containing a slot edge representing e. The reconstructed cycle projects to a feasible matching containing e. Furthermore,

$$D_{n|m}^{\star} = +\infty \quad \Longleftrightarrow \quad \mathcal{P}_e(\mathcal{A}) = \varnothing.$$

Proof. **Step 1: Correctness of the recursion.** For $d = 0$, only the empty continuation at n_m is possible, giving (7). Assume the claim holds at iteration d. A continuation using at most $d+1$ transitions either uses at most d transitions, starts by moving an agent currently assigned to n, or, when n is slack, starts by moving a dummy agent. These three possibilities give the three candidates in (10). The real-agent candidate is exactly (8), and the dummy candidate is exactly (9). The root state remains zero because all lengths are nonnegative. Induction proves the distance interpretation. The retained previous value also gives

$$D_{n|m}^{(d+1)} \leq D_{n|m}^{(d)}.$$

Step 2: Finite convergence. View dummy transitions as paths through the shared state 0. All arc lengths in this graph are nonnegative, so any repeated vertex in a finite walk can be removed together with the intervening closed subwalk without increasing its length. Among minimum-length walks, the transition-count tie-breaking rule therefore selects a simple path.

Such a path visits the slack state at most once. It also uses distinct real agents, because all transitions labeled by a given agent leave that agent's single current arm. Since agent m is excluded, the continuation contains at most $M-1$ real-agent transitions and at most one dummy transition. Thus, at most M transitions suffice, and

$$D_{n|m}^{(d)} = D_{n|m}^{(M)} \qquad \text{for all } d \geq M.$$

Step 3: Correspondence with slot-expanded cycles. Take any slot-expanded alternating cycle containing the forced edge. Each new real-agent edge induces its labeled arm transition. Each new dummy edge induces a transition $s \to 0 \to k$, where the dummy's old slot belongs to a slack arm s. The resulting arm-state walk has the same total length, except that self-loops caused by movements between slots of the same arm may be deleted.

Remove the forced transition $n_m \to n$. The remainder is a walk from n to n_m that does not move m. Removing closed subwalks produces a simple continuation of no greater length. Therefore, the minimum implicit exchange length is no greater than the length of any slot-expanded cycle containing e.

Conversely, prepend $n_m \xrightarrow{m} n$ to a simple continuation returned by the recursion. For each real-agent transition, use that agent's current slot. If a dummy transition $s \to 0 \to k$ occurs, use an unused slot of s and a dummy currently assigned to it. These choices lift the implicit cycle to an alternating cycle of the expanded perfect matching, with the same length. Applying this cycle preserves the perfect matching. Its projection is capacity-feasible and contains e. This proves equality of the two minimum lengths.

Finally, if $\mathcal{P}_e(\mathcal{A}) \neq \varnothing$, the symmetric-difference decomposition with any feasible forced matching provides a cycle containing e, and hence a finite continuation. The converse follows from the feasible matching reconstructed from a finite continuation. □

Proposition 5 (Accuracy of the cycle-induced forced-insertion matching)**.** *Consider $(\widehat{p}, a, \lambda)$ returned by Algorithm 8 with bid increment ϵ_{auc}. For every non-selected edge e with $\mathcal{P}_e(\mathcal{A}) \neq \varnothing$, Algorithm 5 returns $\widehat{p}_e^{\text{ins}} \in \mathcal{P}_e(\mathcal{A})$ satisfying*

$$\widetilde{\Delta}_e^{\text{ins}} \leq \widehat{\Delta}_e^{\text{ins}} \leq \widetilde{\Delta}_e^{\text{ins}} + M\epsilon_{\text{auc}},$$

where

$$\widetilde{\Delta}_e^{\text{ins}} \triangleq \widehat{W}(\widehat{p}) - \max_{p \in \mathcal{P}_e(\mathcal{A})} \widehat{W}(p).$$

If $\mathcal{P}_e(\mathcal{A}) = \varnothing$, the algorithm returns $\widehat{\Delta}_e^{\text{ins}} = +\infty$.

Proof. Fix an edge with a feasible forced-insertion problem. For each real agent i, define its complementary-slackness error

$$\zeta_i \triangleq a_i + \lambda_{n_i} - \widehat{v}_{i,n_i}.$$

By Lemma 14,

$$0 \leq \zeta_i \leq \epsilon_{\text{auc}}.$$

All slots of arm k inherit price λ_k. Give every dummy agent dual value zero. Its reduced cost on a slot of arm k is therefore $\lambda_k \geq 0$. In the current perfect matching, dummy agents occupy unused slots of slack arms, whose prices are zero. Consequently, every dummy agent has zero complementary-slackness error.

For a slot-expanded alternating cycle C, let $I(C)$ be the set of real agents on the cycle. Every agent and every slot appears once among the old edges and once among the new edges. The agent dual values and slot prices therefore cancel, giving

$$\widehat{W}(\widehat{p}) - \widehat{W}(\widehat{p} \oplus C) = L_{\widehat{p}}(C) - \sum_{i \in I(C)} \zeta_i. \tag{11}$$

In particular,

$$L_{\widehat{p}}(C) - M\epsilon_{\mathrm{auc}} \le \widehat{W}(\widehat{p}) - \widehat{W}(\widehat{p} \oplus C) \le L_{\widehat{p}}(C). \tag{12}$$

Let $p_e^\star$ maximize $\widehat{W}$ over $\mathcal{P}_e(\mathcal{A})$. Decompose the symmetric difference between perfect extensions of $\widehat{p}$ and $p_e^\star$ into cycles

$$\mathcal{C} = \{C_e, C_1, \dots, C_J\},$$

where C_e contains the forced edge. Applying all these cycles transforms $\widehat{p}$ into $p_e^\star$. Since their real-agent sets are disjoint, (11) gives

$$\sum_{C\in\mathcal{C}} L_{\widehat{p}}(C) = \widetilde{\Delta}_e^{\mathrm{ins}} + \sum_{i\in\bigcup_{C\in\mathcal{C}} I(C)} \zeta_i \le \widetilde{\Delta}_e^{\mathrm{ins}} + M\epsilon_{\mathrm{auc}}.$$

Every cycle length is nonnegative. Therefore,

$$L_{\widehat{p}}(C_e) \le \sum_{C\in\mathcal{C}} L_{\widehat{p}}(C) \le \widetilde{\Delta}_e^{\mathrm{ins}} + M\epsilon_{\mathrm{auc}}.$$

Let C_e^{alg} be the cycle reconstructed by Algorithm 5. By Proposition 4,

$$L_{\widehat{p}}(C_e^{\mathrm{alg}}) \le L_{\widehat{p}}(C_e).$$

Applying (11) to C_e^{alg} yields

$$\widehat{\Delta}_e^{\mathrm{ins}} = L_{\widehat{p}}(C_e^{\mathrm{alg}}) - \sum_{i\in I(C_e^{\mathrm{alg}})} \zeta_i \le L_{\widehat{p}}(C_e^{\mathrm{alg}}).$$

Since $\widehat{p}_e^{\mathrm{ins}}$ is a feasible forced matching, its loss cannot be smaller than $\widetilde{\Delta}_e^{\mathrm{ins}}$. Combining the inequalities gives

$$\widetilde{\Delta}_e^{\mathrm{ins}} \le \widehat{\Delta}_e^{\mathrm{ins}} \le L_{\widehat{p}}(C_e^{\mathrm{alg}}) \le L_{\widehat{p}}(C_e) \le \widetilde{\Delta}_e^{\mathrm{ins}} + M\epsilon_{\mathrm{auc}}.$$

The infeasible case follows from Proposition 4. □

Corollary 1 (Epoch-q forced-insertion accuracy)**.** *At a shrinking epoch q, let $(\widehat{p}_q, a^{(q)}, \lambda^{(q)})$ be the auction output with $\epsilon_{\mathrm{auc},q} \le \varepsilon_q/M$. For every non-selected edge e such that $\mathcal{P}_e(\mathcal{A}_q) \neq \varnothing$, the returned matching satisfies*

$$\widehat{W}_q(p_{q,e}^{\star,\mathrm{ins,est}}) - \widehat{W}_q(\widehat{p}_{q,e}^{\mathrm{ins}}) \le \varepsilon_q.$$

Proof. By the definitions,

$$\widehat{W}_q(p_{q,e}^{\star,\mathrm{ins,est}}) - \widehat{W}_q(\widehat{p}_{q,e}^{\mathrm{ins}}) = \widehat{\Delta}_{e,q}^{\mathrm{ins}} - \widetilde{\Delta}_{e,q}^{\mathrm{ins}}.$$

Apply Proposition 5 and $M\epsilon_{\mathrm{auc},q} \le \varepsilon_q$. □

Proposition 6 (Complexity comparison: shared propagation vs. per-edge auction recomputation)**.** *Fix a shrinking epoch q, and let*

$$A_q \triangleq |\mathcal{A}_q|, \qquad U_q \triangleq |\mathcal{A}_q \setminus \mathcal{X}(\widehat{p}_q)| = A_q - M.$$

Suppose one auction call has complexity C_q^{auc}. If one recomputes a constrained auction for every non-selected edge, the epoch complexity is

$$C_q^{\mathrm{per}} = O\Big((U_q+1)C_q^{\mathrm{auc}}\Big).$$

By contrast, our algorithm, i.e., one main auction plus Algorithm 5, gives

$$C_q^{\mathrm{ins}} = O\Big(C_q^{\mathrm{auc}} + M^2 A_q\Big).$$

Consequently, the proposed decentralized insertion-gap algorithm replaces (U_q+1) repeated auction calls by one main auction together with a shared propagation step whose cost is linear in the active graph size up to the factor M^2.

Proof. The per-edge recomputation method uses one main auction and one forced-insertion auction per non-selected edge. Our method uses one main auction, and then a shared insertion-gap recursion. By Proposition 4, the recursion stabilizes after at most M rounds. Each round processes all active edges once on the agent side and all matched agents once on the arm side, for total cost $O(MA_q)$. Hence, the total propagation cost is $O(M^2A_q)$, which yields the stated bound. □

D.1.2 Gap Relation and Safe Elimination

Lemma 11 (Matching-level error bound)**.** *Conditioned on $\mathcal{E}$, for every shrinking epoch q and every feasible matching $p \in \mathcal{P}(\mathcal{A}_q)$,*

$$\left|W_q(p) - \widehat{W}_q(p)\right| \le \varepsilon_q,$$

where

$$W_q(p) \triangleq \sum_{(m,n)\in\mathcal{A}_q} w_m U_\alpha(\mu_{mn}) p_{mn}.$$

Proof. Fix an epoch q and an edge $(m,n) \in \mathcal{A}_q$. By the construction of the shrinking epochs, every active edge satisfies $N_{mn}(t_q) \ge n_q$. Moreover, by Lemma 19, on the retained set, $\mu_{mn} \ge u_{\mathrm{se}}$, and the good confidence event gives $\mu_{mn} \le \mathrm{UCB}_{mn}(t_q) \le 1$. By the mean-value theorem and the monotonicity of $U'_\alpha(x) = x^{-\alpha}$,

$$\begin{aligned}
|w_m U_\alpha(\mathrm{UCB}_{mn}(t_q)) - w_m U_\alpha(\mu_{mn})| &\le w_m \mu_{mn}^{-\alpha} \left|\mathrm{UCB}_{mn}(t_q) - \mu_{mn}\right| \\
&\le \left(\sqrt{3} + \sqrt{9+3\sqrt{3}}\right) w_m \mu_{mn}^{\frac{1}{2}-\alpha} \sqrt{\frac{\log(MNT^2)}{n_q}} \\
&\le \left(\sqrt{3} + \sqrt{9+3\sqrt{3}}\right) w_m u_{\mathrm{se}}^{\frac{1}{2}-\alpha} \sqrt{\frac{\log(MNT^2)}{n_q}},
\end{aligned}$$

where the last step uses $\mu_{mn} \ge u_{\mathrm{se}}$ and $\frac{1}{2} - \alpha < 0$.

Therefore, for any $p \in \mathcal{P}(\mathcal{A}_q)$,

$$|W_q(p) - \widehat{W}_q(p)| \le \left(\sqrt{3} + \sqrt{9+3\sqrt{3}}\right) u_{\mathrm{se}}^{\frac{1}{2}-\alpha} \sqrt{\frac{\log(MNT^2)}{n_q}} \sum_{(m,n)\in\mathcal{A}_q} w_m p_{mn} = \varepsilon_q.$$

□

Recall that $W_q(p) = \sum_{(m,n)\in\mathcal{A}_q} w_m U_\alpha(\mu_{mn}) p_{mn}$ is the objective with true means at epoch q. For every edge $e \in \mathcal{A}_q$, define the true insertion subproblem

$$p_{q,e}^{\star,\mathrm{ins,true}} \in \arg\max\left\{W_q(p) : p \in \mathcal{P}(\mathcal{A}_q),\ p_e = 1\right\},$$

and the true insertion gap

$$\Delta_{e,q} \triangleq W_q(p_q^{\star,\mathrm{true}}) - W_q\big(p_{q,e}^{\star,\mathrm{ins,true}}\big),$$

where $p_q^{\star,\mathrm{true}} \in \arg\max_{p\in\mathcal{P}(\mathcal{A}_q)} W_q(p)$. If $\mathcal{P}_e(\mathcal{A}_q) = \varnothing$, define $\Delta_{e,q} = +\infty$. Such an edge cannot belong to any feasible matching and is excluded from the refresh family by its infinite observable insertion gap.

The next proposition is crucial in Phase II: it shows that the observable insertion gaps returned by the decentralized shortest-path algorithm bracket the true insertion gaps up to $O(\varepsilon_q)$. This is exactly what makes the elimination rule safe.

Proposition 7 (Observable insertion gaps vs. true insertion gaps)**.** *Conditioned on $\mathcal{E}$, for every epoch q and every edge $e \in \mathcal{A}_q \setminus \mathcal{X}(\widehat{p}_q)$,*

$$\widehat{\Delta}_{e,q}^{\mathrm{ins}} - 3\varepsilon_q \le \Delta_{e,q} \le \widehat{\Delta}_{e,q}^{\mathrm{ins}} + 3\varepsilon_q.$$

Consequently,

$$\widehat{\Delta}_{e,q}^{\mathrm{ins}} > 3\varepsilon_q \quad\Longrightarrow\quad \Delta_{e,q} > 0,$$

and

$$\widehat{\Delta}_{e,q}^{\mathrm{ins}} \le 3\varepsilon_q \quad\Longrightarrow\quad \Delta_{e,q} \le 6\varepsilon_q.$$

Proof. Fix an epoch q and a non-selected edge $e \in \mathcal{A}_q \setminus \mathcal{X}(\widehat{p}_q)$. If $\mathcal{P}_e(\mathcal{A}_q) = \varnothing$, then both $\Delta_{e,q}$ and $\widehat{\Delta}_{e,q}^{\mathrm{ins}}$ equal $+\infty$, and the claims hold in the extended-real sense. In the remainder, assume $\mathcal{P}_e(\mathcal{A}_q) \ne \varnothing$.

We first prove the lower bound. By Corollary 1,

$$\widehat{W}_q\big(p_{q,e}^{\star,\mathrm{ins,est}}\big) - \widehat{W}_q\big(\widehat{p}_{q,e}^{\mathrm{ins}}\big) \le \varepsilon_q.$$

Moreover, since $\widehat{p}_q^\star$ maximizes $\widehat{W}_q(\cdot)$,

$$\widehat{W}_q(\widehat{p}_q^\star) \geq \widehat{W}_q(\widehat{p}_q).$$

Hence,

$$\Delta_{e,q}^{\star,\mathrm{ins,est}} = \widehat{W}_q(\widehat{p}_q^\star) - \widehat{W}_q\big(p_{q,e}^{\star,\mathrm{ins,est}}\big) \geq \widehat{W}_q(\widehat{p}_q) - \widehat{W}_q\big(\widehat{p}_{q,e}^{\mathrm{ins}}\big) - \varepsilon_q = \widehat{\Delta}_{e,q}^{\mathrm{ins}} - \varepsilon_q.$$

Applying Lemma 11 to both $\widehat{p}_q^\star$ and $p_{q,e}^{\star,\mathrm{ins,est}}$, we obtain

$$\Delta_{e,q} \geq \Delta_{e,q}^{\star,\mathrm{ins,est}} - 2\varepsilon_q \geq \widehat{\Delta}_{e,q}^{\mathrm{ins}} - 3\varepsilon_q.$$

We next prove the upper bound. By ε_q-accuracy of the main auction,

$$\widehat{W}_q(\widehat{p}_q^\star) - \widehat{W}_q(\widehat{p}_q) \leq \varepsilon_q.$$

Since $p_{q,e}^{\star,\mathrm{ins,est}}$ maximizes the estimated constrained problem,

$$\widehat{W}_q\big(p_{q,e}^{\star,\mathrm{ins,est}}\big) \geq \widehat{W}_q\big(\widehat{p}_{q,e}^{\mathrm{ins}}\big).$$

Therefore,

$$\Delta_{e,q}^{\star,\mathrm{ins,est}} = \widehat{W}_q(\widehat{p}_q^\star) - \widehat{W}_q\big(p_{q,e}^{\star,\mathrm{ins,est}}\big) \leq \widehat{W}_q(\widehat{p}_q) - \widehat{W}_q\big(\widehat{p}_{q,e}^{\mathrm{ins}}\big) + \varepsilon_q = \widehat{\Delta}_{e,q}^{\mathrm{ins}} + \varepsilon_q.$$

Applying Lemma 11 again yields

$$\Delta_{e,q} \leq \Delta_{e,q}^{\star,\mathrm{ins,est}} + 2\varepsilon_q \leq \widehat{\Delta}_{e,q}^{\mathrm{ins}} + 3\varepsilon_q.$$

The two consequences follow immediately. □

As a direct consequence, any edge whose observable insertion gap is larger than $3\varepsilon_q$ can be deleted without affecting the true optimal matching.

Corollary 2 (Safe elimination rule). *Conditioned on $\mathcal{E}$, if an edge $e \in \mathcal{A}_q \setminus \mathcal{X}(\widehat{p}_q)$ satisfies*

$$\widehat{\Delta}_{e,q}^{\mathrm{ins}} > 3\varepsilon_q,$$

then e does not belong to any true optimal matching on the active set $\mathcal{A}_q$. Hence, eliminating all such edges at the end of epoch q preserves the true optimal matching.

Proof. By Proposition 7, the condition $\widehat{\Delta}_{e,q}^{\mathrm{ins}} > 3\varepsilon_q$ implies $\Delta_{e,q} > 0$, so any feasible matching containing e is strictly suboptimal on $\mathcal{A}_q$. □

D.1.3 Certified Freezing and Truncation

We first establish the correctness of the freezing rule and then bound the regret accumulated by the shrinking-and-refresh procedure. All statements below are conditioned on the confidence event $\mathcal{E}$, unless otherwise stated.

Lemma 12 (Per-epoch regret of the shrinking-epoch matchings). *For every epoch q,*

$$W_q(p_q^{\star,\mathrm{true}}) - W_q(\widehat{p}_q) \leq 3\varepsilon_q.$$

Moreover, for every surviving non-selected edge $e \in \mathcal{J}_q$,

$$W_q(p_q^{\star,\mathrm{true}}) - W_q(\widehat{p}_{q,e}^{\mathrm{ins}}) \leq 9\varepsilon_q.$$

Proof. For the main matching, the auction accuracy and Lemma 11 give $W_q(p_q^{\star,\mathrm{true}}) - W_q(\widehat{p}_q) \leq 3\varepsilon_q$.

For $e \in \mathcal{J}_q$, Proposition 7 gives $\Delta_{e,q} \leq 6\varepsilon_q$. Moreover, Corollary 1 and Lemma 11 imply

$$W_q(p_{q,e}^{\star,\mathrm{ins,true}}) - W_q(\widehat{p}_{q,e}^{\mathrm{ins}}) \leq 3\varepsilon_q.$$

Adding the true insertion gap proves the second claim. □

Lemma 13 (Correctness of the freezing rule). *If $\widehat{\Gamma}_q > 3\varepsilon_q$, then $\widehat{p}_q$ is the unique true optimal matching on $\mathcal{A}_q$. Consequently, it is the unique original optimal matching, and replaying it incurs no further regret or auction calls.*

Proof. Suppose that another true optimal matching $p \in \mathcal{P}(\mathcal{A}_q)$ exists. Since $p \neq \widehat{p}_q$, there is an edge $e \in \mathcal{X}(p) \setminus \mathcal{X}(\widehat{p}_q)$. Because p is optimal and contains e, we have $\Delta_{e,q} = 0$. Proposition 7 then gives

$\widehat{\Delta}^{\mathrm{ins}}_{e,q} \le 3\varepsilon_q$, contradicting $\widehat{\Gamma}_q > 3\varepsilon_q$. The original optimal matching remains feasible on every active graph by Corollary 2, which proves the second claim. □

For any active graph $\mathcal{A}$ containing p^*, define $\Delta^{\mathrm{ins}}_e(\mathcal{A}) \triangleq F_\alpha(p^*,\mu) - \max_{p\in\mathcal{P}(\mathcal{A}),\, p_e=1} F_\alpha(p,\mu)$, with value $+\infty$ if no feasible matching contains e. We use $1/(+\infty)=0$.

Theorem 6 (Exact freezing and regret)**.** *Assume that the original problem admits a unique optimal matching p^*, and define*

$$\Gamma_{\mathrm{fr}} \triangleq \min_{e\in\mathcal{A}_0\setminus\mathcal{X}(p^*)} \Delta^{\mathrm{ins}}_e(\mathcal{A}_0),$$

with $\min\varnothing = +\infty$. Then $\Gamma_{\mathrm{fr}} > 0$, and the freezing condition is satisfied once $6\varepsilon_q < \Gamma_{\mathrm{fr}}$. Hence, the number of Phase-II auction epochs before freezing is $O(\log\log T)$.

Moreover,

$$R_{\mathrm{II}}(T) \le C_g u_{\mathrm{se}}^{1-2\alpha}\log(MNT^2) \sum_{e\in\mathcal{A}_0\setminus\mathcal{X}(p^*)} \frac{1}{\Delta^{\mathrm{ins}}_e(\mathcal{A}_0)},$$

where $C_g > 0$ depends only on g. Consequently,

$$\mathbb{E}[R(T)] = \widetilde{O}\left(\frac{\Delta(\mathcal{M}\times\mathcal{N})G_{\max}(\mu)}{\mu^*_{\min}} + (\mu^*_{\min})^{1-2\alpha} \sum_{e\in(\mathcal{M}\times\mathcal{N})\setminus\mathcal{X}(p^*)} \frac{1}{\Delta^{\mathrm{ins}}_e(\mathcal{M}\times\mathcal{N})}\right).$$

For every fixed instance, the total number of auction calls is $O(\log\log T)$ on $\mathcal{E}$ and in expectation.

Proof. Because p^* remains feasible on every active graph, uniqueness and finiteness of the feasible matching set imply $\Gamma_{\mathrm{fr}} > 0$, with $\Gamma_{\mathrm{fr}} = +\infty$ when p^* is the only feasible matching on $\mathcal{A}_0$.

Suppose $6\varepsilon_q < \Gamma_{\mathrm{fr}}$. If $\widehat{p}_q \neq p^*$, then $\widehat{p}_q$ contains some $e \notin \mathcal{X}(p^*)$, and therefore

$$W_q(p^*) - W_q(\widehat{p}_q) \ge \Delta^{\mathrm{ins}}_e(\mathcal{A}_0) \ge \Gamma_{\mathrm{fr}}.$$

This contradicts Lemma 12, which bounds the same difference by $3\varepsilon_q$. Hence, $\widehat{p}_q = p^*$.

For every remaining non-selected edge, $\Delta_{e,q} \ge \Delta^{\mathrm{ins}}_e(\mathcal{A}_0) \ge \Gamma_{\mathrm{fr}}$. Proposition 7 gives $\widehat{\Delta}^{\mathrm{ins}}_{e,q} \ge \Gamma_{\mathrm{fr}} - 3\varepsilon_q > 3\varepsilon_q$. Thus, Lemma 13 triggers freezing. Since $\varepsilon_q = \left(\sqrt{3}+\sqrt{9+3\sqrt{3}}\right) u_{\mathrm{se}}^{\frac{1}{2}-\alpha}\sqrt{\log(MNT^2)/n_q}$, the required epoch index is $O(\log\log T)$.

We next bound the regret before freezing. Define $\mathcal{B}_q \triangleq \mathcal{X}(\widehat{p}_q)\setminus\mathcal{X}(p^*)$ and $\mathcal{J}_q^- \triangleq \mathcal{J}_q\setminus\mathcal{X}(p^*)$. Since $\widehat{p}_q$ and p^* both contain M edges,

$$|\mathcal{J}_q \cap \mathcal{X}(p^*)| \le |\mathcal{X}(p^*)\setminus\mathcal{X}(\widehat{p}_q)| = |\mathcal{B}_q|.$$

Thus, Lemma 12 gives

$$R_{\mathrm{II},q} \le 12(g-1)n_q\varepsilon_q\big(|\mathcal{B}_q| + |\mathcal{J}_q^-|\big).$$

For $e\in\mathcal{B}_q$, the main matching itself is a feasible matching containing e, so $\Delta^{\mathrm{ins}}_e(\mathcal{A}_0) \le 3\varepsilon_q$. For $e\in\mathcal{J}_q^-$, Proposition 7 gives $\Delta^{\mathrm{ins}}_e(\mathcal{A}_0) \le \Delta_{e,q} \le 6\varepsilon_q$. Therefore, an edge with $\delta_e \triangleq \Delta^{\mathrm{ins}}_e(\mathcal{A}_0)$ can be charged only while $\delta_e \le 6\varepsilon_q$.

Since $n_q = n_0 g^q$,

$$\sum_{q:\,\delta_e\le 6\varepsilon_q} n_q\varepsilon_q = O\left(\frac{u_{\mathrm{se}}^{1-2\alpha}\log(MNT^2)}{\delta_e}\right).$$

Summing over all nonoptimal edges proves the Phase-II bound.

Finally, $\mathcal{A}_0 \subseteq \mathcal{M}\times\mathcal{N}$ implies

$$\Delta^{\mathrm{ins}}_e(\mathcal{A}_0) \ge \Delta^{\mathrm{ins}}_e(\mathcal{M}\times\mathcal{N}),$$

and Lemma 19 gives $u_{\mathrm{se}} = \Omega(\mu^*_{\min})$. Together with the Phase-I regret bound and $\mathbb{P}(\mathcal{E}^c) \le 5/T$, this yields the stated expected regret bound.

On $\mathcal{E}$, Phase I uses $O(\log\log T)$ auction calls by Corollary 5, and the freezing argument above gives the same order for Phase II. Since the geometric schedules yield at most $O(\log T)$ calls on any sample path, the contribution of $\mathcal{E}^c$ is $O(\log T/T)$, proving the expected auction-call bound. □

We finally give a truncation guarantee that does not require uniqueness.

Corollary 3 (Target-resolution truncation)**.** *Let $H_T = (\log T)^b$ for some $b > 0$, and define $Q_T = \min\{q \geq 0 : \varepsilon_q \leq H_T^{-1}\}$. If Algorithm 2 is truncated at epoch Q_T after shrinking and before refresh, and then replays the current auction matching, the number of auction calls is $O(\log\log T)$ on $\mathcal{E}$ and in expectation.*

On $\mathcal{E}$, the pre-truncation Phase-II regret is

$$O\big(|\mathcal{A}_0| u_{\mathrm{se}}^{1-2\alpha} \log(MNT^2) H_T\big)\,,$$

and the additional tail regret is $O(T/H_T)$. Consequently,

$$\mathbb{E}[R(T)] = \widetilde{O}\left(\frac{\Delta(\mathcal{M}\times\mathcal{N}) G_{\max}(\mu)}{\mu_{\min}^*} + MN(\mu_{\min}^*)^{1-2\alpha} H_T + \frac{T}{H_T}\right).$$

Proof. Only epochs $q < Q_T$ execute refresh. Each refresh family contains at most $1+|\mathcal{A}_0|$ matchings, and Lemma 12 bounds their per-round regret by $O(\varepsilon_q)$. Hence,

$$R_{\mathrm{pre}} = O\left(|\mathcal{A}_0| \sum_{q<Q_T} n_q \varepsilon_q\right).$$

Since $\varepsilon_q = (\sqrt{3}+\sqrt{9+3\sqrt{3}}) u_{\mathrm{se}}^{\frac{1}{2}-\alpha} \sqrt{\log(MNT^2)/n_q}$ and $\varepsilon_q > H_T^{-1}$ for $q < Q_T$, geometric summation gives

$$R_{\mathrm{pre}} = O\big(|\mathcal{A}_0| u_{\mathrm{se}}^{1-2\alpha} \log(MNT^2) H_T\big)\,.$$

At epoch Q_T, Lemma 12 gives per-round regret at most $3\varepsilon_{Q_T} \leq 3/H_T$, and hence $O(T/H_T)$ tail regret. The final expected bound follows from $|\mathcal{A}_0| \leq MN$, $u_{\mathrm{se}} = \Omega(\mu_{\min}^*)$, the Phase-I bound, and $\mathbb{P}(\mathcal{E}^c) \leq 5/T$. The auction-call bound follows from the geometric epoch schedule as in Theorem 6. □

D.2 Phase I Details

We now give the Phase I details omitted from the main text. The role of Phase I is to drive the decentralized system into the same safe regime used in the centralized analysis, while keeping the number of auction calls doubly logarithmic. It consists of four components: a decentralized exploration schedule, a completion routine that turns partial matchings into full feasible matchings, a decentralized auction, and a rare-auction checkpoint rule.

D.2.1 Decentralized Exploration

We first present the decentralized exploration algorithm, as shown in Algorithm 6. Its purpose is to ensure that all agent–arm pairs are explored in a balanced and decentralized fashion. We define

$$C_{\min} \triangleq \min_{n\in\mathcal{N}} C_n, \qquad G \triangleq \left\lceil \frac{M}{C_{\min}} \right\rceil.$$

Partition the agents into G ordered groups

$$\mathcal{G}_g \triangleq \{(g-1)C_{\min}+1,\ (g-1)C_{\min}+2,\ \ldots,\ \min\{gC_{\min}, M\}\}\,, \qquad g = 1,\ldots,G,$$

so that $|\mathcal{G}_g| \leq C_{\min} \leq C_n$ for every $n \in \mathcal{N}$. Next, we partition the G groups into $B \triangleq \lceil G/N \rceil$ ordered batches where batch b contains the groups

$$\mathcal{B}_b \triangleq \{(b-1)N+1,\ (b-1)N+2,\ \ldots,\ \min\{bN, G\}\}\,, \qquad b = 1,\ldots,B.$$

For a group index $o \in \mathcal{B}_b$, define its local position within batch b as

$$\rho(o) \triangleq o - (b-1)N \in \{1,\ldots,|\mathcal{B}_b|\}.$$

Algorithm 6 works as follows. Time is divided into super-cycles of length $L = BN$. At each round, only one batch of groups is active. Within the active batch, different groups are assigned to different arms through a cyclic shift, so no arm receives nominations from two active groups in the same round. Since each group size is at most $C_{\min}$, these nominations never violate any arm capacity. Across one super-cycle, the arm-phase index cycles through all N arms, so every group visits every arm exactly once.

Algorithm 6 Decentralized Exploration

Require: Capacities $\{C_n\}_{n\in\mathcal{N}}$.
Ensure: A partial feasible matching $\widehat{p}^t \in \{0,1\}^{M\times N}$.
1: Initialize $\widehat{p}^t_{mn} \leftarrow 0$ for all $(m,n) \in \mathcal{M}\times\mathcal{N}$;
2: Define the active batch index $b_t \triangleq 1 + ((t-1) \bmod B)$ and the arm-phase index $h_t \triangleq 1 + \left(\lfloor \frac{t-1}{B} \rfloor \bmod N\right)$;
3: **for** each group $o \in \mathcal{B}_{b_t}$ **do**
4: Define its nominated arm $a_o(t) \triangleq 1 + ((h_t + \rho(o) - 2) \bmod N)$;
5: **for** each agent $m \in \mathcal{G}_o$ **do**
6: Set $\widehat{p}^t_{m,a_o(t)} \leftarrow 1$;
7: **end for**
8: **end for**
9: **return** $\widehat{p}^t$.

The next proposition summarizes the properties of Algorithm 6 used later: partial feasibility, exact arm traversal, and uniform exploration counts.

Proposition 8 (Properties of decentralized exploration)**.** *Algorithm 6 satisfies the following properties.*

1. *For every round t, $\widehat{p}^t$ is a partial feasible matching: $\sum_{n\in\mathcal{N}} \widehat{p}^t_{mn} \le 1, \forall m \in \mathcal{M}$, and $\sum_{m\in\mathcal{M}} \widehat{p}^t_{mn} \le C_n, \forall n \in \mathcal{N}$.*
2. *In every super-cycle of length $L = BN$, each group visits every arm exactly once.*
3. *Let $\widetilde{N}^t_{mn}$ be the exploration count of pair (m,n) up to time t. Then after q full super-cycles, $\widetilde{N}^{qL}_{mn} = q, \forall (m,n) \in \mathcal{M}\times\mathcal{N}$.*

Proof. For partial feasibility, each agent belongs to one group and is assigned to at most one arm when its group is active. Within an active batch, the map $o \mapsto a_o(t) = 1 + ((h_t + \rho(o) - 2) \bmod N)$ is injective, so each arm receives agents from at most one active group. Since every group has size at most $C_{\min} \le C_n$, arm capacities are not violated.

For exact traversal, fix a group o in batch $b(o)$. During super-cycle q, the group is active at rounds

$$t_l^{(q,o)} = qL + (l-1)B + b(o), \qquad l = 1,\dots,N.$$

At these rounds, $h_{t_l^{(q,o)}} = l$, and hence

$$a_o(t_l^{(q,o)}) = 1 + ((l + \rho(o) - 2) \bmod N).$$

As l ranges from 1 to N, this visits every arm exactly once. Therefore, each agent in group o explores every arm exactly once per super-cycle, which gives $\widetilde{N}^{qL}_{mn} = q$. □

D.2.2 Decentralized Completion

Algorithm 6 returns a partial matching $\widehat{p}^t$ on the complete graph $\mathcal{M}\times\mathcal{N}$. To derive a full feasible matching at each round, we need to complete $\widehat{p}^t$ without violating arm capacities.

Let the unmatched-agent set at round t be

$$U_1^t \triangleq \left\{ m \in \mathcal{M} : \sum_{n\in\mathcal{N}} \widehat{p}^t_{mn} = 0 \right\},$$

and let the initial residual capacity of arm n be

$$r_n^t(1) \triangleq C_n - \sum_{m\in\mathcal{M}} \widehat{p}^t_{mn}.$$

We fix a common priority order $\pi : \mathcal{M} \to \{1,\dots,M\}$, known to all agents and all arms. For each agent $m \in \mathcal{M}$ and residual iteration $s = 1,\dots,N$, define the arm that m applies to in round s by

$$\phi_m(s) \triangleq 1 + \big((\pi(m) + s - 2) \bmod N\big).$$

Thus, every unmatched agent applies to every arm once within N residual rounds, while each arm accepts applicants up to its remaining capacity.

Algorithm 7 DComplete: Decentralized Completion

Require: A partial matching $\widehat{p}^t$ and capacities $\{C_n\}_{n\in\mathcal{N}}$.
Ensure: A feasible full matching p^t.

```
Set p^t ← p̂^t;
Set U_1^t ← {m ∈ M : Σ_{n∈N} p̂_{mn}^t = 0};
Set r_n^t(1) ← C_n − Σ_{m∈M} p̂_{mn}^t for all n ∈ N;
for s = 1, 2, ..., N do
  if U_s^t = ∅ then
    break;
  end if
  for each unmatched agent m ∈ U_s^t do
    Agent m applies to arm φ_m(s);
  end for
  for each arm n ∈ N do
    Let A_n^t(s) ≜ {m ∈ U_s^t : φ_m(s) = n} be the applicants to arm n;
    Arm n accepts the first min{r_n^t(s), |A_n^t(s)|} agents in A_n^t(s) according to the common
  priority order π;
    Denote the accepted set by B_n^t(s);
    for each m ∈ B_n^t(s) do
      Set p_{mn}^t ← 1;
    end for
    Set r_n^t(s+1) ← r_n^t(s) − |B_n^t(s)|;
  end for
  Set U_{s+1}^t ← U_s^t \ ⋃_{n∈N} B_n^t(s);
end for
return p^t.
```

We next present the three properties of Algorithm 7: it preserves the matching returned by Algorithm 6, returns a full feasible matching, and remains decentralized.

Proposition 9 (Correctness of DComplete)**.** *Algorithm 7 preserves all explored edges in $\widehat{p}^t$, extends $\widehat{p}^t$ to a full feasible matching p^t, and is decentralized.*

Proof. The algorithm initializes $p^t \leftarrow \widehat{p}^t$ and only adds edges for unmatched agents, so all explored edges are preserved. Each unmatched agent leaves the unmatched set after being accepted, and therefore receives at most one added edge. Each arm accepts at most its current residual capacity, and residual capacities are updated locally, so no capacity is violated.

It remains to show that all agents are matched. Suppose some agent remains unmatched after N residual rounds. This agent has applied to every arm once. If it was rejected by every arm, then every arm must have exhausted its residual capacity when it was applied to; since residual capacity never increases, all residual capacities are zero at the end. But the initial total residual capacity is at least the number of initially unmatched agents, because the original capacitated matching problem is feasible. Thus, all unmatched agents must have been accepted, a contradiction. The procedure is decentralized because agents only send local applications and arms only use their residual capacities, received applicants, and the public priority order. □

Corollary 4 (Exploration counts are preserved under completion)**.** *Let $\widetilde{N}_{mn}^t$ be the exploration count under Algorithm 6, and let N_{mn}^t be the actual pull count after Algorithm 7. Then $N_{mn}^t \geq \widetilde{N}_{mn}^t$ for every t and every (m, n).*

Proof. Every explored edge in $\widehat{p}^t$ is preserved in p^t, so every exploration becomes an actual pull. □

D.2.3 Auction Framework

We next describe the decentralized auction that is used both at Phase-I checkpoints and at every Phase-II epoch. The weighted α-fair objective becomes linear on integral matchings, so the optimization problem reduces to a capacitated bipartite matching problem.

At a checkpoint or shrinking epoch, define $\widehat{v}_{mn} \triangleq w_m U_\alpha(\tilde{\mu}_{mn})$ and $v_{mn} \triangleq w_m U_\alpha(\mu_{mn})$, where $\tilde{\mu}_{mn}$ is certain estimation of μ_{mn}, such as LCB, UCB, or $\bar{\mu}^t_{mn}$. Since every feasible matching $p \in \mathcal{P}(\mathcal{A})$ satisfies $\sum_n p_{mn} = 1$ for each agent m, the weighted α-fair objective is equivalent on integral matchings to the linear score

$$W_{\mathcal{A}}(p) \triangleq \sum_{(m,n)\in\mathcal{A}} v_{mn} p_{mn}, \qquad \hat{W}_{\mathcal{A}}(p) \triangleq \sum_{(m,n)\in\mathcal{A}} \widehat{v}_{mn} p_{mn}.$$

Therefore, the active-set problem is

$$\widehat{p}^{\star}_{\mathcal{A}} \in \arg\max_{p\in\mathcal{P}(\mathcal{A})} \hat{W}_{\mathcal{A}}(p). \tag{13}$$

Its LP dual is

$$\min_{a,\lambda} \quad Z_{\mathcal{A}}(a,\lambda) \triangleq \sum_{m\in\mathcal{M}} a_m + \sum_{n\in\mathcal{N}} C_n \lambda_n \tag{14}$$

$$\text{s.t.} \quad a_m + \lambda_n \geq \widehat{v}_{mn}, \qquad \forall (m,n) \in \mathcal{A}, \tag{15}$$

$$\lambda_n \geq 0, \qquad \forall n \in \mathcal{N}. \tag{16}$$

The dual variables have natural economic meanings. The dual variable λ_n is the scarcity price of arm n, and is thus high when arm n is congested and low when its capacity is slack. Given arm prices $(\lambda_n)_{n\in\mathcal{N}}$, we have

$$a_m = \max_{k:(m,k)\in\mathcal{A}} \big(\widehat{v}_{mk} - \lambda_k\big),$$

which is the best priced value that agent m can obtain. We define

$$\widehat{r}_{mn} \triangleq a_m + \lambda_n - \widehat{v}_{mn} = \max_{k:(m,k)\in\mathcal{A}} \big(\widehat{v}_{mk} - \lambda_k\big) - \big(\widehat{v}_{mn} - \lambda_n\big) \geq 0.$$

Hence, $\widehat{r}_{mn}$ is the loss in the priced value when agent m is forced to use arm n instead of its priced best choice.

Before running the auction, we repeatedly fix every singleton agent to its unique incident arm and reduce the corresponding residual capacity. Whenever an arm's residual capacity becomes zero, remove that arm and its incident edges from the residual instance. Record the removal order and run the auction on the remaining instance (if no agents remain, use an empty matching and agent-value vector, with $\lambda_n = 0$ for every remaining arm). After termination, undo the preprocessing in reverse order, restoring the agents, capacities, and edges at each step. When restoring a fixed assignment $m \to n$, keep all existing prices and agent net values unchanged. If arm n was removed when its residual capacity became zero, first reintroduce it and set

$$\lambda_n \leftarrow \max\Big(\{0\} \cup \{\widehat{v}_{in} - a_i : i \in \mathcal{I}_n\}\Big),$$

where $\mathcal{I}_n$ consists of the already restored or residual agents whose edges to n are reintroduced at this step. Otherwise, retain the current λ_n. Restore $m \to n$ and set $a_m \leftarrow \widehat{v}_{mn} - \lambda_n$. The maximum above is computed by arm n from messages of its incident agents and requires no additional auction call. All returned quantities are defined on the original active graph $\mathcal{A}$.

The basic idea of Algorithm 8 is as follows. At each iteration, every unmatched agent first finds its best arm under the current prices, and then compares the best arm with its second-best option. The gap between these two priced values determines how much the agent is willing to bid. Each arm keeps its best C_n bids, and its price becomes the smallest kept bid.

We next introduce the auction properties needed in both Phase I and Phase II: basic properties, dual feasibility, an auction lower bound, and near-optimality.

Lemma 14 (Basic properties of the decentralized auction)**.** *Suppose that Algorithm 8, with the singleton preprocessing and reverse restoration above, is applied to a feasible active instance with finite transformed values. Then:*

1. *Algorithm 8 terminates in finite iterations;*
2. *its output matching $\widehat{p}^{\,\mathrm{auc}}_{\mathcal{A}}$ belongs to $\mathcal{P}(\mathcal{A})$;*
3. *the arm prices are non-decreasing over iterations, i.e.,*

$$\lambda_n^{(j+1)} \geq \lambda_n^{(j)}, \qquad \forall n \in \mathcal{N},\ \forall j \geq 0;$$

Algorithm 8 Price-Threshold Dual-Margin Auction (PTDMA)

Require: Active edge set $\mathcal{A}$, transformed weights $\widehat{v} = \{\widehat{v}_{mn}\}_{(m,n)\in\mathcal{A}}$, capacities $\{C_n\}_{n\in\mathcal{N}}$, bid increment $\varepsilon > 0$.

Ensure: A feasible matching $\widehat{p}^{\mathrm{auc}}_{\mathcal{A}} \in \mathcal{P}(\mathcal{A})$, arm prices $(\lambda_n)_{n\in\mathcal{N}}$, and agent net values $(a_m)_{m\in\mathcal{M}}$.

Initialize $\lambda_n^{(0)} \leftarrow 0$, $\mathcal{S}_n^{(0)} \leftarrow \emptyset$ for all n, and $j \leftarrow 0$;
while there exists an unmatched agent **do**
  Each arm n broadcasts $\lambda_n^{(j)}$;
  for each currently unmatched agent m in parallel **do**
    Compute $a_m^{(j)} \leftarrow \max_{k\in\mathcal{N}_m(\mathcal{A})}(\widehat{v}_{mk} - \lambda_k^{(j)})$;
    Choose $n_m^\star \in \arg\max_{k\in\mathcal{N}_m(\mathcal{A})}(\widehat{v}_{mk} - \lambda_k^{(j)})$;
    Compute $s_m^{(j)} \leftarrow \max_{k\in\mathcal{N}_m(\mathcal{A})\setminus\{n_m^\star\}}(\widehat{v}_{mk} - \lambda_k^{(j)})$;
    Send bid $\beta_m^{(j)} \leftarrow \lambda_{n_m^\star}^{(j)} + a_m^{(j)} - s_m^{(j)} + \varepsilon$ to arm $n_m^\star$;
  end for
  for each arm n in parallel **do**
    Let $\mathcal{I}_n^{(j)}$ be the newly received bid pairs $(m, \beta_m^{(j)})$;
    Set $\mathcal{C}_n^{(j)} \leftarrow \mathcal{S}_n^{(j)} \cup \mathcal{I}_n^{(j)}$;
    Keep the top C_n pairs in $\mathcal{C}_n^{(j)}$, denoted by $\mathcal{S}_n^{(j+1)}$;
    Set $\lambda_n^{(j+1)} \leftarrow 0$ if $|\mathcal{S}_n^{(j+1)}| < C_n$; otherwise set $\lambda_n^{(j+1)}$ to the smallest kept bid;
  end for
  $j \leftarrow j + 1$;
end while
Recover $\widehat{p}^{\mathrm{auc}}_{\mathcal{A}}$ from the final winner sets $\{\mathcal{S}_n^{(j)}\}_{n\in\mathcal{N}}$;
Set $a_m \leftarrow \max_{n\in\mathcal{N}_m(\mathcal{A})}(\widehat{v}_{mn} - \lambda_n^{(j)})$ for all m;
return $(\widehat{p}^{\mathrm{auc}}_{\mathcal{A}}, a, \lambda)$.

4. *for every matched edge* (m, n) *with* $\widehat{p}^{\mathrm{auc}}_{mn} = 1$,
$$\widehat{v}_{mn} - \lambda_n \geq a_m - \varepsilon = \max_{k\in\mathcal{N}_m(\mathcal{A})}(\widehat{v}_{mk} - \lambda_k) - \varepsilon;$$
5. *if* $\lambda_n > 0$, *then* $\sum_{m\in\mathcal{M}} \widehat{p}^{\mathrm{auc}}_{mn} = C_n$.

Proof. The first property follows from the standard finite-termination property of the synchronized auction for similar objects / transportation assignment Bertsekas & Castanon (1989). We first verify the remaining properties on the residual instance, and then show that reverse restoration preserves them.

For the second property, feasibility follows because each arm keeps at most C_n winners and only unmatched agents bid.

For the third property, if $|\mathcal{S}_n^{(j+1)}| < C_n$, then $\lambda_n^{(j+1)} = 0$. If $|\mathcal{S}_n^{(j)}| < C_n$ as well, then $\lambda_n^{(j)} = 0$. If $|\mathcal{S}_n^{(j)}| = C_n$, then $\lambda_n^{(j)}$ is the smallest kept bid at round j. Since $\mathcal{S}_n^{(j+1)}$ is selected from the union of the old kept bids and the new bids, the smallest kept bid cannot decrease. Hence, $\lambda_n^{(j+1)} \geq \lambda_n^{(j)}$ in all cases.

For the fourth property, fix a matched edge (m, n) with $\widehat{p}^{\mathrm{auc}}_{mn} = 1$. Let τ be the last iteration at which agent m entered the winner set of arm n, and let b_m be the bid of agent m at that iteration. Then
$$b_m = \lambda_n^{(\tau)} + a_m^{(\tau)} - s_m^{(\tau)} + \varepsilon,$$
where $a_m^{(\tau)} = \widehat{v}_{mn} - \lambda_n^{(\tau)}$ and $s_m^{(\tau)} = \max_{k\in\mathcal{N}_m(\mathcal{A})\setminus\{n\}}(\widehat{v}_{mk} - \lambda_k^{(\tau)})$. Hence,
$$\widehat{v}_{mn} - b_m = s_m^{(\tau)} - \varepsilon.$$
Since agent m remains in the final winner set of arm n, the final arm price satisfies $\lambda_n \leq b_m$. Therefore,
$$\widehat{v}_{mn} - \lambda_n \geq \widehat{v}_{mn} - b_m = s_m^{(\tau)} - \varepsilon.$$
By the third property, all arm prices are non-decreasing, so for every $k \neq n$,
$$\widehat{v}_{mk} - \lambda_k \leq \widehat{v}_{mk} - \lambda_k^{(\tau)} \leq s_m^{(\tau)}.$$

Thus,

$$\widehat{v}_{mn} - \lambda_n \geq \max_{k \in \mathcal{N}_m(\mathcal{A}) \setminus \{n\}} (\widehat{v}_{mk} - \lambda_k) - \varepsilon.$$

Since $\widehat{v}_{mn} - \lambda_n$ is itself one candidate in $\max_{k \in \mathcal{N}_m(\mathcal{A})}(\widehat{v}_{mk} - \lambda_k)$, we have

$$\widehat{v}_{mn} - \lambda_n \geq \max_{k \in \mathcal{N}_m(\mathcal{A})} (\widehat{v}_{mk} - \lambda_k) - \varepsilon = a_m - \varepsilon.$$

For the fifth property, the price update rule sets $\lambda_n = 0$ when fewer than C_n winners are kept at arm n. Therefore, $\lambda_n > 0$ implies $|\mathcal{H}_n| = C_n$, which indicates $\sum_{m \in \mathcal{M}} \widehat{p}^{\text{auc}}_{mn} = C_n$.

Finally, consider one reverse-restoration step for a singleton $m \to n$. Immediately before its removal, m had only the neighbor n in that residual instance. If n remained present, its occupancy and capacity both increase by one, and all old prices and net values are unchanged. If n is reintroduced, its capacity and occupancy are both one, and the chosen price gives $\widehat{v}_{in} - \lambda_n \leq a_i$ for every newly restored edge (i, n) with $i \neq m$. Each existing agent retains an old maximizer, so its unchanged a_i is still its maximum priced value. In both cases the new agent satisfies $a_m = \widehat{v}_{mn} - \lambda_n$ with zero complementary-slackness error. Thus feasibility, dual feasibility, the ε-complementary-slackness bound, and the property that positive-price arms are full are preserved. Existing prices never change during restoration, and newly reintroduced prices are nonnegative and finite. Induction over the reverse removal order establishes these properties on the original instance. At most M restoration steps are needed, so the entire procedure terminates in finite time. □

The following lemma converts the auction prices into a feasible dual solution.

Lemma 15 (Dual feasibility from auction prices)**.** *Let $\lambda \geq 0$, and define $a_m = \max_{n \in \mathcal{N}_m(\mathcal{A})}(\widehat{v}_{mn} - \lambda_n)$. Then (a, λ) is feasible for the dual problem (14)–(15).*

Proof. For every $(m, n) \in \mathcal{A}$, $a_m \geq \widehat{v}_{mn} - \lambda_n$, so $a_m + \lambda_n \geq \widehat{v}_{mn}$. Together with $\lambda_n \geq 0$, this gives dual feasibility. □

The next lemma shows that the auction output and the auction prices are close enough to primal-dual optimality.

Lemma 16 (Auction lower bound from ε-complementary slackness)**.** *Under Lemmas 14 and 15,*

$$\hat{W}_{\mathcal{A}}(\widehat{p}^{\text{auc}}_{\mathcal{A}}) \geq Z_{\mathcal{A}}(a, \lambda) - M\varepsilon.$$

Proof. For every matched edge (m, n), Lemma 14 gives $\widehat{v}_{mn} \geq a_m + \lambda_n - \varepsilon$. Summing over the M matched edges yields

$$\hat{W}_{\mathcal{A}}(\widehat{p}^{\text{auc}}_{\mathcal{A}}) \geq \sum_m a_m + \sum_n \lambda_n \sum_m \widehat{p}^{\text{auc}}_{mn} - M\varepsilon.$$

If $\lambda_n > 0$, then arm n is full; if $\lambda_n = 0$, the equality $\lambda_n \sum_m \widehat{p}^{\text{auc}}_{mn} = C_n \lambda_n$ is trivial. Thus, the right-hand side equals $Z_{\mathcal{A}}(a, \lambda) - M\varepsilon$. □

Finally, combining ε-complementary slackness with dual feasibility yields a near-optimal primal matching.

Theorem 7 (Near-optimality of the decentralized auction)**.** *Under Lemmas 14–16,*

$$0 \leq \hat{W}_{\mathcal{A}}(\widehat{p}^{\star}_{\mathcal{A}}) - \hat{W}_{\mathcal{A}}(\widehat{p}^{\text{auc}}_{\mathcal{A}}) \leq M\varepsilon,$$

where $\widehat{p}^{\star}_{\mathcal{A}}$ is an optimal solution to (13).

Proof. By Lemma 15, (a, λ) is dual feasible, so weak duality gives

$$\hat{W}_{\mathcal{A}}(\widehat{p}^{\star}_{\mathcal{A}}) \leq Z_{\mathcal{A}}(a, \lambda).$$

Combining with Lemma 16 yields

$$\hat{W}_{\mathcal{A}}(\widehat{p}^{\star}_{\mathcal{A}}) - \hat{W}_{\mathcal{A}}(\widehat{p}^{\text{auc}}_{\mathcal{A}}) \leq M\varepsilon.$$

The left inequality is trivial because $\widehat{p}^{\star}_{\mathcal{A}}$ is optimal for the estimated problem. □

D.2.4 RARE-AUCTION EXPLORATION IN PHASE I

We combine decentralized exploration, completion, and the auction to obtain the safe scale and retained edge set while limiting the number of auction calls. Fix $g \geq 2$ and an initial checkpoint level $n_{\mathrm{ckpt}} \geq 1$. Define

$$n_\gamma \triangleq n_{\mathrm{ckpt}} g^\gamma, \qquad t_\gamma \triangleq L n_\gamma,$$

where $L = BN$ is the super-cycle length of Algorithm 6.

Positive-LCB restriction. At checkpoint γ, define

$$\mathcal{A}_\gamma^+ \triangleq \{(m,n) \in \mathcal{M} \times \mathcal{N} : \mathrm{LCB}_{mn}^{t_\gamma} > 0\}.$$

For a full feasible matching p, define the extended checkpoint objective by summing only over its selected edges:

$$W_\gamma^{\mathrm{LB}}(p) \triangleq \sum_{(m,n)\in\mathcal{X}(p)} w_m U_\alpha(\mathrm{LCB}_{mn}^{t_\gamma}).$$

If $\mathcal{P}(\mathcal{A}_\gamma^+) = \varnothing$, every full feasible matching contains a zero-LCB edge and therefore has objective value $-\infty$. In this case, set

$$\mathrm{LB}(t_\gamma) = -\infty, \qquad \underline{\mu}(t_\gamma) = \theta(t_\gamma) = 0,$$

skip the auction, and continue exploration.

Otherwise, define the finite scores

$$v_{mn}^{\mathrm{LB},\gamma} \triangleq w_m U_\alpha(\mathrm{LCB}_{mn}^{t_\gamma}), \qquad (m,n) \in \mathcal{A}_\gamma^+,$$

and run Algorithm 8 only on $\mathcal{A}_\gamma^+$. Let $p^{\mathrm{LB}}(t_\gamma)$ be its output, and set

$$\mathrm{LB}(t_\gamma) \triangleq W_\gamma^{\mathrm{LB}}(p^{\mathrm{LB}}(t_\gamma)),$$

$$\underline{\mu}(t_\gamma) \triangleq U_\alpha^{-1}\left(\frac{\mathrm{LB}(t_\gamma) - (1 - w_{\min})U_\alpha(1)}{w_{\min}}\right), \qquad \theta(t_\gamma) \triangleq \underline{\mu}(t_\gamma)/15.$$

This restriction does not truncate the objective. If $\mathcal{A}_\gamma^+$ admits a full feasible matching, every finite-valued matching lies in this graph, so

$$\max_{p\in\mathcal{P}(\mathcal{M}\times\mathcal{N})} W_\gamma^{\mathrm{LB}}(p) = \max_{p\in\mathcal{P}(\mathcal{A}_\gamma^+)} W_\gamma^{\mathrm{LB}}(p).$$

If no such matching exists, the left-hand side is $-\infty$. The restriction is recomputed at every checkpoint and is used only for the LCB auction. It does not remove edges from the exploration graph.

The feasibility check uses only edge incidences and capacities. Equivalently, connect a source to every agent with capacity one, connect agents to their incident arms in $\mathcal{A}_\gamma^+$ with capacity one, and connect arm n to a sink with capacity C_n. A full matching exists exactly when this network admits a flow of value M. The check can be implemented by capacity-feasible augmenting-path messages and does not invoke the numerical auction.

The following results justify the rare-auction phase.

Lemma 17 (Uniform checkpoint counts)**.** *For every checkpoint γ reached within the horizon, $t_\gamma = Ln_\gamma$ satisfies*

$$\widetilde{N}_{mn}^{t_\gamma} = n_\gamma, \qquad N_{mn}^{t_\gamma} \geq n_\gamma, \qquad \forall (m,n) \in \mathcal{M} \times \mathcal{N}.$$

Proof. By Proposition 8, each base-arm is nominated once in every super-cycle. There are exactly n_γ complete super-cycles before checkpoint t_γ. The actual count is no smaller than the nomination count by Corollary 4. □

Proposition 10 (Checkpoint lower bound from the auction)**.** *Define*

$$\mathrm{OPT}_\gamma^{\mathrm{LB}} \triangleq \max_{p\in\mathcal{P}(\mathcal{M}\times\mathcal{N})} W_\gamma^{\mathrm{LB}}(p).$$

At every checkpoint reached within the horizon,

$$\mathrm{LB}(t_\gamma) \geq \mathrm{OPT}_\gamma^{\mathrm{LB}} - M\epsilon_{\mathrm{ckpt}},$$

where the inequality is interpreted in the extended real numbers when $\mathrm{OPT}_\gamma^{\mathrm{LB}} = -\infty$.

Algorithm 9 Phase-I Rare-Auction Exploration

Require: Horizon T, capacities $\{C_n\}_{n\in\mathcal{N}}$, growth factor $g \geq 2$, initial checkpoint level $n_{\mathrm{ckpt}} \geq 1$, checkpoint bid increment $\epsilon_{\mathrm{ckpt}} > 0$.
Ensure: Stopping time t_{se}, safe scale u_{se}, and retained set $\mathcal{A}_{\mathrm{se}}$.
Initialize $\gamma \leftarrow 0$, $t \leftarrow 0$, and all empirical statistics;
while $t < T$ **do**
  Set $n_\gamma \leftarrow n_{\mathrm{ckpt}} g^\gamma$ and $t_\gamma \leftarrow L n_\gamma$;
  Run Algorithms 6 and 7, updating the empirical statistics, until $t = \min\{t_\gamma, T\}$;
  Compute LCB^{t_γ} and UCB^{t_γ};
  Set $\mathcal{A}_\gamma^+ \leftarrow \{(m,n) : \mathrm{LCB}_{mn}^{t_\gamma} > 0\}$;
  if $\mathcal{P}(\mathcal{A}_\gamma^+) = \varnothing$ **then**
    Set $\mathrm{LB}(t_\gamma) \leftarrow -\infty$ and $\underline{\mu}(t_\gamma) \leftarrow 0$, $\theta(t_\gamma) \leftarrow 0$;
  else
    Set $v_{mn}^{\mathrm{LB},\gamma} \leftarrow w_m U_\alpha(\mathrm{LCB}_{mn}^{t_\gamma})$ for $(m,n) \in \mathcal{A}_\gamma^+$;
    Run Algorithm 8 on $\mathcal{A}_\gamma^+$ with scores $v^{\mathrm{LB},\gamma}$ and bid increment ϵ_{ckpt}, obtaining $p^{\mathrm{LB}}(t_\gamma)$;
    Set $\mathrm{LB}(t_\gamma) \leftarrow \sum_{(m,n)\in\mathcal{A}_\gamma^+} v_{mn}^{\mathrm{LB},\gamma} p_{mn}^{\mathrm{LB}}(t_\gamma)$;
    Set $\underline{\mu}(t_\gamma) \leftarrow U_\alpha^{-1}\big((\mathrm{LB}(t_\gamma) - (1-w_{\min})U_\alpha(1))/w_{\min}\big)$ and $\theta(t_\gamma) \leftarrow \underline{\mu}(t_\gamma)/15$;
  end if
  if $\theta(t_\gamma) > 0$ **then**
    if $n_\gamma \geq 5\log(MNT^2)/\theta(t_\gamma)$ **then**
      Set $t_{\mathrm{se}} \leftarrow t_\gamma$ and $u_{\mathrm{se}} \leftarrow \frac{9}{5}\theta(t_\gamma)$;
      Set $\overline{\mathrm{UCB}}_m(t_{\mathrm{se}}) \leftarrow \max_{n\in\mathcal{N}} \mathrm{UCB}_{mn}^{t_{\mathrm{se}}}$ for every m;
      Compute $\mathrm{UB}_{mn}(t_{\mathrm{se}}) \leftarrow w_m U_\alpha(\mathrm{UCB}_{mn}^{t_{\mathrm{se}}}) + \sum_{i\neq m} w_i U_\alpha(\overline{\mathrm{UCB}}_i(t_{\mathrm{se}}))$ for every (m,n);
      Set $\mathcal{A}_{\mathrm{se}} \leftarrow \{(m,n) : \mathrm{UB}_{mn}(t_{\mathrm{se}}) \geq \mathrm{LB}(t_{\mathrm{se}})\}$;
      return $(t_{\mathrm{se}}, u_{\mathrm{se}}, \mathcal{A}_{\mathrm{se}})$;
    end if
  end if
  Set $\gamma \leftarrow \gamma + 1$;
end while

Proof. If $\mathcal{P}(\mathcal{A}_\gamma^+) = \varnothing$, then $\mathrm{LB}(t_\gamma) = \mathrm{OPT}_\gamma^{\mathrm{LB}} = -\infty$. Otherwise, restriction to $\mathcal{A}_\gamma^+$ preserves the optimal checkpoint value. Apply Theorem 7 to this finite-score feasible instance. □

Lemma 18 (Safe stopping under rare checkpoints)**.** *Define*

$$\beta_{\mathrm{auc}}^* \triangleq \frac{1}{15} U_\alpha^{-1}\left(\frac{U_\alpha(c_{\mathrm{L}}\mu_{\min}^*) - M\epsilon_{\mathrm{ckpt}} - (1-w_{\min})U_\alpha(1)}{w_{\min}}\right),$$

$$N_{B,\mathrm{auc}} \triangleq \left\lceil \frac{5\log(MNT^2)}{\beta_{\mathrm{auc}}^*} \right\rceil, \qquad \gamma^\dagger \triangleq \max\left\{0, \left\lceil \log_g\left(\frac{N_{B,\mathrm{auc}}}{n_{\mathrm{ckpt}}}\right)\right\rceil\right\},$$

and

$$t^\dagger \triangleq L n_{\mathrm{ckpt}} g^{\gamma^\dagger}.$$

Then

$$L N_{B,\mathrm{auc}} \leq t^\dagger < Lg\max\{n_{\mathrm{ckpt}}, N_{B,\mathrm{auc}}\}.$$

Conditioned on $\mathcal{E}_1 \cap \mathcal{E}_2$, if $t^\dagger \leq T$, Algorithm 9 triggers the stopping rule no later than $t^\dagger$.

Proof. If $n_{\mathrm{ckpt}} \geq N_{B,\mathrm{auc}}$, then $\gamma^\dagger = 0$ and $t^\dagger = L n_{\mathrm{ckpt}}$. Otherwise, the ceiling in $\gamma^\dagger$ gives

$$N_{B,\mathrm{auc}} \leq n_{\mathrm{ckpt}} g^{\gamma^\dagger} < g N_{B,\mathrm{auc}}.$$

These two cases establish the stated bounds on $t^\dagger$.

Suppose the algorithm has not stopped before $t^\dagger$. By Lemma 17,

$$N_{mn}^{t^\dagger} \geq n_{\gamma^\dagger} \geq N_{B,\mathrm{auc}}$$

for every base-arm. Since $\beta_{\mathrm{auc}}^* \leq \beta^*$, the count condition of Lemma 6 holds for every optimal edge. Therefore,

$$\mathrm{LCB}_{m,n^*(m)}^{t^\dagger} \geq c_{\mathrm{L}}\mu_{m,n^*(m)} \geq c_{\mathrm{L}}\mu_{\min}^* > 0.$$

The true optimal matching thus belongs to $\mathcal{P}(\mathcal{A}^+_{\gamma^\dagger})$. In particular, the feasibility test succeeds and the checkpoint auction is executed.

Moreover,

$$\mathrm{OPT}^{\mathrm{LB}}_{\gamma^\dagger} \geq W^{\mathrm{LB}}_{\gamma^\dagger}(p^*) \geq U_\alpha(c_{\mathrm{L}}\mu^*_{\min}).$$

Proposition 10 implies

$$\mathrm{LB}(t^\dagger) \geq U_\alpha(c_{\mathrm{L}}\mu^*_{\min}) - M\epsilon_{\mathrm{ckpt}}.$$

By the definition of β^*_{auc},

$$\underline{\mu}(t^\dagger) \geq 15\beta^*_{\mathrm{auc}}, \qquad \theta(t^\dagger) \geq \beta^*_{\mathrm{auc}}.$$

Consequently,

$$n_{\gamma^\dagger} \geq N_{B,\mathrm{auc}} \geq \frac{5\log(MNT^2)}{\theta(t^\dagger)},$$

so the stopping condition is satisfied. □

Lemma 19 (Safe retained scale)**.** *Conditioned on $\mathcal{E}$, at the stopping checkpoint of Algorithm 9, all optimal edges are retained, and every retained edge satisfies*

$$\mu_{mn} > u_{\mathrm{se}} \qquad \textit{and} \qquad u_{\mathrm{se}} = \Omega(\mu^*_{\min}),$$

where the hidden constant depends only on α, w, and $M\epsilon_{\mathrm{ckpt}}$.

Proof. The proof of Proposition 2 applies because $p^{\mathrm{LB}}(t_{\mathrm{se}})$ is feasible and $\mathrm{LCB}^{t_{\mathrm{se}}} \leq \mu$, so $\mathrm{LB}(t_{\mathrm{se}}) \leq F_\alpha(p^*, \mu)$. Hence, no optimal edge is eliminated.

The proof of Proposition 3 also applies with the checkpoint lower bound, and shows that every edge with $\mu_{mn} \leq \frac{9}{5}\theta(t_{\mathrm{se}})$ is eliminated. Therefore, every retained edge satisfies $\mu_{mn} > u_{\mathrm{se}}$.

Finally, suppose $\theta(t_{\mathrm{se}}) < \beta^*_{\mathrm{auc}}$. The stopping condition then gives $N^{t_{\mathrm{se}}}_{mn} > 5\log(MNT^2)/\beta^*_{\mathrm{auc}}$ for every edge. Repeating the argument in Lemma 18 yields $\mathrm{LB}(t_{\mathrm{se}}) \geq U_\alpha(c_{\mathrm{L}}\mu^*_{\min}) - M\epsilon_{\mathrm{ckpt}}$, and hence $\theta(t_{\mathrm{se}}) \geq \beta^*_{\mathrm{auc}}$, a contradiction. Since $\beta^*_{\mathrm{auc}} = \Theta(\mu^*_{\min})$, we obtain $u_{\mathrm{se}} = \Omega(\mu^*_{\min})$. □

Corollary 5 (Main auction complexity of Phase I)**.** *For every fixed instance, conditioned on $\mathcal{E}$, Algorithm 9 uses $O(\log\log T)$ checkpoint auction calls.*

Proof. Each checkpoint invokes the auction at most once. Checkpoints without a full positive-LCB matching invoke no auction. For fixed instance parameters, $N_{B,\mathrm{auc}} = O(\log T)$, and hence $\gamma^\dagger = O(\log\log T)$. The result follows from Lemma 18. □

E Additional Experiments

E.1 Additional Experiments for the Centralized Algorithm

Figures 5–7 provide additional centralized experiments under different system sizes and fairness parameters. Our algorithm consistently achieves the lowest cumulative regret, while FE performs poorly because its max-min objective is not aligned with the weighted α-fair objective. UCB, ETU, and SE also incur larger regret, since they either lack adaptive safe-entry detection or rely on arm-wise elimination that can be unsafe under capacity coupling. Figure 7 further shows that the Phase-I regret of our algorithm increases sharply as $\mu^*_{\min}$ decreases, confirming that smaller safe scales require substantially more exploration in the non-Lipschitz regime.

E.2 Additional Experiments for the Decentralized Algorithm

We compare our decentralized algorithm with two auction-based baselines. The first is the *Per-edge Auction* baseline, which follows the same shrinking-and-refresh structure as ours but computes each insertion gap by solving one constrained auction for every non-selected edge. The second is the *Naive Auction* baseline, which replaces the centralized oracle by a decentralized auction and calls the auction at every round. We present two results under different system sizes.

Figures 8–10 present the auction complexity and regret–complexity tradeoff under $M = 8$, $N = 5$, $\alpha = 1.5$, and $T = 4 \times 10^6$. Figures 8 and 9 show that our method requires only 20 total auction calls

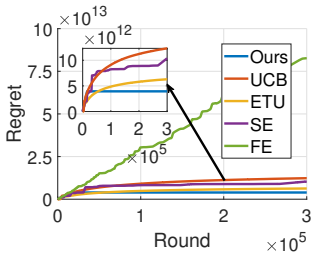


Figure 5: Regret with $M = 8, N = 5, \alpha = 20$

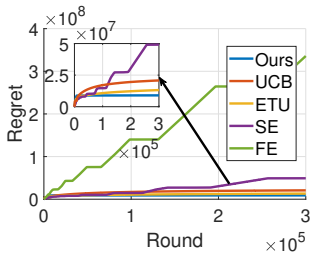


Figure 6: Regret with $M = 10, N = 3, \alpha = 10$

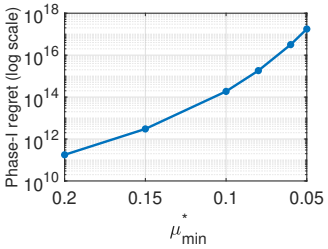


Figure 7: Phase-I regret for different $\mu^*_{\min}$

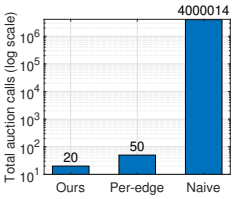


Figure 8: Total auction calls

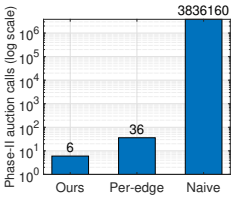


Figure 9: Phase-II auction calls

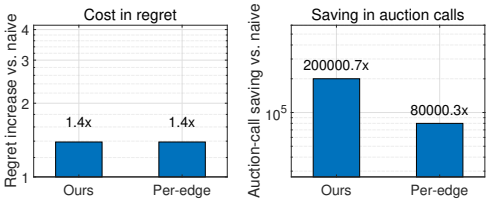


Figure 10: Regret and auction calls tradeoff

and 6 Phase-II auction calls, while the Per-edge Auction baseline needs 50 and 36 calls, respectively. This validates the benefit of computing all insertion gaps through the shared shortest-path reallocation procedure rather than through one constrained auction per edge. The Naive Auction baseline achieves lower regret because it calls an auction at every round, but this comes at the cost of about 4×10^6 auction calls. Figure 10 further illustrates this tradeoff: our method incurs about 1.4 times the regret of the Naive Auction baseline while reducing the auction-call count by a factor of approximately 2×10^5. Thus, our algorithm substantially reduces the number of auction calls while maintaining controlled regret.

Figures 11–13 show the same trend under $M = 12$, $N = 5$, $\alpha = 2$, and $T = 6 \times 10^6$. As shown in Figures 11 and 12, our method uses only 21 total auction calls and 7 Phase-II auction calls, whereas the Per-edge Auction baseline requires 56 total calls and 42 Phase-II calls. The Naive Auction baseline again achieves lower regret by calling the auction almost every round, but it requires about 6×10^6 total auction calls. Figure 13 shows that our method incurs about 1.3 times the regret of the Naive Auction baseline while reducing the auction-call count by a factor of approximately 2.86×10^5.

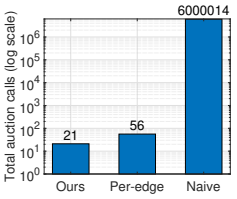


Figure 11: Total auction calls

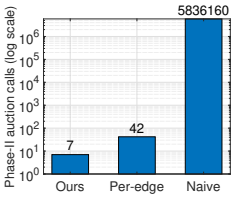


Figure 12: Phase-II auction calls

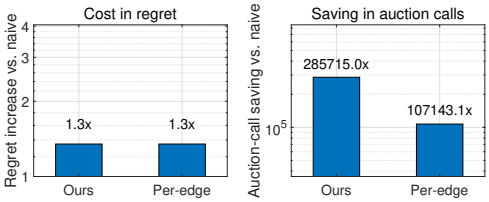


Figure 13: Regret and auction calls tradeoff