arXiv:2503.09762v2 [cs.DS] 23 Aug 2025

Achieving constant regret for dynamic matching via state-independent policies

Süleyman Kerimov, Pengyu Qian, Mingwei Yang, and Sophie H. Yu*

August 26, 2025

Abstract

We study a centralized discrete-time dynamic two-way matching model with finitely many agent types. Agents arrive stochastically over time and join their type-dedicated queues waiting to be matched. We focus on state-independent greedy policies that achieve constant regret at all times by making matching decisions based solely on agent availability across types, rather than requiring complete queue-length information. Such policies are particularly appealing for life-saving applications such as kidney exchange, as they require less information and provide more transparency compared to state-dependent policies.

First, for acyclic matching networks, we analyze a deterministic priority policy proposed by [KAG23] that follows a static priority order over matches. We derive the first explicit regret bound in terms of the general position gap (GPG) parameter ϵ—which measures the distance of the fluid relaxation from degeneracy. Second, for general two-way matching networks, we design a randomized state-independent greedy policy that achieves constant regret with optimal scaling $O(\epsilon^{-1})$, matching the existing lower bound established by [KAG24].

Contents

1 Introduction **2**
1.1 Related literature . . . 4
1.2 Notation . . . 5

2 Model setup **5**
2.1 Optimality criterion . . . 6
2.2 Static-planning and general position gap . . . 7
2.3 Greedy policies . . . 9

3 Main results **10**
3.1 Static priority policy on acyclic graphs . . . 10
3.2 Randomized state-independent greedy policy . . . 12

4 Analysis of static priority policy on acyclic graphs **13**
4.1 Warm-up: regret analysis on paths . . . 13
4.2 Proof of Theorem 1 . . . 16

*S. Kerimov is with the Jones Graduate School of Business, Rice University, Houston TX, USA, kerimov@rice.edu. P. Qian is with the Questrom School of Business, Boston University, Boston MA, USA, pqian@bu.edu. M. Yang is with the Department of Management Science and Engineering, Stanford University, Stanford CA, USA, mwyang@stanford.edu. S. H. Yu is with The Wharton School of Business, University of Pennsylvania, Philadelphia PA, USA, hysophie@wharton.upenn.edu.

5 Analysis of randomized state-independent greedy policy **19**

6 Consistent greedy policy **21**

7 Conclusion **22**

A Preliminaries on Lyapunov function analysis **27**

B Postponed proofs in Section 4 **28**
B.1 Proof of Proposition 3 . . . 28
B.2 Proof of Claim 1 . . . 30
B.3 Proof of Claim 2 . . . 30
B.4 Proof of Lemma 7 . . . 31
B.5 Proof of Lemma 8 . . . 31
B.6 Proof of Lemma 9 . . . 32
B.7 Proof of Lemma 10 . . . 33
B.8 Proof of Lemma 11 . . . 34

C Postponed proofs in Section 6 **35**
C.1 Proof of Proposition 4 . . . 35
C.2 Proof of Corollary 1 . . . 37

D Extension to non-empty initial state **37**

1 Introduction

Dynamic matching markets, where agents arrive over time and must be matched based on availability and compatibility, arise in numerous applications including kidney exchange markets (e.g., see [ABB+18]), online carpooling platforms (e.g., see [ÖW20]), and logistics (e.g. see [ET23]). A fundamental challenge in these markets is to design matching policies that are both efficient and practical to implement. While state-dependent policies that utilize complete queue-length information can achieve strong theoretical guarantees, they may be impractical or undesirable in many settings; they require maintaining and communicating detailed state information, can be complex to implement, and may be vulnerable to manipulation when participants strategically misreport queue-lengths to receive higher priorities.

As opposed to state-dependent policies, *state-independent policies* are more difficult to manipulate, and they may provide more transparency for agents in life-saving applications such as organ transplantation.[1] In the context of kidney exchange, hospitals may misreport information regarding the exact length of queues (the number of donor-patients waiting to exchange a kidney with certain blood types and sensitivity levels) if state-dependent policies are employed by the national kidney exchange platforms that hospitals are participating in. Whereas, state-independent policies only require the information on whether certain queues are empty or not at a particular time, which

[1] On March 2024, Memorial Hermann hospital in Houston, Texas, has halted their liver and kidney transplant programs over "a pattern of irregularities" [Chr24]. Recently, a New York Times investigation reported that American organ transplant system is "in chaos"; although federal rules try to ensure that donated organs are offered to the patients with respect to a priority order, officials tend to skip patients in the waitlists [Tim25]. It is also reported that skipping patients happens often so that a national kidney registry committee is too overwhelmed to examine each case closely.

indeed requires less information to perform matches. However, relying on less information usually comes with costs in performance guarantees [AAS25].

Following [KAG23], we focus on state-independent policies in two-way matching networks, where each match consists of exactly two agent types. Two-way matching captures many practical scenarios such as two-way kidney exchanges, where two incompatible patient-donor pairs swap their intended donors, and ride-hailing platforms, where each driver matches with exactly one rider. Under this model, agents arrive stochastically over a time horizon T and join their type-dedicated queues waiting to be matched. The performance of a policy is measured through an *all-time regret* performance metric—the maximum difference between the online policy's accumulated reward and the optimal offline matching value up to time t for all $t \in [T]$. Achieving *all-time constant regret* is particularly significant as it indicates the policy successfully balances short-term and long-term rewards, maintaining its performance guarantees at all times, not only at the end of the time horizon T. In this setting, *greedy policies*, which perform matches whenever possible based on current availability, are particularly appealing due to their simplicity in implementation, and effectiveness in balancing immediate and future rewards.

A key condition that enables this strong performance guarantee is the *general position gap* (GPG) condition, which ensures that the fluid relaxation of the dynamic matching problem has a unique non-degenerate optimal solution. The GPG parameter ϵ measures the distance of the fluid relaxation from degeneracy [KAG24, KAG23, Gup24, WXY23], quantifying both the market's inherent thickness and operational stability. Understanding how a policy's all-time regret scales with network parameters, e.g., the GPG parameter ϵ and network structure parameters, is crucial, as these relationships reveal the policy's stability across different operating environments. Remarkably, under the GPG condition, no policy (even with complete state information) can achieve a regret better than $O(\epsilon^{-1})$ at all times [KAG24]. In this work, we attempt to match this lower bound with state-independent greedy policies.

Under the GPG condition, [KAG23] introduced a state-independent static priority policy for two-way acyclic networks. This policy follows a fixed priority order over matches and achieves constant regret at all times, independent of the time horizon T. While the policy is state-independent and relies only on local queue availability information (i.e., whether each neighboring queue of the arriving agent is empty or not) rather than full queue-length information, its regret scaling in terms of problem parameters is unclear. In contrast, the longest-queue greedy policy of [KAG23] for two-way matching networks achieves an optimal regret scaling of $O(\epsilon^{-1})$ but requires complete queue-length information of neighboring queues, making it state-dependent. Meanwhile, there exist other greedy-like approaches for multi-way matching, such as the primal-dual policy of [WXY23] and the sum-of-square policy of [Gup24]. However, these methods neither maintain state-independence nor align with our natural greedy matching framework. While these state-dependent policies achieve the optimal scaling $O(\epsilon^{-1})$, the landscape for state-independent policies remains largely unexplored.

This raises two fundamental questions: (1) How does the regret of the static priority policy scale with the GPG and network structure parameters? (2) Can we design state-independent policies that achieve the optimal scaling for general two-way matching networks?

In this paper, we answer the above two questions affirmatively by providing a comprehensive analysis of state-independent policies in two-way matching networks, and we summarize our contributions as follows. First, we derive an explicit regret bound for the static priority policy on acyclic networks, showing that the regret scales as $O(\epsilon^{-1}(1 + 1/\epsilon)^{\lfloor (d_r-1)/2 \rfloor})$, where d_r is the depth of the network (Theorem 1). While the scaling with network depth suggests limitations of our techniques, this characterization represents a significant step toward understanding the fundamental trade-offs in state-independent policy design. We achieve this by simplifying and improving the Lyapunov techniques from [KAG23].

Second, we introduce a novel randomized state-independent greedy policy that achieves the optimal $O(\epsilon^{-1})$ regret scaling for general networks, including those with cycles (Theorem 2). This result demonstrates that state independence does not inherently require compromising performance guarantees. However, compared to static priority policies that only need local queue availability information, our randomized policy requires global queue availability information when making decisions at each step. This illustrates a fundamental trade-off between implementation complexity and performance guarantees for state-independent policies.

To derive these regret bounds, our strategy is to first establish a regret bound under the stationary distribution of the system and then translate it into a regret bound at all times. To complete this translation, one natural approach is to couple the Markov chain starting from an arbitrary initial state with the Markov chain starting from the steady-state, and argue that the distance between the states in these coupled chains can be controlled over time.[2] This motivates us to introduce a concept of *consistency* (Definition 4), and a greedy policy is consistent if under this policy, for an arbitrary sequential arrival process, when starting from two different initial states, the distance between the states in these two systems does not grow over time.

[KAG23] claim that all deterministic greedy policies are consistent, which is not correct as we illustrate via an example (Example 1). Instead, we identify a sufficient and easily-verifiable condition for a greedy policy to satisfy consistency (Proposition 4), via which we further show that various natural greedy policies are consistent (Corollary 1). Even though it is common to study long-run objectives and asymptotical behaviors of matching policies in the literature (see, e.g., [GW15, NS19, ADSW24]), it is of practical importance to derive all-time regret bounds, since a large error might occur before the Markov chain is sufficiently mixed. We believe that our results for consistency can be further applied in other contexts.

1.1 Related literature

Dynamic matching has been extensively studied in various settings. We review several streams of literature most relevant to our work.

Multi-way dynamic matching. Starting from [KAG24], the multi-way dynamic matching problem has received extensive attention recently, and the optimal regret scaling of $O(\epsilon^{-1})$ has been achieved via different policies. [KAG24] propose a batching policy, which performs a maximum weighted matching periodically. [Gup24] shows that the prominent sum-of-square policy achieves an optimal regret scaling even under a more general model. Later on, [WXY23] design a policy based on the primal-dual framework, which can achieve the optimal regret scaling even when the arrival rates are unknown. [KAG23] consider two-way matching networks and focus on greedy policies. They show that the longest-queue policy achieves the optimal regret scaling, and they propose a static priority policy that achieves constant regret at all times only for acyclic matching networks without characterizing how the regret depends on the problem parameters, e.g., the GPG parameter ϵ. We remark that all the above policies, except for the static priority policy, are state-dependent.

Dynamic matching on fixed graphs. Similar to our work, numerous papers study the dynamic matching problem with a fixed matching configuration by imposing different modeling assumptions. Most literature assumes that each match consists of exactly two agents, where the

[2]There are also other approaches to complete the translation. For example, [KAG24] achieve this via studying an exponential Lyapunov function to establish the geometric recurrence of the Markov chain.

underlying matching network can be either bipartite [ADSW24, ÖW20, CHS24, KSSW22] or non-bipartite [AS22, CILB+20]. A common feature of these literature is that they all assume that agents depart stochastically, and different objectives are considered, such as minimizing the holding costs [BM15], maximizing the match-specific rewards [CILB+20], or maximizing the generated rewards minus the holding costs [ADSW24].

Dynamic matching on random graphs. There is a vast literature on dynamic matching models based on random graphs, where agents arrive over time and can possibly form edges with the existing agents with fixed probabilities [AAGK17, ALG20, ABJM19]. A common assumption in this literature is that match values are homogeneous so that the focus is on minimizing the number of unmatched agents, and the general finding of this literature is that acting greedily is asymptotically optimal. Recently, [BRS+22] consider a setting where match values are heterogeneous, and as the market grows large, they show that greedy threshold policies are asymptotically optimal. Another line of research assumes that the agents lie in a metric space, and the weight of an edge between two agents is determined by the distance between them [Kan21, BFP23, SYY24].

Network revenue management. In the network revenue management (NRM) problem, there exist offline resources, and online requests that consume certain amounts of offline resources arrive dynamically. [TVR98] achieve $O(\sqrt{T})$ regret in the (quantity-based) NRM model via the bid-price policy. Later on, [JK12] improve the regret to be $O(1)$ via a re-solving policy under the GPG condition. Since then, constant regret is achieved when the arrival rates are unknown [Jas15], when the GPG condition does not hold [AG19, BW20, VB19, VBG21, JMZ22], and for a variety of other related problems [BBP24]. In contrast, for the multi-way dynamic matching problem considered in this paper, without the GPG condition, a regret lower bound of $\Omega(\sqrt{T})$ exists [KAG24].

Stability of stochastic matching systems. The multi-way dynamic matching model is closely related to the study of stability of stochastic matching systems [MM16, RM21, JMRSL22]. The connection is established by, e.g., [KAG24, Lemma 4.1] and [Gup24, Lemma 1], which assert that bounded all-time regret is implied by the stability of queues that are fully utilized by the fluid relaxation, provided that only the matches actively utilized by the fluid relaxation are performed. For the special case of two-way matching systems, [MM16] identify sufficient and necessary conditions for stability, and [JMRSL22] show that several policies, including the longest-queue policy and a generalized max-weight policy, achieve the maximal stability region.

1.2 Notation

For $x, y \in \mathbb{R}$, we use $x \wedge y$ to denote $\min\{x, y\}$ and x^+ to denote $\max\{x, 0\}$. For $n \in \mathbb{Z}_{>0}$, we use $[n]$ to denote the set $\{1, \ldots, n\}$. For a vector $v \in \mathbb{R}^n$ and an index set $J \subset [n]$, we use v^+ to denote $(v_1^+, \ldots, v_n^+)$ and use $v_J \in \mathbb{R}^{|J|}$ to denote the vector obtained by restricting v on J. We use $\mathbf{e}_i$ to denote the i-th standard basis, i.e., the i-th entry of $\mathbf{e}_i$ is 1 with all other entries being 0. We use standard asymptotic notation: for two positive sequences $\{x_n\}$ and $\{y_n\}$, we write $x_n = O(y_n)$ if $x_n \leq C y_n$ for an absolute constant C and for all n; $x_n = \Omega(y_n)$ if $y_n = O(x_n)$.

2 Model setup

We study a centralized dynamic matching market. There are n types of agents $\mathcal{A} = [n]$ and d types of matches $\mathcal{M} = [d]$. For each $m \in \mathcal{M}$, the value of performing a match m is denoted by

$r_m > 0$. Following [KAG23], in this paper, we focus on two-way matching structures, i.e., each match $m \in \mathcal{M}$ consists of exactly two agent types. We encode the matching structure into a *matching matrix* $M \in \{0,1\}^{\mathcal{A}\times\mathcal{M}}$, where $M_{im} = 1$ if and only if match m involves agent type i for all $i \in \mathcal{A}$ and $m \in \mathcal{M}$. Assume without loss of generality that each agent type participates in at least one match. We denote $(\mathcal{A}, \mathcal{M})$ as the (undirected) graph where vertices are formed by agent types and edges are formed by matches. For each agent type $i \in \mathcal{A}$, denote $\mathcal{N}(i)$ as the set of neighbors of i in the graph $(\mathcal{A}, \mathcal{M})$. If there is a match in $\mathcal{M}$ that contains agent types i and j, we will use $m(i, j)$ to denote this match.

We consider discrete-time arrivals where exactly one agent arrives and joins the type-dedicated queue at each time $t \in [T]$. For simplicity, we assume that there are no preexisting agents in the system.[3] Let $\lambda \in \mathbb{R}^n_{>0}$ denote a probability distribution over agent types, where λ_i represents the probability of an arriving agent being type i, and $\sum_{i=1}^n \lambda_i = 1$. For each type i, let $A_i(t)$ denote the cumulative number of type i arrivals up to and including time t, with $\Delta A_i(t) := A_i(t) - A_i(t-1) \in \{0, 1\}$ indicating whether an agent of type i arrives at time $t > 0$. At each period t, the central decision maker can perform match $m \in \mathcal{M}$ only if there are waiting agents of each type contained in m. This match then generates a reward r_m, and the agents participating in the match leave the system. For the sequence of events within each period, we always assume that matches are performed after the arrival of an agent. We refer to the tuple $\mathcal{G} = (M, \lambda, r)$ as the *matching network*, which captures both the matching structure and the arrival process.

A (randomized) *matching policy* decides how to (randomly) perform matches at each period. We will restrict our attention to non-anticipative policies, whose decisions at each moment only depend on the events happened so far. Given a policy, for all match $m \in \mathcal{M}$ and time $t \geq 0$, we denote $D_m(t)$ as the number of match m performed by the policy up to and including time t, and denote $\Delta D_m(t) := D_m(t) - D_m(t-1)$ as the number of match m performed at time $t > 0$. Define $Q_i(t) := (A(t) - MD(t))_i$ as the length of each queue $i \in \mathcal{A}$ after time $t \geq 0$, and we refer to $Q(t)$ as the *state* of the system after time t.

2.1 Optimality criterion

The expected total value generated by a policy Π during the first t periods is denoted by $\mathcal{R}^\Pi(t) := \mathbb{E}[r^T D(t)]$. Let $\mathcal{R}^*(t)$ be the expected value attained by the policy that takes no action before time t and performs matches that maximize the overall rewards up to time t. Formally,

$$\mathcal{R}^*(t) := \mathbb{E}\left[\begin{array}{cl} \max & r^T y \\ \text{s.t.} & My \leq A(t) \\ & y \in \mathbb{Z}^d_{\geq 0} \end{array}\right], \tag{1}$$

and $\mathcal{R}^*(t)$ is straightforwardly an upper bound for $\mathcal{R}^\Pi(t)$ for every policy Π. Define the *regret* of a policy Π after time t as $\mathcal{R}^*(t) - \mathcal{R}^\Pi(t)$, and define the *all-time regret* of Π after time T as

$$\text{Regret}(\Pi, T) := \sup_{0 \leq t \leq T} (\mathcal{R}^*(t) - \mathcal{R}^\Pi(t)).$$

We say that Π achieves constant regret at all times (or all-time constant regret), if $\text{Regret}(\Pi, T)$ is upper bounded by a constant that does not depend on T. Note that this performance metric differs from simply achieving exact hindsight optimality at the end of time horizon T, which can be attained by a trivial policy that delays all matches until the end of the time horizon T and then

[3] This is a common assumption made by the literature [KAG23, WXY23], and we refer to Appendix D for discussions on an extension to non-empty initial state, i.e., with the presence of preexisting agents.

solves an optimization problem to maximize overall rewards. In other words, to achieve all-time constant regret, a policy must effectively balance short-term and long-term rewards, ensuring that decisions made at each step do not compromise overall performance.

2.2 Static-planning and general position gap

We consider the fractional relaxation of the integer programming in (1). In particular, by applying Jensen's inequality,

$$\mathcal{R}^*(t) = \mathbb{E}\left[\begin{array}{ll} \max & r^T y \\ \text{s.t.} & My \leq A(t) \\ & y \in \mathbb{Z}^d_{\geq 0} \end{array}\right] \leq \begin{array}{ll} \max & r^T x \\ \text{s.t.} & Mx \leq t\lambda. \\ & x \in \mathbb{R}^d_{\geq 0} \end{array} \tag{2}$$

Replacing $z = x/t$ and adding slack variables $(s_i)_{i\in\mathcal{A}}$, the linear programming in the RHS can be rewritten as

$$\mathrm{SPP}(\lambda) := \begin{array}{ll} \max & r^T z \\ \text{s.t.} & Mz + s = \lambda \\ & x \in \mathbb{R}^d_{\geq 0}, s \in \mathbb{R}^n_{\geq 0} \end{array},$$

which we refer to as the *static-planning problem.*

Next, we introduce the notion of general position gap that captures the stability level of $\mathrm{SPP}(\lambda)$. This condition is proved necessary for any policy to achieve constant regret at all times (see, e.g., [KAG24, Example 3.1]). We note that the definition of general position gap we adopt coincides with that in [KAG23, KAG24].

Definition 1 (General position gap)**.** A matching network $\mathcal{G}$ satisfies the *general position gap* (GPG) condition if $\mathrm{SPP}(\lambda)$ has a unique non-degenerate optimal solution (z^*, s^*), i.e., all n basic variables in this solution are strictly positive. When $\mathcal{G}$ satisfies the GPG condition, define the GPG parameter ϵ as the minimum value of all basic variables, i.e.,

$$\epsilon := \min_{m\in\mathcal{M}, z^*_m>0} z^*_m \wedge \min_{i\in\mathcal{A}, s^*_i>0} s^*_i. \tag{3}$$

The GPG condition is a standard assumption in online revenue management and dynamic matching literature (see, e.g., [JK12, CLY24, KAG24]), and any linear programming can satisfy this condition with an arbitrarily small perturbation [MC89]. The power of the GPG condition comes from the following important property (see, e.g., [KAG23, Gup24, WXY23]).

Proposition 1 (Corollary 4.1 in [KAG23])**.** *Suppose that $\mathcal{G}$ satisfies the GPG condition with ϵ defined as (3), and let (z^*, s^*) be the unique non-degenerate optimal solution of* $\mathrm{SPP}(\lambda)$*. Then, for every $\lambda' \in \mathbb{R}^n_{\geq 0}$ with $\|\lambda - \lambda'\|_1 \leq \epsilon$, there exists an optimal basic feasible solution to* $\mathrm{SPP}(\lambda')$ *with the same basic activities, i.e., non-zero components, as (z^*, s^*).*

Given a matching network $\mathcal{G}$ that satisfies the GPG condition with ϵ defined as (3), let (z^*, s^*) be the unique non-degenerate optimal solution of $\mathrm{SPP}(\lambda)$. Define $\mathcal{M}_+ := \{m \in \mathcal{M} \mid z^*_m > 0\}$ and $\mathcal{M}_0 := \mathcal{M} \setminus \mathcal{M}_+$ as the set of active matches and the set of redundant matches, respectively. Also, define $\mathcal{A}_+ := \{j \in \mathcal{A} \mid s^*_j > 0\}$ and $\mathcal{A}_0 := \mathcal{A} \setminus \mathcal{A}_+$ as the set of under-demanded queues and the set of over-demanded queues, respectively. Note that

$$\epsilon = \min_{m\in\mathcal{M}_+} z^*_m \wedge \min_{j\in\mathcal{A}_+} s^*_j > 0.$$

Note that if there exist multiple connected components in the graph $(\mathcal{A}, \mathcal{M})$, we can analyze the policy on each connected component separately, since the actions on one connected component do not affect the performance on another one. As a result, we assume that the graph $(\mathcal{A}, \mathcal{M})$ only consists of one connected component without loss of generality.

As an important consequence of the GPG condition, which is also commonly used by prior work (see, e.g., [KAG23, Lemma 5.1] and [Gup24, Lemma 1]), bounding the all-time regret of a policy can be boiled down to analyzing the total length of the over-demanded queues, provided that the policy is restricted to active matches.

Lemma 1. *Suppose that $\mathcal{G}$ satisfies the GPG condition, and let (z^*, s^*) be the unique non-degenerate optimal solution of* $\mathrm{SPP}(\lambda)$*. Suppose that the following conditions hold under a policy* Π*:*

1. *only matches in $\mathcal{M}_+$ are used, and*
2. *$\sum_{i \in \mathcal{A}_0} \mathbb{E}[Q_i(t)] \le B$ for every $t > 0$, where $B > 0$ does not depend on t.*

Then, Π *achieves constant regret at all times, and* $\mathrm{Regret}(\Pi, T) \le r_{\max} n B$*, where*

$$r_{\max} := \max_{m \in \mathcal{M}_+} r_M \,.$$

Proof. Recall that the constraints of $\mathrm{SPP}(\lambda)$ can be written as

$$\begin{bmatrix} M & I \end{bmatrix} \begin{bmatrix} z \\ s \end{bmatrix} = \lambda,$$

where $I \in \mathbb{R}^{n \times n}$ is the identity matrix. Let $M_B \in \mathbb{R}^{n \times n}$ be the matrix obtained by selecting the columns of $[M \ I]$ corresponding to basic variables of $\mathrm{SPP}(\lambda)$, which implies that M_B is full-rank. Hence, $(z^*_{\mathcal{M}_+}, s^*_{\mathcal{A}_+})^T = M_B^{-1}\lambda$. For each vector $v \in \mathbb{R}^n$, we use $(M_B^{-1} v)_{\mathcal{M}_+}$ to denote the first $|\mathcal{M}_+|$ components of $M_B^{-1} v$. Fix $t > 0$. Note that

$$\mathcal{R}^*(t) \le t \cdot r_{\mathcal{M}_+}^T z^* = t \cdot r_{\mathcal{M}_+}^T \left(M_B^{-1} \lambda \right)_{\mathcal{M}_+}$$

and

$$\mathcal{R}^{\Pi}(t) = \mathbb{E}\left[r^T D(t) \right] = \mathbb{E}\left[r_{\mathcal{M}_+}^T \left(M_B^{-1} (A(t) - Q(t)) \right)_{\mathcal{M}_+} \right] = \mathbb{E}\left[r_{\mathcal{M}_+}^T \left(M_B^{-1} (t\lambda - Q(t)) \right)_{\mathcal{M}_+} \right].$$

Therefore, the regret of Π after time t is

$$\mathcal{R}^*(t) - \mathcal{R}^{\Pi}(t) \le \mathbb{E}\left[r_{\mathcal{M}_+}^T \left(M_B^{-1} Q(t) \right)_{\mathcal{M}_+} \right] \le r_{\max} \cdot \left\| M_B^{-1} \right\|_\infty \cdot \mathbb{E}\left[\|Q(t)\|_1 \right] \le r_{\max} B \cdot \left\| M_B^{-1} \right\|_\infty .$$

By [KAG23, Theorem 4.1], each entry of M_B^{-1} lies between $[-1, 1]$, which implies $\left\| M_B^{-1} \right\|_\infty \le n$, concluding the proof. ∎

In view of Lemma 1, to achieve constant regret at all times, it suffices to restrict the policies to only use active matches and control the lengths of over-demanded queues. We note that it is a common strategy to ignore redundant matches [KAG23, KAG24, Gup24] with few exceptions [WXY23]. Hence, without loss of generality, we assume $\mathcal{M} = \mathcal{M}_+$. Moreover, we allow policies to discard available agents in under-demanded queues, i.e., queues in $\mathcal{A}_+$, at the end of each period, where discarding agents can be equivalently viewed as setting these agents aside and never matching them.

2.3 Greedy policies

Greedy policies form an important family of policies extensively studied by the dynamic matching literature [KAG23, MP17], which perform matches whenever possible and discard available agents in under-demanded queues at the end of each period.

Definition 2 (Greedy policy)**.** A policy is greedy if

1. a match is performed whenever at least one match becomes available, and
2. all available agents in under-demanded queues $\mathcal{A}_+$, if not matched after arrival, are discarded at the end of each period.

Note that we did not specify in Definition 2 which match to perform when there are multiple available matches, and the tie-breaking rule is allowed to be policy-specific. Next, we list several important properties possessed by greedy policies, which can be directly deduced from the definition:

(P1) At most one match can be performed at each period.

(P2) If a match is performed at time t, then it must contain the agent that arrives at time t.

(P3) For every $t > 0$, $Q_i(t) \cdot Q_j(t) = 0$ for all $i, j \in \mathcal{A}$ such that $m(i, j) \in \mathcal{M}$, and $Q_i(t) = 0$ for every $i \in \mathcal{A}_+$.

We say that a state is *valid* if it satisfies (P3), and valid states are states that can be potentially reached by greedy policies starting from an empty state.

Next, we introduce state-independent greedy policies, whose matching decisions depend solely on agent availability across different types - specifically, whether queues are empty or not - rather than requiring knowledge of the complete state information about exact queue-lengths.

Definition 3 (State-independent greedy policy)**.** A greedy policy is state-independent if its matching decision at time t only depends on the type of the agent arrived at time t, and the availability of all types of agents, i.e., $(\mathbf{1}_{\{Q_i(t-1)>0\}})_{i\in\mathcal{A}}$.

In this paper, we only consider greedy policies that are time-homogeneous Markovian, i.e., the matching rule only depends on the current state and does not change over time, and we will implicitly assume this from now on. Given a greedy policy Π and a valid state $q \in \mathbb{Z}^n_{\geq 0}$, since Π is time-homogeneous Markovian, we can define $x^\Pi_m(q)$ as the probability that each match $m \in \mathcal{M}$ is performed at the current period when this period starts with state q. Moreover, let $x^\Pi_m(q, i)$ denote the probability that each match $m \in \mathcal{M}$ is performed at the current period when this period starts with state q and the agent arriving at this period is of type $i \in \mathcal{A}$.

We then define the consistency property of greedy policies, which asserts that when starting from different initial states with the same arrival process, the difference between the resulting states does not grow over time.

Definition 4 (Consistency)**.** We say that a greedy policy is consistent if for all valid initial states $Q(0)$ and $Q'(0)$, for every possible arrival at time 1, there exists a coupling P between $Q(1)$ and $Q'(1)$ such that

$$\mathbb{E}_{(Q(1),Q'(1))\sim P}\left[\left\|Q(1) - Q'(1)\right\|_1\right] \leq \left\|Q(0) - Q'(0)\right\|_1,$$

where the randomness of $Q(1)$ and $Q'(1)$ comes from the randomness used by the policy.

As a consequence of a greedy policy being consistent, every bound for the expected total queue-length under the stationary distribution can be seamlessly translated into a bound for the expected total queue-length at all times, as stated by the following lemma.

Lemma 2. *Let Π be a consistent greedy policy. Suppose that the Markov chain $(Q(t))_{t\geq 0}$ is ergodic, and let π be its stationary distribution. If $\mathbb{E}_\pi[\|Q(0)\|_1] \leq B$, then $\mathbb{E}[\|Q(t)\|_1] \leq 2B$ for every $t \geq 0$.*

Proof. Recall that $(Q(t))_{t\geq 0}$ are the states that result from Π starting from $Q(0) = \mathbf{0}$. Let $(Q'(t))_{t\geq 0}$ be the states resulting from Π when we start from $Q'(0) \sim \pi$. Fix $t \geq 0$. Since Π is consistent, by induction, there exists a coupling P between $Q(t)$ and $Q'(t)$ such that

$$\mathbb{E}_{(Q(t),Q'(t))\sim P}\left[\left\|Q(t) - Q'(t)\right\|_1\right] \leq \mathbb{E}\left[\left\|Q(0) - Q'(0)\right\|_1\right].$$

Hence,

$$\begin{aligned}
\mathbb{E}\left[\|Q(t)\|_1\right] - \mathbb{E}\left[\left\|Q'(t)\right\|_1\right] &= \mathbb{E}_{(Q(t),Q'(t))\sim P}\left[\|Q(t)\|_1 - \left\|Q'(t)\right\|_1\right] \\
&\leq \mathbb{E}_{(Q(t),Q'(t))\sim P}\left[\left\|Q(t) - Q'(t)\right\|_1\right] \\
&\leq \mathbb{E}\left[\left\|Q(0) - Q'(0)\right\|_1\right] = \mathbb{E}\left[\left\|Q'(0)\right\|_1\right].
\end{aligned}$$

Therefore,

$$\mathbb{E}\left[\|Q(t)\|_1\right] \leq \mathbb{E}\left[\left\|Q'(t)\right\|_1\right] + \mathbb{E}\left[\left\|Q'(0)\right\|_1\right] = 2\mathbb{E}\left[\left\|Q'(0)\right\|_1\right] \leq 2B,$$

where the equality holds since $Q'(0)$ follows the stationary distribution π. $\square$

In Section 6, we show that several natural greedy policies are consistent via presenting a sufficient and easily-verifiable condition for consistency.

3 Main results

In this section, we formally present our main results by describing the policies and their regret bounds. In Section 3.1, we focus on acyclic graphs and state our improved regret bound for the static priority policy of [KAG23], which is a state-independent policy only requiring availability information of neighboring queues of the arriving agent. In Section 3.2, we turn to general graphs and describe our randomized state-independent policy, which requires global queue availability information yet achieves an optimal regret scaling.

3.1 Static priority policy on acyclic graphs

When the graph $(\mathcal{A}, \mathcal{M})$ forms a tree, [KAG23] propose a static priority policy, which is formally defined below, that achieves constant regret at all times without explicitly upper bounding its regret in terms of the network parameters. In this section, we improve their analysis and provide an explicit upper bound for the regret of this policy.

We first introduce static priority policies, which constitute a simple yet powerful family of state-independent greedy policies widely considered in prior work [KAG23, MP17, ADSW24].

Definition 5 (Static priority policy)**.** Given a strict ordering $\succ$ over $\mathcal{M}$, we say that a match $m \in \mathcal{M}$ has a higher priority than $m' \in \mathcal{M}$ if $m \succ m'$. The *static priority policy with priority order* $\succ$ is a greedy policy that performs available matches according to their priorities, choosing the highest priority match whenever one or more matches are available.

Next, we formally describe the priority order $\succ$ adopted by the static priority policy of [KAG23] for acyclic matching networks, which we denote as **SP**. When $(\mathcal{A}, \mathcal{M})$ forms a tree, by [KAG23, Lemma 3.1], there is precisely one under-demanded queue, i.e., $|\mathcal{A}_+| = 1$, and we denote it as r. We view $(\mathcal{A}, \mathcal{M})$ as a tree rooted at r. Given any directed path starting from r to $i \in \mathcal{A}_0$, for any two matches (edges) $m, m' \in \mathcal{M}$ on this path, we have $m \succ m'$ if and only if m is farther away from r than m'. In other words, for each node $i \in \mathcal{A}_0$, each match connecting i and one of its children is prioritized over the match connecting i and its parent. An illustration of the construction of this priority order is given in Figure 1.

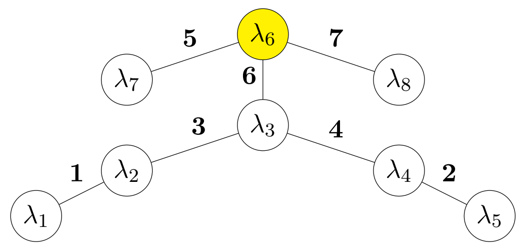

Figure 1: A matching network with root node $r = 6$. One possible priority order is labeled on the matches, where smaller numbers indicate higher priorities. For example, if there is an arriving agent of type 2 and both queues of type 1 and type 3 agents are non-empty, then **SP** performs the match $(1, 2)$ instead of $(2, 3)$ based on the priority order.

SP achieves elegant simplicity in acyclic networks through its natural construction. The key step is identifying the unique under-demanded node by solving the static planning problem SPP(λ). Using this node as the root of the tree, priorities are assigned systematically from root to leaves, with matches further from the root receiving higher priority. This construction aligns with the following intuition: nodes further from the under-demanded root are more demanded, thus warranting higher priority for their matches.

The stability property of this policy is notable – under the GPG condition, even when arrival rates vary (as long as they stay within an ϵ-neighborhood of the original rates), the optimal matching structure and performance guarantees of the **SP** policy remain valid. Following this specific priority structure is crucial, as [KAG23, Example 6.2] demonstrates that even small deviations from this order can fail to achieve constant regret at all times.

Proposition 2 (Theorem 3.2 in [KAG23])**.** *Suppose that $\mathcal{G}$ satisfies the GPG condition, and the graph $(\mathcal{A}, \mathcal{M})$ forms a tree. Then,* **SP** *is state-independent and achieves constant regret at all times.*

While [KAG23] establishes that the regret of **SP** remains bounded independent of T, their analysis does not characterize how this regret depends on the network structure or the GPG parameter ϵ. This limitation leaves a critical aspect of the policy's performance uncharacterized, as the GPG parameter ϵ measures the distance of the fluid relaxation from degeneracy and captures both market thickness and operational stability. Such characterization would reveal how the policy performs across different operating environments and market conditions, information that is crucial for practical implementations.

To address this analytical gap, we provide the first explicit regret bound for **SP** in the following theorem, whose proof is presented in Section 4. The bound reveals how the policy's performance depends on both structural parameters (e.g., the tree depth d_r) and the GPG parameter ϵ.

Theorem 1. *Suppose that $\mathcal{G}$ satisfies the GPG condition with ϵ defined as (3), and the graph $(\mathcal{A}, \mathcal{M})$ forms a tree. Then,*

$$\mathrm{Regret}(\mathbf{SP}, T) \leq r_{\max} \cdot \frac{n^2 2^{d_r}}{\epsilon} \left(1 + \frac{1}{\epsilon}\right)^{\lfloor (d_r - 1)/2 \rfloor},$$

where $r_{\max} = \max_{m \in \mathcal{M}_+} r_m$ and d_r is the depth of the tree when rooted at r.

3.2 Randomized state-independent greedy policy

In this section, we design a randomized state-independent greedy policy that achieves constant regret at all times. We formally present this policy in Algorithm 1. Fix $t > 0$, and we describe the behaviors of the policy at time t. Define

$$\mathcal{U}_+(t) := \{i \in \mathcal{A} \mid Q_i(t-1) > 0\} \quad \text{and} \quad \mathcal{U}_0(t) := \mathcal{A} \setminus \mathcal{U}_+(t)$$

as the sets of non-empty and empty queues at the beginning of time t, respectively. Since the policy discards unmatched agents in under-demanded queues $\mathcal{A}_+$ at the end of each period, we have $\mathcal{A}_+ \cap \mathcal{U}_+(t) = \emptyset$. Define a new arrival vector $\widetilde{\lambda}(t)$ such that

$$\widetilde{\lambda}_i(t) = \begin{cases} \lambda_i, & i \in \mathcal{U}_0(t), \\ \lambda_i + \epsilon/n, & i \in \mathcal{U}_+(t), \end{cases}$$

which does not necessarily satisfy $\sum_{i \in \mathcal{A}} \widetilde{\lambda}_i(t) = 1$. Let (z^*, s^*) be the unique non-degenerate optimal solution of $\mathrm{SPP}(\lambda)$,[4] and let ϵ be the GPG parameter of $\mathcal{G}$. Since $\left\|\lambda - \widetilde{\lambda}(t)\right\|_1 \leq \epsilon$, by Proposition 1, there exists an optimal basic feasible solution $(\widetilde{z}(t), \widetilde{s}(t))$ to $\mathrm{SPP}(\widetilde{\lambda}(t))$ with the same basic activities as (z^*, s^*), which implies $(M\widetilde{z}(t))_i = \widetilde{\lambda}_i(t)$ for every $i \in \mathcal{A}_0$.

Suppose an agent of type $j \in \mathcal{A}$ arrives at time t, and j has at least one non-empty neighboring queue, i.e., $\mathcal{U}_+(t) \cap \mathcal{N}(j) \neq \emptyset$; otherwise, there are no matches we can perform. By (P3), we have $Q_j(t-1) = 0$ and hence $\widetilde{\lambda}_j(t) = \lambda_j$. For each non-empty neighboring queue $i \in \mathcal{N}(j) \cap \mathcal{U}_+(t)$ of the arriving agent, we match this agent to an agent of type i with probability

$$\frac{\widetilde{z}_{m(i,j)}(t)}{\sum_{k \in \mathcal{N}(j) \cap \mathcal{U}_+(t)} \widetilde{z}_{m(k,j)}(t)}.$$

Finally, we update $Q(t)$ accordingly. It is easy to verify that the matching rule described above ensures **RG** to be a greedy policy.

The following theorem upper bounds the regret of **RG**, whose proof is presented in Section 5.

Theorem 2. *Suppose that $\mathcal{G}$ satisfies the GPG condition with ϵ defined by (3). Then, Algorithm 1, denoted as **RG**, is state-independent and achieves constant regret at all times. Moreover,*

$$\mathrm{Regret}(\mathbf{RG}, T) \leq \frac{3 r_{\max} n^2}{\epsilon},$$

where $r_{\max} = \max_{m \in \mathcal{M}_+} r_m$.

[4] Note that for any fixed availability configuration of agents, the optimal matching probabilities remain the same. Since there are at most 2^n possible availability configurations with n agent types, we can precompute the optimal solutions for all $O(2^n)$ cases in advance.

Algorithm 1: Randomized Greedy (**RG**)

$Q(0) \leftarrow \mathbf{0}$;
for $t = 1, \dots, T$ **do**
  $\mathcal{U}_+(t) \leftarrow \{i \in \mathcal{A} \mid Q_i(t-1) > 0\}$; $\mathcal{U}_0(t) \leftarrow \mathcal{A} \setminus \mathcal{U}_+(t)$;
  Define $\widetilde{\lambda}(t) \in \mathbb{R}^n$ such that $\widetilde{\lambda}_i(t) = \lambda_i$ for $i \in \mathcal{U}_0(t)$ and $\widetilde{\lambda}_i(t) = \lambda_i + \epsilon/n$ for $i \in \mathcal{U}_+(t)$;
  Let $(\widetilde{z}(t), \widetilde{s}(t))$ be the optimal solution to SPP$(\widetilde{\lambda}(t))$;
  An agent of type $j \in \mathcal{A}$ arrives;
  Match the arriving agent to an agent in each queue $i \in \mathcal{N}(j) \cap \mathcal{U}_+(t)$ with probability proportional to $\widetilde{z}_{m(i,j)}(t)$;
  if *the arriving agent is matched to an agent in queue* i **then**
    $Q_i(t) \leftarrow Q_i(t) - 1$;
  else
    $Q_j(t) \leftarrow Q_j(t) + \mathbf{1}_{\{j \notin \mathcal{A}_+\}}$;
  end
end

4 Analysis of static priority policy on acyclic graphs

In this section, we analyze the regret of the static priority policy on acyclic graphs. We first demonstrate in Section 4.1 our main proof idea by focusing on a simpler example when the matching network is a path. Then, we present the proof of Theorem 1 in Section 4.2.

4.1 Warm-up: regret analysis on paths

As a warm-up, when $(\mathcal{A}, \mathcal{M})$ forms a path with at least four nodes, we show how to bound the expected queue-lengths for the three nodes farthest away from the root. Combining with Lemma 1, this implies a regret upper bound consistent with Theorem 1 when the path consists of at most four nodes. In particular, we will iteratively bound the queue-lengths in a bottom-up manner starting from the leaf node, and our analysis clearly illustrates the necessity of the dependence on network depth in the regret bound achieved by our approach.

Let $\mathcal{A} = \{1, 2, \cdots, n\}$ and $\mathcal{M} = \{1, 2, \cdots, n-1\}$, where $j \in \mathcal{M}$ denote the match of $(j, j+1)$. Here, we consider the case when n is the under-demanded node with $s_n^* > 0$ and $\mathcal{A}_+ = \{n\}$. Then, the root of $(\mathcal{A}, \mathcal{M})$ is node n by construction. We illustrate the constructed matching network in Figure 2.

Figure 2: A path network where $\mathcal{A}_+ = \{n\}$ (indicated with the yellow vertex).

Note that under **SP**, the priority order $\succ$ is unique: for all $k, l \in \mathcal{M}$, $k \succ l$ if and only if $k < l$. We introduce the following family of (artificial) systems $\mathcal{S} = \{\mathcal{S}_i \mid 1 \le i < n\}$ and construct a coupling with our original system such that each system is equipped with the same arrival process as in the original system. In system $\mathcal{S}_i$, at the end of each period, all available agents of type j with $i + 1 \le j \le n$ are removed from the system. In particular, whenever there is an arrival of agent type $i+1$, if there is no agent of type i present in the system, then the arriving agent of type $i+1$ leaves the system unmatched. Under $\mathcal{S}_i$, denote the number of agents of type j in the queue at the end of time t by $\overline{Q}^i_j(t)$ for $1 \le j \le n$. Note that for all $t \ge 0$ and $1 \le j \le i$, we have

$$\overline{Q}^i_j(t) = A_j(t) - \overline{D}^i_{j-1}(t)\mathbf{1}_{\{j>1\}} - \overline{D}^i_j(t)\mathbf{1}_{\{j<n\}}\,, \tag{4}$$

where for any $1 \le j \le n-1$, $\overline{D}^i_j(t)$ denotes the number of match of $(j, j+1)$ performed up to and including time t in $\mathcal{S}_i$, and we set $\overline{D}^i_0(t) = \overline{D}^i_n(t) = 0$. In particular, by the property of $\mathcal{S}_i$, $\overline{Q}^i_j(t) = 0$ for any $j \in \{i+1, \dots, n\}$.

The following proposition establishes a fundamental alternating pattern in queue-length comparisons between each $\mathcal{S}_i$ and the original system under **SP**. Remarkably, this pattern's structure depends on whether i is odd or even, revealing an intrinsic connection between the static priority ordering and queue-length behaviors.

Proposition 3. *Under* **SP***, for any $t \ge 0$ and $1 \le i < n$, if i is odd,*

$$\overline{Q}^i_{2m}(t) \le Q_{2m}(t)\,, \quad \overline{Q}^i_{2m+1}(t) \ge Q_{2m+1}(t)\,, \quad \forall 0 \le 2m \le i-1\,; \tag{5}$$

if i is even,

$$\overline{Q}^i_{2m}(t) \ge Q_{2m}(t), \quad \overline{Q}^i_{2m+1}(t) \le Q_{2m+1}(t)\,, \quad \forall 0 \le 2m \le i\,. \tag{6}$$

The proof is based on an induction argument and is given in Section B.1. Our goal is to use this coupling to characterize queue-lengths under **SP**. We also note that this coupling ensures the processes we are going to analyze to be Markovian; e.g., the process $(Q_1(t))_{t\ge0}$ itself is not a Markov chain, since the transition probabilities depend on the state of queue 2, whereas $(\overline{Q}^1_1(t))_{t\ge0}$ is a Markov chain. In general, by the construction of the artificial systems, $(\overline{Q}^i_{[i]}(t))_{t\ge0}$ is a Markov chain for all $1 \le i < n$.

We then introduce the following lemma that characterizes the optimal solution of the static planning problem SPP(λ).

Lemma 3 (Theorem 4.1 in [KAG23])**.** *For any $1 \le i < n$, $z^*_i + z^*_{i-1}\mathbf{1}_{\{i\ge2\}} = \lambda_i$.*

Following Lemma 3, we get $z^*_1 = \lambda_1$, $z^*_2 = \lambda_2 - \lambda_1$, and $z^*_3 = \lambda_3 - z^*_2 = \lambda_3 - \lambda_2 + \lambda_1$, and note that $z^*_1, z^*_2, z^*_3 \ge \epsilon$, which will be useful to prove the following results. Next, we discuss the intuition behind the construction of our Lyapunov functions. In $\mathcal{S}_1$, we only need to focus on the length of queue 1, and we naturally adopt the quadratic Lyapunov function $\mathcal{L}(t) := (Q_1(t))^2$. When it comes to $\mathcal{S}_2$, to achieve all-time constant regret via Lemma 1, we should have $D_1(t) \approx A_1(t)$ and $D_2(t) \approx A_2(t) - A_1(t)$. This implies that ideally we want both

$$A_1(t) - D_1(t) = Q_1(t) \quad \text{and} \quad A_2(t) - A_1(t) - D_2(t) = Q_2(t) - Q_1(t)$$

to be small. Hence, a natural choice of the Lyapunov function would be $\mathcal{L}(t) := \beta(Q_1(t))^2 + (Q_2(t) - Q_1(t))^2$ for appropriately chosen coefficient $\beta \ge 0$. However, we always have $Q_1(t) \cdot Q_2(t) = 0$ by (P3), implying that we can safely drop the first term in $\mathcal{L}(t)$, giving rise to our final choice of the Lyapunov function for $\mathcal{S}_2$. Similar derivations also lead to our construction of the Lyapunov function for $\mathcal{S}_3$, which will become clear momentarily.

In the following two lemmas, we show that $\mathbb{E}[Q_1(t)]$ and $\mathbb{E}[Q_2(t)]$ can be bounded by $O(\epsilon^{-1})$ at all times respectively.

Lemma 4. $\mathbb{E}[Q_1(t)] \le \epsilon^{-1}$ *for all $t \ge 0$ under* **SP***.*

Proof. Consider the Lyapunov function $\mathcal{L}(t) := (\overline{Q}_1^1(t))^2$. Conditioned on $\mathcal{L}(t) > 0$, when an agent of type 1 arrives, we have $\overline{Q}_1^1(t+1) = \overline{Q}_1^1(t) + 1$; when an agent of type 2 arrives, match 1 is performed under **SP** and $\overline{Q}_1^1(t+1) = \overline{Q}_1^1(t) - 1$. Thus, for all $t \geq 0$, we have

$$\begin{aligned}\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid \overline{Q}^1(t), \mathcal{L}(t) > 0] &= \mathbb{E}[(\overline{Q}_1^1(t+1) + \overline{Q}_1^1(t))(\overline{Q}_1^1(t+1) - \overline{Q}_1^1(t)) \mid \overline{Q}^1(t), \mathcal{L}(t) > 0] \\ &\leq -2\overline{Q}_1^1(t)(\lambda_2 - \lambda_1) + 1,\end{aligned}$$

where $\lambda_2 - \lambda_1 = z_2^* \geq \epsilon$. Per Lemma 14, the Markov chain $(\overline{Q}_1^1(t))_{t\geq 0}$ is ergodic, and we denote its stationary distribution by π. Then by Lemma 12, we have

$$\mathbb{E}_\pi[\overline{Q}_1^1(0)] \leq \frac{1}{2(\lambda_2 - \lambda_1)} \leq \frac{1}{2\epsilon}.$$

Per Lemma 2 and Corollary 1, we have $\mathbb{E}[\overline{Q}_1^1(t)] \leq \epsilon^{-1}$ for all $t \geq 0$. Finally, it follows from Proposition 3 that $\mathbb{E}[Q_1(t)] \leq \epsilon^{-1}$ for all $t \geq 0$, since $Q_1(t) \leq \overline{Q}_1^1(t)$ for all $t \geq 0$. ☐

Lemma 5. *$\mathbb{E}[Q_2(t)] \leq 2\epsilon^{-1}$ for all $t > 0$ under* **SP**.

Proof. Consider the Lyapunov function $\mathcal{L}(t) := (\overline{Q}_2^2(t) - \overline{Q}_1^2(t))^2$. Conditioned on $\mathcal{L}(t) > 0$, we can either have $\overline{Q}_1^2(t) > 0$ or $\overline{Q}_2^2(t) > 0$ by (P3) given that **SP** is a greedy policy.

Claim 1. For all $t \geq 0$ and $i = 1, 2$, we have

$$\mathbb{E}\left[(\mathcal{L}(t+1) - \mathcal{L}(t) \mid \overline{Q}^2(t), \mathcal{L}(t) > 0, \overline{Q}_i^2(t) > 0\right] \leq -2\epsilon\overline{Q}_i^2(t) + 1. \tag{7}$$

The proof of Claim 1 is deferred to Appendix B.2. Using Claim 1, we have

$$\begin{aligned}\mathbb{E}[(\mathcal{L}(t+1) - \mathcal{L}(t) \mid \overline{Q}^2(t), \mathcal{L}(t) > 0] &\leq -(2\epsilon\overline{Q}_1^2(t) + 1) \cdot \mathbf{1}_{\{\overline{Q}_1^2(t)>0\}} - (2\epsilon\overline{Q}_2^2(t) + 1) \cdot \mathbf{1}_{\{\overline{Q}_2^2(t)>0\}} \\ &\leq -2\epsilon|\overline{Q}_2^2(t) - \overline{Q}_1^2(t)| + 1,\end{aligned}$$

where the first inequality follows from Claim 1 and the second inequality follows from the fact that $\overline{Q}_1^2(t) \cdot \overline{Q}_2^2(t) = 0$ for all $t \geq 0$ by (P3). Denote the stationary distribution of the Markov Chain $(\overline{Q}_{[2]}^2(t))_{t\geq 0}$ by π, which is granted by Lemma 14. Per Lemma 12, we have

$$\mathbb{E}_\pi[|\overline{Q}_2^2(0) - \overline{Q}_1^2(0)|] \leq \frac{1}{2\epsilon},$$

Per Lemma 2 and Corollary 1, we have $\mathbb{E}[|\overline{Q}_2^2(t) - \overline{Q}_1^2(t)|] \leq \epsilon^{-1}$ for all $t \geq 0$. Since $\overline{Q}_2^2(t) \geq Q_2(t)$ and $\overline{Q}_1^2(t) \leq Q_1(t)$ for all $t \geq 0$ by Proposition 3, together with Lemma 4, we get $\mathbb{E}[\overline{Q}_1^2(t)] \leq \epsilon^{-1}$ and $\mathbb{E}[Q_2(t)] \leq \mathbb{E}[\overline{Q}_2^2(t)] \leq 2\epsilon^{-1}$ for all $t \geq 0$. ☐

Next, we upper bound $\mathbb{E}[Q_3(t)]$, for which we can no longer achieve the scaling of $O(\epsilon^{-1})$.

Lemma 6. *$\mathbb{E}[Q_3(t)] \leq O(\epsilon^{-2})$ for all $t \geq 0$ under* **SP**.

Proof. Consider the following Lyapunov function

$$\mathcal{L}(t) := \beta_1 \left(\overline{Q}_1^3(t)\right)^2 + \beta_2 \left(\overline{Q}_2^3(t) - \overline{Q}_1^3(t)\right)^2 + \left(\overline{Q}_3^3(t) - \overline{Q}_2^3(t) + \overline{Q}_1^3(t)\right)^2, \tag{8}$$

where we will determine $\beta_1, \beta_2 \in \mathbb{R}_{\geq 0}$ momentarily. Define $\mathcal{B}(t) := \{i \in \mathcal{A} \mid \overline{Q}_i^3(t) > 0\}$ as the set of non-empty queues at the end of time t. Define the following events $\mathcal{E}_1(t) := \{\mathcal{B}(t) = \{1\}\}, \mathcal{E}_2(t) := \{\mathcal{B}(t) = \{2\}\}, \mathcal{E}_3(t) := \{\mathcal{B}(t) = \{3\}\}$, and $\mathcal{E}_4(t) := \{\mathcal{B}(t) = \{1, 3\}\}$. Note that by (P3), the union of these events forms a partition when $\mathcal{L}(t) > 0$. Next, we introduce the following claim on bounding $\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid \overline{Q}^3(t), \mathcal{L}(t) > 0, \mathcal{E}_i(t)]$ for $1 \leq i \leq 4$.

Claim 2. For all $t \geq 0$, we have

$$\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid \overline{Q}^3(t), \mathcal{L}(t) > 0, \mathcal{E}_1(t)]$$
$$\leq -2\left[(\beta_1 + \beta_2)(\lambda_2 - \lambda_1) - (\lambda_3 - \lambda_2 + \lambda_1)\right] \cdot \left|\overline{Q}_3^3(t) - \overline{Q}_2^3(t) + \overline{Q}_1^3(t)\right| + \beta_1 + \beta_2 + 1.$$

Moreover, for all $k = 2, 3, 4$, and $t \geq 0$, we have

$$\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid \overline{Q}^3(t), \mathcal{L}(t) > 0, \mathcal{E}_k(t)] \leq -2\epsilon \left|\overline{Q}_3^3(t) - \overline{Q}_2^3(t) + \overline{Q}_1^3(t)\right| + 3(\beta_1 + \beta_2 + 1).$$

The proof of Claim 2 is deferred to Appendix B.3. Note that per Claim 2, $\mathcal{L}(t)$ has a negative drift under $\mathcal{E}_2(t), \mathcal{E}_3(t), \mathcal{E}_4(t)$ regardless of the choices of β_1 and β_2, but the sign of the drift is unclear under $\mathcal{E}_1(t)$. Let $\delta := (\beta_1 + \beta_2)(\lambda_2 - \lambda_1) - (\lambda_3 - \lambda_2 + \lambda_1)$ be the coefficient in the drift of $\mathcal{L}(t)$ under $\mathcal{E}_1(t)$. In order to ensure that $\mathcal{L}(t)$ has a negative drift under $\mathcal{E}_1(t)$ as well, we will pick β_1 and β_2 to ensure that $\delta > 0$. Therefore, the overall drift is given by

$$\begin{aligned}\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid \overline{Q}^3(t), \mathcal{L}(t) > 0] &\leq -2\min\{\delta, \epsilon\} \left|\overline{Q}_3^3(t) - \overline{Q}_2^3(t) + \overline{Q}_1^3(t)\right| + 3(\beta_1 + \beta_2 + 1) \\ &= -2\min\{\delta, \epsilon\} \left|\overline{Q}_3^3(t) - \overline{Q}_2^3(t) + \overline{Q}_1^3(t)\right| + \frac{3(\delta + \lambda_3)}{\lambda_2 - \lambda_1}.\end{aligned}$$

Since the third condition of Lemma 14 holds by (P3), it follows that the stationary distribution π of the Markov chain $(\overline{Q}_{[3]}^3(t))_{t \geq 0}$ exists by Lemma 14. Moreover, by Lemma 12, and if we choose appropriate β_1 and β_2 such that $\delta = \epsilon$,

$$\mathbb{E}_\pi\left[\left|\overline{Q}_3^3(0) - \overline{Q}_2^3(0) + \overline{Q}_1^3(0)\right|\right] \leq \frac{3(\delta + \lambda_3)}{2(\lambda_2 - \lambda_1)\min\{\delta, \epsilon\}} = \frac{3(\epsilon + \lambda_3)}{2\epsilon(\lambda_2 - \lambda_1)} \leq O(\epsilon^{-2}), \tag{9}$$

where the last inequality holds since $\lambda_3 \leq 1$ and $\lambda_2 - \lambda_1 \geq \epsilon$. To translate the expected queue-length under the steady-state to that at all times, by Lemma 2 and Corollary 1, we have

$$\mathbb{E}\left[\left|\overline{Q}_3^3(t) - \overline{Q}_2^3(t) + \overline{Q}_1^3(t)\right|\right] \leq O(\epsilon^{-2})$$

for all $t \geq 0$. Moreover, by Lemma 5, we have $\mathbb{E}[Q_2(t)] \leq 2\epsilon^{-1}$ for all $t \geq 0$. Therefore, by Proposition 3, we conclude that $\mathbb{E}[Q_3(t)] \leq \mathbb{E}[\overline{Q}_3^3(t)] \leq O(\epsilon^{-2})$ for all $t \geq 0$. □

Remark 1. Note that in the analysis of Lemma 6, we are unable to achieve the optimal scaling of $O(\epsilon^{-1})$ for the expected length of queue 3 in certain cases by using the generalized quadratic Lyapunov function defined in (8). Specifically, when $\lambda_3 = \Omega(1)$ and $\lambda_2 - \lambda_1 = O(\epsilon)$, the last inequality in (9) will be tight. Furthermore, when generalizing our analysis to matching networks with an arbitrary depth, similar situations would repetitively occur as the depth grows, indicating that it is inevitable for the resulting regret to depend on the depth.

4.2 Proof of Theorem 1

Now, we generalize the arguments in Section 4.1 to prove Theorem 1. For each node $i \in \mathcal{A}$, let $\mathcal{C}(i)$ be the set of children of i; denote $\mathcal{T}(i)$ as the set of nodes in the subtree rooted at i (including i), and denote $\mathcal{T}^-(i) := \mathcal{T}(i) \setminus \{i\}$. For all $i, j \in \mathcal{A}$, let $d(i,j)$ be the (unweighted) distance between i and j. For each $i \in \mathcal{A}$, define $d_i := \max_{j \in \mathcal{T}(i)} d(i,j)$ as the depth of the subtree rooted at i. Let $\mathcal{A}^- := \{i \in \mathcal{A} \mid d_i > 0\}$ denote the set of non-leaf nodes.

Given the optimal solution (z^*, s^*) of SPP(λ), for each $i \in \mathcal{T}^-(r)$, define $w_i := z_m^*$ with m being the match connecting i and its parent; define $w_r := s_r^* > 0$. By (3), $\epsilon \leq w_i \leq 1$ for every $i \in \mathcal{A}$.

For each $i \in \mathcal{A}$, denote $\mathcal{P}(i) := \{j \in \mathcal{A} \mid i \in \mathcal{T}^-(j) \text{ with } d(j,i) \equiv 0 \pmod 2\}$ as the set of ancestors of i whose depth has the same parity with i. We then recursively set α_i, which will be the coefficient of our Lyapunov function, as

$$\alpha_i := 1 + \frac{1}{w_i} \sum_{j \in \mathcal{P}(i)} \alpha_j (\lambda_j - w_j) , \quad \forall i \in \mathcal{A}^- . \tag{10}$$

Note that for every $i \in \{r\} \cup \mathcal{C}(r)$, we have $\mathcal{P}(i) = \emptyset$ and then $\alpha_i = 1$. Next, we define the following generalized quadratic Lyapunov function

$$\mathcal{L}(t) := \sum_{i \in \mathcal{A}^-} \alpha_i \left(f_i(Q(t))^+ \right)^2 , \quad \forall t \geq 0 \tag{11}$$

where for every $i \in \mathcal{A}^-$ and $v \in \mathbb{R}^n$,

$$f_i(v) := \sum_{j \in \mathcal{T}^-(i)} (-1)^{d(i,j)+1} v_j . \tag{12}$$

Notably, our Lyapunov function is the same as the one used in [KAG23] if we do not take the positive parts. For their Lyapunov function, it is difficult to keep track of the coefficients $\{\alpha_i\}_{i \in \mathcal{A}^-}$, as mentioned in [KAG23]. Instead, we will show that this modification leads to simplified analysis, allowing us to derive an explicit upper bound for the regret.

The following lemma upper bounds $\{\alpha_i\}_{i \in \mathcal{A}^-}$.

Lemma 7. *For every $i \in \mathcal{A}^-$, $\alpha_i \leq (1 + \epsilon^{-1})^{\lfloor d(r,i)/2 \rfloor}$.*

Fix $t \geq 0$, and we aim to upper bound $\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid Q(t)]$, where the expectation is taken over the randomness of the arrival at time $t+1$. Let $x := x^{\mathbf{SP}}(Q(t))$ denote the probabilities that matches in $\mathcal{M}$ are performed by **SP** at time $t+1$. Let $\mathcal{E}_1 := \{i \in \mathcal{A}^- \mid f_i(Q(t)) > 0\}$ be the set of nodes with a strictly positive $f_i(Q(t))$. The next lemma simplifies the Lyapunov drift that we hope to upper bound.

Lemma 8. *It holds that*

$$\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid Q(t)] \leq 2 \sum_{i \in \mathcal{E}_1} \alpha_i \cdot f_i(Q(t)) \cdot f_i(\lambda - Mx) + n \left(1 + \frac{1}{\epsilon}\right)^{\lfloor (d_r - 1)/2 \rfloor} . \tag{13}$$

Now, our goal is to upper bound the RHS of (13). Define $\mathcal{E}_2 := \{i \in \mathcal{A}^- \mid \left\| Q_{\mathcal{C}(i)}(t) \right\|_1 > 0\}$ as the set of nodes i such that at least one of its child nodes has a non-empty queue under $Q(t)$. Observe that $f_i(Q(t)) > 0$ for every $i \in \mathcal{E}_1$, and the following lemma establishes that $f_i(\lambda - Mx) < 0$ for all $i \in \mathcal{E}_2$, and $f_i(\lambda - Mx) \leq \lambda_i - w_i$ for every $i \in \mathcal{E}_1 \setminus \mathcal{E}_2$. This ensures that the sum $\sum_{i \in \mathcal{E}_1 \cap \mathcal{E}_2} \alpha_i \cdot f_i(Q(t)) \cdot f_i(\lambda - Mx)$ is negative, which partially contributes to establishing the negative Lyapunov drift.

Lemma 9. *$f_i(\lambda - Mx) = -w_i$ for every $i \in \mathcal{E}_2$, and $f_i(\lambda - Mx) \leq \lambda_i - w_i$ for every $i \in \mathcal{E}_1 \setminus \mathcal{E}_2$.*

To further ensure the Lyapunov drift being negative, it suffices to further upper bound $\sum_{i \in \mathcal{E}_1 \setminus \mathcal{E}_2} \alpha_i \cdot f_i(Q(t)) \cdot f_i(\lambda - Mx)$, which leads to upper bounding $f_i(Q(t))$ for $i \in \mathcal{E}_1 \setminus \mathcal{E}_2$. For every $i \in \mathcal{A}^- \setminus \mathcal{E}_2$, let $\mathcal{H}(i)$ be the set of nodes $j \in \mathcal{T}^-(i) \cap \mathcal{E}_2$ that satisfy the following conditions:

1. $d(i,j) \equiv 0 \pmod 2$, and

2. for every $k \in \mathcal{P}(j) \cap \mathcal{T}^-(i)$ (node on the path between i and j (excluding i and j) with $d(i,k) \equiv 0 \pmod 2$), $k \notin \mathcal{E}_2$.

To visualize the construction of $\mathcal{H}(i)$, one can imagine walking from i down to the leaf nodes in the subtree rooted at i with step size 2. If we encounter a node in $\mathcal{E}_2$, then we add the current node into $\mathcal{H}(i)$ and stop; otherwise, we continue walking. One can also see by this description that $\mathcal{T}^-(j) \cap \mathcal{T}^-(k) = \emptyset$ for all $j,k \in \mathcal{H}(i)$ with $j \neq k$, i.e., there is no overlapping between any two rooted subtrees with roots $j,k \in \mathcal{H}(i)$.

The following lemma states that for every $i \in \mathcal{A}^- \setminus \mathcal{E}_2$, we can upper bound $f_i(Q(t))$ by the sum of $f_j(Q(t))$ for all $j \in \mathcal{H}(i)$.

Lemma 10. *For every $i \in \mathcal{A}^- \setminus \mathcal{E}_2$,*

$$f_i(Q(t)) \le \sum_{j \in \mathcal{H}(i)} f_j(Q(t)).$$

Now, we put the above three lemmas together and upper bound the first term in the RHS of (13), establishing a negative drift for the Lyapunov function. Note that

$$\begin{aligned}
\sum_{i \in \mathcal{E}_1} \alpha_i \cdot f_i(Q(t)) \cdot f_i(\lambda - Mx) &= \sum_{i \in \mathcal{E}_1 \cap \mathcal{E}_2} \alpha_i \cdot f_i(Q(t)) \cdot f_i(\lambda - Mx) + \sum_{i \in \mathcal{E}_1 \setminus \mathcal{E}_2} \alpha_i \cdot f_i(Q(t)) \cdot f_i(\lambda - Mx) \\
&\le \sum_{i \in \mathcal{E}_1 \cap \mathcal{E}_2} \alpha_i \cdot f_i(Q(t)) \cdot (-w_i) + \sum_{i \in \mathcal{E}_1 \setminus \mathcal{E}_2} \alpha_i \cdot f_i(Q(t)) \cdot (\lambda_i - w_i) \\
&\le \sum_{i \in \mathcal{E}_1 \cap \mathcal{E}_2} \alpha_i \cdot f_i(Q(t)) \cdot (-w_i) + \sum_{i \in \mathcal{E}_1 \setminus \mathcal{E}_2} \alpha_i (\lambda_i - w_i) \sum_{j \in \mathcal{H}(i)} f_j(Q(t)) \\
&= \sum_{i \in \mathcal{E}_2} f_i(Q(t)) \left(-\mathbf{1}_{\{i \in \mathcal{E}_1\}} \cdot \alpha_i w_i + \sum_{j \in \mathcal{E}_1 \setminus \mathcal{E}_2} \mathbf{1}_{\{i \in \mathcal{H}(j)\}} \cdot \alpha_j (\lambda_j - w_j) \right),
\end{aligned}$$

where the first inequality holds by Lemma 9, the second inequality holds by Lemma 10, and the last equality holds since $\mathcal{H}(i) \subseteq \mathcal{E}_2$ for every $i \in \mathcal{E}_1 \setminus \mathcal{E}_2 \subseteq \mathcal{A}^- \setminus \mathcal{E}_2$. To upper bound the above quantity, we consider each $i \in \mathcal{E}_2$ separately. On one hand, for every $i \in \mathcal{E}_2 \setminus \mathcal{E}_1$, since $f_i(Q(t)) \le 0$ by the definition of $\mathcal{E}_1$, we have

$$\begin{aligned}
&f_i(Q(t)) \left(-\mathbf{1}_{\{i \in \mathcal{E}_1\}} \cdot \alpha_i w_i + \sum_{j \in \mathcal{E}_1 \setminus \mathcal{E}_2} \mathbf{1}_{\{i \in \mathcal{H}(j)\}} \cdot \alpha_j (\lambda_j - w_j) \right) \\
&= f_i(Q(t)) \left(\sum_{j \in \mathcal{E}_1 \setminus \mathcal{E}_2} \mathbf{1}_{\{i \in \mathcal{H}(j)\}} \cdot \alpha_j (\lambda_j - w_j) \right) \le 0.
\end{aligned}$$

On the other hand, for every $i \in \mathcal{E}_1 \cap \mathcal{E}_2$, which implies $f_i(Q(t)) > 0$, it holds that

$$\begin{aligned}
-\mathbf{1}_{\{i \in \mathcal{E}_1\}} \cdot \alpha_i w_i + \sum_{j \in \mathcal{E}_1 \setminus \mathcal{E}_2} \mathbf{1}_{\{i \in \mathcal{H}(j)\}} \cdot \alpha_j (\lambda_j - w_j) &\le -\alpha_i w_i + \sum_{j \in \mathcal{P}(i) \cap (\mathcal{E}_1 \setminus \mathcal{E}_2)} \alpha_j (\lambda_j - w_j) \\
&\le -\alpha_i w_i + \sum_{j \in \mathcal{P}(i)} \alpha_j (\lambda_j - w_j) = -w_i,
\end{aligned}$$

where the first inequality holds since $i \in \mathcal{H}(j)$ implies $j \in \mathcal{P}(i)$, and the last equality holds by (10). Combining the above three displayed equations, we conclude that

$$\sum_{i \in \mathcal{E}_1} \alpha_i \cdot f_i(Q(t)) \cdot f_i(\lambda - Mx) \leq - \sum_{i \in \mathcal{E}_1 \cap \mathcal{E}_2} f_i(Q(t)) \cdot w_i. \tag{14}$$

By Lemma 8, (14), and the fact that $\epsilon \leq w_i \leq 1$ for every $i \in \mathcal{A}$, we get

$$\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid Q(t)] \leq -2\epsilon \sum_{i \in \mathcal{E}_1 \cap \mathcal{E}_2} f_i(Q(t)) + n \left(1 + \frac{1}{\epsilon}\right)^{\lfloor (d_r - 1)/2 \rfloor}. \tag{15}$$

Then, we translate the above drift of the Lyapunov function in terms of f to a drift in terms of the total queue-length via the following lemma.

Lemma 11. *For every $q \in \mathbb{Z}^n_{\geq 0}$,*

$$\frac{1}{2^{d_r}} \sum_{i \in \mathcal{A}_0} q_i \leq \sum_{i \in \mathcal{E}_1 \cap \mathcal{E}_2} f_i(q).$$

Combining (15) and Lemma 11, we obtain

$$\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid Q(t)] \leq -\frac{\epsilon}{2^{d_r - 1}} \left\| Q(t) \right\|_1 + n \left(1 + \frac{1}{\epsilon}\right)^{\lfloor (d_r - 1)/2 \rfloor},$$

which implies that the Markov chain $(Q(t))_{t \geq 0}$ is ergodic by Lemma 14. Denote the stationary distribution of this Markov chain to be π. By applying Lemma 12 with $f(Q(t)) = \frac{\epsilon}{2^{d_r - 1}} \left\| Q(t) \right\|_1$, $g(Q(t)) = \mathcal{L}(t)$, and $c = n(1 + \epsilon^{-1})^{\lfloor (d_r - 1)/2 \rfloor}$, we get

$$\mathbb{E}_\pi \left[\left\| Q(0) \right\|_1 \right] \leq \frac{n 2^{d_r - 1}}{\epsilon} \left(1 + \frac{1}{\epsilon}\right)^{\lfloor (d_r - 1)/2 \rfloor}.$$

By Corollary 1 and Lemma 2, we obtain

$$\mathbb{E} \left[\left\| Q(t) \right\| \right] \leq \frac{n 2^{d_r}}{\epsilon} \left(1 + \frac{1}{\epsilon}\right)^{\lfloor (d_r - 1)/2 \rfloor}$$

for every $t \geq 0$. Finally, the regret bound of **SP** follows from Lemma 1. All the omitted proofs in this subsection can be found in Appendix B.

5 Analysis of randomized state-independent greedy policy

We prove Theorem 2 in this section. The state-independence of **RG** comes from its description. Then, we analyze the Markov chain $(Q(t))_{t \geq 0}$ by using the following quadratic Lyapunov function for $t \geq 0$:

$$\mathcal{L}(t) := \left\| Q(t) \right\|_2^2 = \sum_{i \in \mathcal{A}_0} (Q_i(t))^2,$$

where the second equality holds since $Q_i(t) = 0$ for every $i \in \mathcal{A}_+$. Fix $t \geq 0$, and we upper bound $\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid Q(t)]$, where the expectation is taken over the randomness of the arrival process and **RG**. Let $x := x^{\mathbf{RG}}(Q(t))$ denote probabilities that matches in $\mathcal{M}$ are performed by **RG** at time $t+1$. By standard calculation (see, e.g., [KAG23, Proposition 5.1]),

$$\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid Q(t)] \le 2\langle Q(t), \lambda - Mx\rangle + 1.$$

For each match $m \in \mathcal{M}$ that contains types i and j such that $i \in \mathcal{U}_+(t)$ and $j \in \mathcal{U}_0(t)$, m is performed at time $t+1$ if and only if an agent of type j arrives and **RG** decides to match this agent with an agent in queue i, which implies that

$$\begin{aligned} x_m &= \lambda_j \cdot \frac{\widetilde{z}_m(t)}{\sum_{k \in \mathcal{N}(j) \cap \mathcal{U}_+(t)} \widetilde{z}_{m(k,j)}(t)} \ge \lambda_j \cdot \frac{\widetilde{z}_m(t)}{\sum_{k \in \mathcal{N}(j)} \widetilde{z}_{m(k,j)}(t)} \\ &= \lambda_j \cdot \frac{\widetilde{z}_m(t)}{(M\widetilde{z}(t))_j} = \lambda_j \cdot \frac{\widetilde{z}_m(t)}{\widetilde{\lambda}_j(t)} = \lambda_j \cdot \frac{\widetilde{z}_m(t)}{\lambda_j} = \widetilde{z}_m(t). \end{aligned}$$

Hence, for each queue $i \in \mathcal{U}_+(t)$,

$$(Mx)_i = \sum_{j \in \mathcal{N}(i)} x_{m(i,j)} \ge \sum_{j \in \mathcal{N}(i)} \widetilde{z}_{m(i,j)}(t) = (M\widetilde{z}(t))_i = \widetilde{\lambda}_i(t).$$

As a result,

$$\begin{aligned} \mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid Q(t)] &\le 2\langle Q(t), \lambda - Mx\rangle + 1 \\ &\le 2 \sum_{i \in \mathcal{U}_+(t)} Q_i(t)(\lambda_i - \widetilde{\lambda}_i(t)) + 1 = -2\frac{\epsilon}{n} \left\|Q(t)\right\|_1 + 1, \end{aligned} \tag{16}$$

where the both equalities hold since $Q_i(t) = 0$ for every $i \in \mathcal{U}_0(t)$.

Next, we combine the negative drift (16) with Lemma 13 to upper bound the expected total queue-length. Fix $t \ge 0$. By (P1) and (P2), the variation of $\|Q(t)\|_2$ is bounded by

$$\|Q(t+1)\|_2 - \|Q(t)\|_2 \le \|Q(t+1) - Q(t)\|_2 = \|\Delta A(t+1) - M\Delta D(t+1)\|_2 \le 1.$$

Then, we bound the expected decrease of $\|Q(t)\|_2$, i.e., $\mathbb{E}[\|Q(t+1)\|_2 - \|Q(t)\|_2 \mid Q(t)]$. When $\|Q(t)\|_2 \ge n/\epsilon$, we get

$$\begin{aligned} \mathbb{E}\left[\|Q(t+1)\|_2 - \|Q(t)\|_2 \mid Q(t)\right] &\le \mathbb{E}\left[\left.\frac{\|Q(t+1)\|_2^2 - \|Q(t)\|_2^2}{2\left\|Q(t)\right\|_2} \,\right|\, Q(t)\right] \\ &\le \mathbb{E}\left[\left.\frac{-2\frac{\epsilon}{n}\left\|Q(t)\right\|_1 + 1}{2\left\|Q(t)\right\|_2} \,\right|\, Q(t)\right] \le -\frac{\epsilon}{2n}, \end{aligned}$$

where the first inequality holds by $x - y \le (x^2 - y^2)/(2y)$ for $x \in \mathbb{R}$ and $y > 0$ [CJK+06, Lemma 3.6], the second inequality holds by (16), and the last inequality holds since $\|Q(t)\|_1 \ge \|Q(t)\|_2$.

Now, we apply Lemma 13 on $\Psi(t) = \|Q(t)\|_2$ with $K = 1$, $\eta = \epsilon/(2n)$, and $B = n/\epsilon$, which gives

$$\mathbb{E}\left[\|Q(t)\|_2\right] = \mathbb{E}\left[\|\Psi(t)\|\right] \le 2 + \frac{n}{\epsilon} + \frac{1 - \frac{\epsilon}{2n}}{\frac{\epsilon}{n}} = \frac{3}{2} + \frac{2n}{\epsilon} \le \frac{3n}{\epsilon},$$

where the last inequality holds since $\epsilon \le 1$ and $n \ge 2$. Finally, the all-time constant regret of **RG** follows from Lemma 1.

6 Consistent greedy policy

In this section, we investigate the sufficient conditions for a greedy policy to be consistent (Definition 4). Recall that [MP17, Lemma 4] establish the consistency for static priority policies when no agents are discarded. Also, [KAG23, Lemma 5.3] claim that all (deterministic) greedy policies are consistent, which is not accurate as we will demonstrate. We present a general sufficient condition, which is easy to verify, for a (randomized) greedy policy to be consistent. Then, we show that all static priority policies and the longest-queue policy of [KAG23], which discard available agents in under-demanded queues, meet this condition. All the omitted proofs in this section can be found in Appendix C.

We start with an example to illustrate that certain deterministic greedy policies do not satisfy consistency, refuting [KAG23, Lemma 5.3].

Example 1. Assume that n is sufficiently large and the graph $(\mathcal{A}, \mathcal{M})$ forms a path such that for every $i \in [n-1]$, there is a match in $\mathcal{M}$ containing types i and $i+1$. To describe a greedy policy, it suffices to specify, under an arbitrary valid state q, which match to perform when an agent of type $i > 1$ arrives with queues $i-1$ and $i+1$ being non-empty at the same time. Under this situation, our greedy policy Π works as follows: if $q_1 = 0$, then Π performs the match $m(i-1, i)$; otherwise, Π performs the match $m(i, i+1)$. In other words, when queue 1 is empty, Π acts like the static priority policy with priority order $\succ$ such that $m(i-1,i) \succ m(i, i+1)$ for every $i > 1$; otherwise, Π acts like the static priority policy with priority order $\succ'$ such that $m(i, i+1) \succ' m(i-1, i)$ for every $i > 1$.

Next, we give the initial states and specify the arrival process. Let $Q(0) = \mathbf{0}$ and $Q'(0) = (1, 0, \ldots, 0)$. The first four arriving agents are of types $3, 5, 4, 6$, respectively. During the first four periods, Π will perform $m(3,4)$ and $m(5,6)$ under Q, and perform $m(4,5)$ under Q'. Hence, $\|Q(4)\|_1 = 0$ and $\|Q'(4)\|_1 = 3$, which implies

$$\left\|Q(4) - Q'(4)\right\|_1 = 3 > \left\|Q(0) - Q'(0)\right\|_1 .$$

Therefore, Π is not consistent.

We remark that in the above example, one can repeat the arrival pattern to make the distance between $Q(t)$ and $Q'(t)$ arbitrarily large as t increases. That is, the next four arriving agents are respectively of types $8, 10, 9, 11$, and so on.

Next, we present our sufficient condition for a greedy policy to be consistent as follows.

Proposition 4. *Let Π be a greedy policy such that for all valid states q and q', for every possible type i of arrival at time 1 such that $q_{\mathcal{N}(i)} \neq \mathbf{0}$ and $q'_{\mathcal{N}(i)} \neq \mathbf{0}$, defining $x := x^{\Pi}(q, i)$ and $x' := x^{\Pi}(q', i)$, we have*

$$\sum_{j \in \mathcal{N}_=} \left(x_{m(i,j)} - x'_{m(i,j)} \right)^+ + \sum_{j \in \mathcal{N}_<} x_{m(i,j)} \leq \sum_{j \in \mathcal{N}_<} x'_{m(i,j)}, \tag{17}$$

where $\mathcal{N}_= := \{j \in \mathcal{N}(i) \mid q_j = q'_j\}$ and $\mathcal{N}_< := \{j \in \mathcal{N}(i) \mid q_j < q'_j\}$. Then, Π is consistent.

Here, we give some high-level explanations of the condition (17). Recall that the definition of consistency (Definition 4) asks for a coupling between $Q(1)$ and $Q'(1)$, which is equivalent to coupling $m \sim x^{\Pi}(Q(0), i)$ and $m' \sim x^{\Pi}(Q'(0), i)$, that satisfies certain conditions. Then, in view of the structure of the desired couplings (see Lemma 16) and the optimal way to construct such couplings, (17) appears naturally as a sufficient condition to achieve this.

We note that similar results also appear in prior work, e.g., [SYY24, Theorem 6]. Since agents are divided into offline and online sides in their setting, whereas all agents in our model are online, their result cannot be directly applied to ours.

Recall that the longest-queue policy, which is a greedy policy that achieves constant regret at all times [KAG23], adopts the following matching rule: When the arriving agent has multiple non-empty neighboring queues, select the one with the largest length to match to, with ties broken in favor of the queue with the smallest index.[5] Now, we apply Proposition 4 to show that all static priority policies and the longest-queue policy are consistent.

Corollary 1. *All static priority policies and the longest-queue policy are consistent.*

7 Conclusion

In this paper, we investigate the performance of state-independent greedy policies in two-way matching networks. We first give an explicit regret bound for the static priority policy on acyclic matching networks, which grows with the depth of the network, and show that this is inevitable using our Lyapunov approach. We leave the problem of pinpointing the tight regret scaling of the static priority policy as an open problem, and it would also be interesting to find a static priority policy for cyclic matching networks that achieves constant regret at all times.

Next, by presenting a randomized state-independent greedy policy with an optimal regret scaling of $O(\epsilon^{-1})$, we reveal that state-independence does not necessarily lead to a compromise in performance and does not require the matching network to be acyclic, which we believe will inspire future research on state-independent policies. However, compared to static priority policies, it requires the information on availability of each agent type rather than merely the availability of agent types neighboring to the arriving agent. Hence, it would be promising to provide a state-independent greedy policy with both the optimal regret scaling and less needed information.

Our findings reveal an interesting informational trade-off between state-dependent and state-independent policies. While both the longest-queue greedy policy and the randomized greedy policy achieve the optimal regret scaling, the former requires complete queue-length information of neighbors of the arriving agent's queue, while the latter requires global queue availability information (whether a queue is empty or not). In contrast, while we establish an explicit constant regret bound for the static priority policy which does not necessarily achieve the optimal regret scaling, the policy only requires the local queue availability information. In particular, in settings where querying and communicating the exact queue-length information is more costly than simply querying availability information of queues, the state-independent policies we study in this work might be preferred to state-dependent policies without compromising a strong performance guarantee.

For multi-way matching networks where each match might contain more than two agent types, both the primal-dual policy of [WXY23], which greedily schedules matches, and the sum-of-square policy of [Gup24], which greedily commits agents to matches, can achieve constant regret at all times. Nevertheless, both policies do not fit into our greedy framework since they may not necessarily perform matches even with the presence of sufficient available agents. Hence, another intriguing future direction is to devise state-independent policies for multi-way matching networks that achieve constant regret at all times. Notably, there exists an example with a multi-way match-

[5] In [KAG23], ties are broken arbitrarily when there are multiple longest queues. However, the longest-queue policy with an arbitrary tie-breaking rule is not necessarily consistent. To illustrate, under two different states, when the sets of longest neighboring queues under two states are identical, if the policy breaks ties differently under two states, then the consistency condition would be violated.

ing network such that the regret of any greedy policy grows linearly with T [KAG24, Example 3.2], indicating that the concept of state-independence also needs to be correspondingly generalized.

While our work establishes strong theoretical guarantees for state-independent policies, an important direction for future research is to formalize their practical benefits of transparency and resistance against manipulation. Combining our results with the literature on strategic behaviors in queueing systems (e.g., [ERIZ25]) could help demonstrate how state-independent policies provide robustness against strategic agents under precise game-theoretic assumptions.

References

[AAGK17] Ross Anderson, Itai Ashlagi, David Gamarnik, and Yash Kanoria. Efficient dynamic barter exchange. *Operations Research*, 65(6):1446–1459, 2017.

[AAS25] Alireza AmaniHamedani, Ali Aouad, and Amin Saberi. Adaptive approximation schemes for matching queues, 2025.

[ABB+18] Itai Ashlagi, Adam Bingaman, Maximilien Burq, Vahideh Manshadi, David Gamarnik, Cathi Murphey, Alvin E Roth, Marc L Melcher, and Michael A Rees. Effect of match-run frequencies on the number of transplants and waiting times in kidney exchange. *American Journal of Transplantation*, 18(5):1177–1186, 2018.

[ABJM19] Itai Ashlagi, Maximilien Burq, Patrick Jaillet, and Vahideh Manshadi. On matching and thickness in heterogeneous dynamic markets. *Operations Research*, 67(4):927–949, 2019.

[ADSW24] Angelos Aveklouris, Levi DeValve, Maximiliano Stock, and Amy Ward. Matching impatient and heterogeneous demand and supply. *Operations Research*, 2024.

[AG19] Alessandro Arlotto and Itai Gurvich. Uniformly bounded regret in the multisecretary problem. *Stochastic Systems*, 9(3):231–260, 2019.

[ALG20] Mohammad Akbarpour, Shengwu Li, and Shayan Oveis Gharan. Thickness and information in dynamic matching markets. *Journal of Political Economy*, 128(3):783–815, 2020.

[AS22] Ali Aouad and Ömer Sarıtaç. Dynamic stochastic matching under limited time. *Operations Research*, 70(4):2349–2383, 2022.

[BBP24] Santiago R Balseiro, Omar Besbes, and Dana Pizarro. Survey of dynamic resource-constrained reward collection problems: Unified model and analysis. *Operations Research*, 72(5):2168–2189, 2024.

[BFP23] Eric Balkanski, Yuri Faenza, and Noémie Périvier. The power of greedy for online minimum cost matching on the line. In *Proceedings of the 24th ACM Conference on Economics and Computation*, pages 185–205, 2023.

[BM15] Ana Buŷić and Sean Meyn. Approximate optimality with bounded regret in dynamic matching models. *ACM SIGMETRICS Performance Evaluation Review*, 43(2):75–77, 2015.

[BRS+22] Jose H Blanchet, Martin I Reiman, Virag Shah, Lawrence M Wein, and Linjia Wu. Asymptotically optimal control of a centralized dynamic matching market with general utilities. *Operations Research*, 70(6):3355–3370, 2022.

[BW20] Pornpawee Bumpensanti and He Wang. A re-solving heuristic with uniformly bounded loss for network revenue management. *Management Science*, 66(7):2993–3009, 2020.

[Chr24] Houston Chronicle. Memorial hermann stops kidney transplants, days after announcing liver transplant stoppage, May 2024.

[CHS24] Ziyun Chen, Zhiyi Huang, and Enze Sun. Stochastic online correlated selection. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 2275–2294. IEEE, 2024.

[CILB+20] Natalie Collina, Nicole Immorlica, Kevin Leyton-Brown, Brendan Lucier, and Neil Newman. Dynamic weighted matching with heterogeneous arrival and departure rates. In *Web and Internet Economics: 16th International Conference, WINE 2020, Beijing, China, December 7–11, 2020, Proceedings 16*, pages 17–30. Springer, 2020.

[CJK+06] János Csirik, David S. Johnson, Claire Kenyon, James B. Orlin, Peter W. Shor, and Richard R. Weber. On the sum-of-squares algorithm for bin packing. *J. ACM*, 53(1):1–65, 2006.

[CLY24] Guanting Chen, Xiaocheng Li, and Yinyu Ye. Technical note - an improved analysis of lp-based control for revenue management. *Oper. Res.*, 72(3):1124–1138, 2024.

[ERIZ25] Arturo Estrada Rodriguez, Rouba Ibrahim, and Dongyuan Zhan. On customer (dis-) honesty in unobservable queues: The role of lying aversion. *Management Science*, 71(1):844–860, 2025.

[ET23] Myungeun Eom and Alejandro Toriello. Batching and greedy policies: How good are they in dynamic matching? *Available at SSRN 4357375*, 2023.

[Gup24] Varun Gupta. Greedy algorithm for multiway matching with bounded regret. *Operations Research*, 72(3):1139–1155, 2024.

[GW15] Itai Gurvich and Amy Ward. On the dynamic control of matching queues. *Stochastic Systems*, 4(2):479–523, 2015.

[GZ08] Peter W Glynn and Assaf Zeevi. Bounding stationary expectations of markov processes. In *Markov processes and related topics: a Festschrift for Thomas G. Kurtz*, volume 4, pages 195–215. Institute of Mathematical Statistics, 2008.

[Jas15] Stefanus Jasin. Performance of an lp-based control for revenue management with unknown demand parameters. *Operations Research*, 63(4):909–915, 2015.

[JK12] Stefanus Jasin and Sunil Kumar. A re-solving heuristic with bounded revenue loss for network revenue management with customer choice. *Mathematics of Operations Research*, 37(2):313–345, 2012.

[JMRSL22] Matthieu Jonckheere, Pascal Moyal, Claudia Ramírez, and Nahuel Soprano-Loto. Generalized max-weight policies in stochastic matching. *Stochastic Systems*, 2022.

[JMZ22] Jiashuo Jiang, Will Ma, and Jiawei Zhang. Degeneracy is ok: Logarithmic regret for network revenue management with indiscrete distributions. *arXiv preprint arXiv:2210.07996*, 2022.

[KAG23] Süleyman Kerimov, Itai Ashlagi, and Itai Gurvich. On the optimality of greedy policies in dynamic matching. *Operations Research*, 2023.

[KAG24] Süleyman Kerimov, Itai Ashlagi, and Itai Gurvich. Dynamic matching: Characterizing and achieving constant regret. *Management Science*, 70(5):2799–2822, 2024.

[Kan21] Yash Kanoria. Dynamic spatial matching. *arXiv preprint arXiv:2105.07329*, 2021.

[KSSW22] Kristen Kessel, Ali Shameli, Amin Saberi, and David Wajc. The stationary prophet inequality problem. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 243–244, 2022.

[MC89] Nimrod Megiddo and Ramaswamy Chandrasekaran. On the ε-perturbation method for avoiding degeneracy. *Operations Research Letters*, 8(6):305–308, 1989.

[MM16] Jean Mairesse and Pascal Moyal. Stability of the stochastic matching model. *Journal of Applied Probability*, 53(4):1064–1077, 2016.

[MP17] Pascal Moyal and Ohad Perry. On the instability of matching queues. 2017.

[MS87] Ovil Mangasarian and Tzong Shiau. Lipschitz Continuity of Solutions of Linear Inequalities, Programs and Complementarity Problems. *SIAM Journal on Control and Optimization*, 25(3):583–595, 1987.

[NS19] Mohammadreza Nazari and Alexander L Stolyar. Reward maximization in general dynamic matching systems. *Queueing Systems*, 91:143–170, 2019.

[ÖW20] Erhun Özkan and Amy R Ward. Dynamic matching for real-time ride sharing. *Stochastic Systems*, 10(1):29–70, 2020.

[RM21] Youssef Rahme and Pascal Moyal. A stochastic matching model on hypergraphs. *Advances in Applied Probability*, 53(4):951–980, 2021.

[Rob03] Philippe Robert. Stochastic networks and queues. *(No Title)*, 2003.

[SYY24] Amin Saberi, Mingwei Yang, and Sophie H. Yu. Stochastic online metric matching: Adversarial is no harder than stochastic. *CoRR*, abs/2407.14785, 2024.

[Tim25] The New York Times. Organ transplant system 'in chaos' as waiting lists are ignored, May 2025.

[TVR98] Kalyan Talluri and Garrett Van Ryzin. An analysis of bid-price controls for network revenue management. *Management science*, 44(11-part-1):1577–1593, 1998.

[VB19] Alberto Vera and Siddhartha Banerjee. The bayesian prophet: A low-regret framework for online decision making. *ACM SIGMETRICS Performance Evaluation Review*, 47(1):81–82, 2019.

[VBG21] Alberto Vera, Siddhartha Banerjee, and Itai Gurvich. Online allocation and pricing: Constant regret via bellman inequalities. *Operations Research*, 69(3):821–840, 2021.

[WXY23] Yehua Wei, Jiaming Xu, and Sophie H Yu. Constant regret primal-dual policy for multi-way dynamic matching. *Available at SSRN 4357216*, 2023.

A Preliminaries on Lyapunov function analysis

Lyapunov function analysis is one of the prevalent approaches to bound the expected value of some function with respect to the stationary state of a Markov chain. Our analysis heavily relies on the following general tool.

Lemma 12 (Corollary 4 in [GZ08]). *Let $X = (X(t))_{t\geq 0}$ be a discrete-time $\mathcal{S}$-valued Markov chain with transition kernel P, and suppose $f : \mathcal{S} \to \mathbb{R}_{\geq 0}$. If there exists a function $g : \mathcal{S} \to \mathbb{R}_{\geq 0}$ and a constant c for which*

$$\int_S P(x, dy)g(y) - g(x) \leq -f(x) + c \text{ for every } x \in \mathcal{S},$$

then

$$\int_S \pi(dx)f(x) \leq c$$

for every stationary distribution π of X.

The function g in the above lemma is usually referred to as a *Lyapunov function.* We will also apply the following lemma, which enables us to directly bound the expectation of the Lyapunov function at all times instead of only under the stationary distribution.

Lemma 13 (Lemma 5 in [WXY23]). *Let $\Psi(t)$ be an $\{\mathcal{F}_t\}$-adapted stochastic process satisfying:*

- *Bounded variation: $|\Psi(t + 1) - \Psi(t)| \leq K$;*
- *Expected decrease: $\mathbb{E}[\Psi(t + 1) - \Psi(t) \mid \mathcal{F}_t] \leq -\eta$, when $\Psi(t) \geq B$;*
- $\Psi(0) \leq K + B$.

Then, we have

$$\mathbb{E}[\Psi(t)] \leq K\left(1 + \left\lceil \frac{B}{K} \right\rceil\right) + K\left(\frac{K - \eta}{2\eta}\right).$$

The following lemma introduces a generic way of establishing ergodicity of Markov chains.

Lemma 14 ([Rob03], Corollary 8.7). *Let $(M_t)_{t\geq 0}$ be a discrete-time, homogeneous, irreducible and aperiodic Markov chain with values in a countable state space $\mathcal{X}$. If there exist a function $f : \mathcal{X} \to \mathbb{R}_+$ and constants $K, \eta > 0$ such that*

(i) $\mathbb{E}[f(M_1) - f(M_0) \mid f(M_0) > K] \leq -\eta$,

(ii) $\mathbb{E}[f(M_1) \mid f(M_0) \leq K] < \infty$, *and*

(iii) the set $\{x \in \mathcal{X} \mid f(x) \leq K\}$ is finite,

then the Markov chain $(M_t)_{t\geq 0}$ is ergodic.

B Postponed proofs in Section 4

B.1 Proof of Proposition 3

Assume that i is odd. Let us denote by $(t_k)_{k=1}^{\infty}$ the sequence of times when there is an agent of type $j \le i+1$ arrives at $\mathcal{S}_i$. For any $0 \le t < t_0$, we have $\overline{Q}_j^i(t) = Q_j(t) = 0$ for any $1 \le j \le i+1$, and (5) follows directly. Since there is no arrivals of any type $1 \le j \le i$ at time t for any $t_k < t < t_{k+1}$, then $\overline{Q}_j^i(t) = \overline{Q}_j^i(t_k)$ and $Q_j(t) = Q_j(t_k)$. Also, by the property of $\mathcal{S}_i$, we have $\overline{Q}_{i+1}^i(t) = 0$, and hence $\overline{Q}_{i+1}^i(t) = 0 \le Q_{i+1}(t)$. Hence, it suffices to prove by induction that for any $k \ge 0$, we have (5) holds for $t = t_k$ and $1 \le j \le i$.

We then prove by induction. Suppose $k = 1$. If the arriving agent at t_1 is of type l, where $1 \le l \le i$, then clearly $\overline{Q}_j^i(t_1) = Q_j(t_1) = 0$ for all $1 \le j \le i+1, j \ne l$, and $\overline{Q}_l^i(t_1) = Q_l(t_1) = 1$. Then, (5) holds. If the arriving agent at t_1 is of type $i+1$, then $\overline{Q}_j^i(t_1) = Q_j(t_1) = 0$ for all $1 \le j \le i$, and $Q_{i+1}(t_1) = 1$. By the property of $\mathcal{S}_i$, $\overline{Q}_{i+1}^i(t_1) = 0$. Then, (5) holds. For $k \ge 1$, suppose that (5) holds for $t = t_k$ and consider t_{k+1}. We have the following cases:

- **Case 1:** Assume that the arriving agent at t_{k+1} is of type 1.
  - **Case 1.1:** If the arriving agent is not matched to an agent of type 2 in $\mathcal{S}_i$ at time t_{k+1}, this implies $\overline{Q}_2^i(t_{k+1}) = \overline{Q}_2^i(t_{k+1}-1) = 0$. Then, $Q_2^i(t_{k+1}) \ge 0 = \overline{Q}_2^i(t_{k+1})$. Also, by inductive hypothesis, $Q_1(t_k) \le \overline{Q}_1^i(t_k)$, and then $Q_1(t_{k+1}) \le Q_1(t_k) + 1 \le \overline{Q}_1^i(t_k) + 1 = \overline{Q}_1^i(t_{k+1})$ given that the arriving agent of type 1 is not matched in $\mathcal{S}_i$ at time t_{k+1}.
  - **Case 1.2:** If the arriving agent is matched to agent of type 2 in $\mathcal{S}_i$ at time t_{k+1}, then we must have $Q_2(t_k) \ge \overline{Q}_2^i(t_k) \ge 1$, which implies that the arriving agent is also matched to an agent type of 2 in the original system. Then, by inductive hypothesis, $Q_2^i(t_{k+1}) = Q_2^i(t_k) - 1 \ge \overline{Q}_2^i(t_k) - 1 = \overline{Q}_2^i(t_{k+1})$, and $Q_1^i(t_{k+1}) = Q_1^i(t_k) \le \overline{Q}_1^i(t_k) = \overline{Q}_1^i(t_{k+1})$.

  Hence, we get $Q_2^i(t_{k+1}) \ge \overline{Q}_2^i(t_{k+1})$, and $Q_1^i(t_{k+1}) \le \overline{Q}_1^i(t_{k+1})$. For any $3 \le j \le i$, $Q_j(t_{k+1}) = \overline{Q}_j^i(t_{k+1})$. By inductive hypothesis, (5) holds for $t = t_{k+1}$.

- **Case 2:** Assume that the arriving agent at t_{k+1} is of type l, where $2 \le l \le i-1$, and l is odd.
  - **Case 2.1:** Suppose the arriving agent is not matched in the original system at time t_{k+1}. Then, $Q_{l-1}(t_{k+1}-1) = Q_{l-1}(t_k) = 0$ and $Q_{l+1}(t_{k+1}-1) = Q_{l+1}(t_k) = 0$. By inductive hypothesis, $Q_{l-1}(t_k) \ge \overline{Q}_{l-1}^i(t_k)$ and $Q_{l+1}(t_k) \ge \overline{Q}_{l+1}^i(t_k)$, and then $\overline{Q}_{l-1}^i(t_{k+1}-1) = \overline{Q}_{l-1}^i(t_k) = 0$ and $\overline{Q}_{l+1}^i(t_{k+1}-1) = \overline{Q}_{l+1}^i(t_k) = 0$. It implies that the arriving agent is also not matched in $\mathcal{S}_i$ at time t_{k+1}. Hence, we have $\overline{Q}_{l+1}^i(t_{k+1}) = \overline{Q}_{l+1}^i(t_{k+1}-1) = 0 \le Q_{l+1}(t_{k+1})$, and $\overline{Q}_{l-1}^i(t_{k+1}) = \overline{Q}_{l-1}^i(t_{k+1}-1) = 0 \le Q_{l-1}(t_{k+1})$. By inductive hypothesis, we have $\overline{Q}_l^i(t_k) \ge Q_l(t_k)$, and then $\overline{Q}_l^i(t_{k+1}) = \overline{Q}_l^i(t_{k+1}-1)+1 = \overline{Q}_l^i(t_k)+1 \ge Q_l(t_k)+1 = Q_l(t_{k+1}-1)+1 = Q_l(t_{k+1})$.
  - **Case 2.2:** Suppose the arriving agent is matched to an agent of type $l-1$ in the original system at time t_{k+1}.
    - **Case 2.2.1:** If the arriving agent is not matched in $\mathcal{S}_i$ at time t_{k+1}, then $\overline{Q}_{l-1}^i(t_{k+1}) = \overline{Q}_{l-1}^i(t_{k+1}-1) = \overline{Q}_{l-1}^i(t_k) = 0$ and $\overline{Q}_{l+1}^i(t_{k+1}) = \overline{Q}_{l+1}^i(t_{k+1}-1) = \overline{Q}_{l+1}^i(t_{k+1}) = 0$. It follows that $\overline{Q}_{l+1}^i(t_{k+1}) = 0 \le Q_{l+1}(t_{k+1})$, and $\overline{Q}_{l-1}^i(t_{k+1}) = 0 \le Q_{l-1}(t_{k+1})$. By inductive hypothesis, we have $\overline{Q}_l^i(t_k) \ge Q_l(t_k)$, and then $\overline{Q}_l^i(t_{k+1}) = \overline{Q}_l^i(t_{k+1}-1)+1 = \overline{Q}_l^i(t_k) + 1 \ge Q_l(t_k) = Q_l(t_{k+1}-1) = Q_l(t_{k+1})$, in view of that the arriving agent is not matched in $\mathcal{S}_i$ but matched in the original system at time t_{k+1}.

* **Case 2.2.2:** If the arriving agent is matched to an agent of type $l-1$ in $\mathcal{S}_i$ at time t_{k+1}, we have $\overline{Q}^i_{l-1}(t_{k+1}) = \overline{Q}^i_{l-1}(t_{k+1}-1) - 1 = \overline{Q}^i_{l-1}(t_k) - 1 \le Q_{l-1}(t_k) - 1 = Q_{l-1}(t_{k+1}-1) - 1 = Q_{l-1}(t_{k+1})$, where the inequality holds by inductive hypothesis ($\overline{Q}^i_{l-1}(t_k) \le Q_{l-1}(t_k)$). Similarly, by inductive hypothesis, we can show that $\overline{Q}^i_l(t_{k+1}) \ge Q_l(t_{k+1})$, and $\overline{Q}^i_{l+1}(t_{k+1}) \le Q_{l+1}(t_{k+1})$.
* **Case 2.2.3:** If the arriving agent is matched to an agent of type $l+1$ in $\mathcal{S}_i$ at time t_{k+1}, then we can show that $\overline{Q}_{l-1}(t_{k+1}) \le Q_{l-1}(t_{k+1})$ $\overline{Q}^i_l(t_{k+1}) \ge Q_l(t_{k+1})$, and $\overline{Q}^i_{l+1}(t_{k+1}) \le Q_{l+1}(t_{k+1})$. The proof is analogous to the proof for Case 2.2.2 and hence omitted here.

– **Case 2.3:** Suppose the arriving agent is matched to an agent of type $l+1$ in the original system at time t_{k+1}. By the property of static priority policy, we must have $Q_{l-1}(t_{k+1}-1) = Q_{l-1}(t_k) = 0$. By inductive hypothesis, we have $\overline{Q}_{l-1}(t_k) \le Q_{l-1}(t_k)$, and then $\overline{Q}_{l-1}(t_{k+1}-1) = \overline{Q}_{l-1}(t_k) = 0$. Hence, the arriving agent cannot match to an agent of type $l-1$ in $\mathcal{S}_i$ at time t_{k+1}, and $Q_{l-1}(t_{k+1}) \ge \overline{Q}_{l-1}(t_{k+1}) = \overline{Q}_{l-1}(t_{k+1}-1) = 0$.

* **Case 2.3.1:** If the arriving agent is not matched in $\mathcal{S}_i$ at time t_{k+1}, then $\overline{Q}^i_{l+1}(t_{k+1}) = \overline{Q}^i_{l+1}(t_{k+1}-1) = \overline{Q}^i_{l+1}(t_k) = 0$. By inductive hypothesis, we have $Q_{l+1}(t_k) \ge \overline{Q}^i_{l+1}(t_k)$, and then $Q_{l+1}(t_{k+1}) \ge 0 = \overline{Q}^i_{l+1}(t_{k+1})$. By inductive hypothesis, we have $\overline{Q}^i_l(t_k) \ge Q_l(t_k)$, and then $\overline{Q}^i_l(t_{k+1}) = \overline{Q}^i_l(t_{k+1}-1) + 1 = \overline{Q}^i_l(t_k) + 1 \ge Q_l(t_k) = Q_l(t_{k+1}-1) = Q_l(t_{k+1})$, in view that the arriving agent is not matched in $\mathcal{S}_i$ but matched in the original system at time t_{k+1}.
* **Case 2.3.2:** If the arriving agent is matched to an agent of type $l+1$ in $\mathcal{S}_i$ at time t_{k+1}, we have $\overline{Q}^i_{l+1}(t_{k+1}) = \overline{Q}^i_{l+1}(t_{k+1}-1) - 1 = \overline{Q}^i_{l+1}(t_k) - 1 \le Q_{l+1}(t_k) - 1 = Q_{l+1}(t_{k+1}-1) - 1 = Q_{l+1}(t_{k+1})$, where the inequality holds by inductive hypothesis ($\overline{Q}^i_{l+1}(t_k) \le Q_{l+1}(t_k)$). Similarly, by inductive hypothesis, we can show that $\overline{Q}^i_l(t_{k+1}) \ge Q_l(t_{k+1})$.

Hence, we get $\overline{Q}^i_{l-1}(t_{k+1}) \le Q_{l-1}(t_{k+1})$, $\overline{Q}^i_l(t_{k+1}) \ge Q_l(t_{k+1})$, and $\overline{Q}^i_{l+1}(t_{k+1}) \le Q_{l+1}(t_{k+1})$. For any $j \ne \ell-1, \ell, \ell$ where $1 \le j \le i$, $Q_j(t_k+1) = \overline{Q}^i_j(t_k+1)$. By inductive hypothesis, (5) holds for $t = t_{k+1}$.

- **Case 3:** Assume that the arriving agent at t_{k+1} is of type l, where $2 \le l \le i-1$, and l is even. Then, (5) holds. The proof is analogous to the proof for Case 2 and is therefore omitted.

- **Case 4:** Assume that the arriving agent at t_{k+1} is of type i. By the property of $\mathcal{S}_i$, we have $\overline{Q}^i_{i+1}(t) = 0$ for any t. Hence, we have $\overline{Q}^i_{i+1}(t_{k+1}) = 0 \le Q_{i+1}(t_{k+1})$. The rest of the proof for Case 4 is analogous to the proof of Case 2 and is therefore omitted.

- **Case 5:** Assume that the arriving agent at t_{k+1} is of type $i+1$. By the property of $\mathcal{S}_i$, we have $\overline{Q}^i_{i+1}(t) = 0$ for any t. Hence, we have $\overline{Q}^i_{i+1}(t_{k+1}) = 0 \le Q_{i+1}(t_{k+1})$.

  – **Case 5.1:** If the arriving agent is not matched in the original system, then $Q_i(t_{k+1}-1) = 0$, and hence $Q_i(t_{k+1}) = 0$. It follows that $Q_i(t_{k+1}) = 0 \le \overline{Q}^i_i(t_{k+1}) = 0$.

  – **Case 5.2:** If the arriving agent is matched to an agent of type i in the original system, $Q_i(t_{k+1}-1) = Q_i(t_k) \ge 1$. By inductive hypothesis, $\overline{Q}^i_i(t_k) \ge Q_i(t_k) \ge 1$, and then $\overline{Q}^i_i(t_{k+1}-1) \ge 1$. Hence, the arriving agent is also matched to an agent of type i in $\mathcal{S}_i$, and then $\overline{Q}^i_i(t_{k+1}) = \overline{Q}^i_i(t_k) - 1 \ge Q_i(t_k) - 1 = Q_i(t_{k+1})$, given that the arriving agent is matched to an agent of type i in both the original system and $\mathcal{S}_i$.

– **Case 5.3:** If the arriving agent is matched to $i+2$ in the original system. Then we must have $Q_i(t_{k+1}-1) = Q_i(t_k) = 0$, and then $Q_i(t_{k+1}) = 0$. Hence, we have $\overline{Q}_i^i(t_{k+1}) \geq Q_i(t_{k+1}) = 0$.

Hence, we get $\overline{Q}_i^i(t_{k+1}) \leq Q_i(t_{k+1})$, and $\overline{Q}_{i+1}^i(t_{k+1}) \geq Q_{i+1}(t_{k+1})$. For any $j \leq i-1$, $Q_j(t_{k+1}) = Q_j(t_k)$ and $\overline{Q}_j^i(t_{k+1}) = \overline{Q}_j^i(t_k)$. By inductive hypothesis, (5) holds for $t = t_{k+1}$.

The proof for i being even is analogous to the proof for i being odd and is therefore omitted.

B.2 Proof of Claim 1

First, assume that $\overline{Q}_1^2(t) > 0$. If an agent of type 1 arrives, then we have

$$(\overline{Q}_2^2(t+1) - \overline{Q}_1^2(t+1)) - (\overline{Q}_2^2(t) - \overline{Q}_1^2(t)) = -1,$$
$$(\overline{Q}_2^2(t+1) - \overline{Q}_1^2(t+1)) + (\overline{Q}_2^2(t) - \overline{Q}_1^2(t)) = -2\overline{Q}_1^2(t) - 1.$$

If an agent of type 2 arrives, then we have

$$(\overline{Q}_2^2(t+1) - \overline{Q}_1^2(t+1)) - (\overline{Q}_2^2(t) - \overline{Q}_1^2(t)) = 1,$$
$$(\overline{Q}_2^2(t+1) - \overline{Q}_1^2(t+1)) + (\overline{Q}_2^2(t) - \overline{Q}_1^2(t)) = -2\overline{Q}_1^2(t) + 1.$$

And, if an agent of type 3 arrives, $\mathcal{L}(t+1) = \mathcal{L}(t)$. Thus, we have for all $t \geq 0$,

$$\begin{aligned}\mathbb{E}\left[(\mathcal{L}(t+1) - \mathcal{L}(t) \mid \overline{Q}^2(t), \mathcal{L}(t) > 0, \overline{Q}_1^2(t) > 0\right] &\leq -2\overline{Q}_1^2(t)(\lambda_2 - \lambda_1) + \lambda_1 + \lambda_2 \\ &\leq -2\epsilon\overline{Q}_1^2(t) + 1,\end{aligned}$$

where we used the fact that $\lambda_2 - \lambda_1 \geq \epsilon$ and $\lambda_1 + \lambda_2 \leq 1$.

Now assume that $\overline{Q}_2^2(t) > 0$. If an agent of type 1 or 3 arrives, then we have

$$(\overline{Q}_2^2(t+1) - \overline{Q}_1^2(t+1)) - (\overline{Q}_2^2(t) - \overline{Q}_1^2(t)) = -1,$$
$$(\overline{Q}_2^2(t+1) - \overline{Q}_1^2(t+1)) + (\overline{Q}_2^2(t) - \overline{Q}_1^2(t)) = 2\overline{Q}_2^2(t) - 1,$$

and if an agent of type 2 arrives, then we have

$$(\overline{Q}_2^2(t+1) - \overline{Q}_1^2(t+1)) - (\overline{Q}_2^2(t) - \overline{Q}_1^2(t)) = 1,$$
$$(\overline{Q}_2^2(t+1) - \overline{Q}_1^2(t+1)) + (\overline{Q}_2^2(t) - \overline{Q}_1^2(t)) = 2\overline{Q}_2^2(t) + 1.$$

Thus, we have for all $t \geq 0$,

$$\begin{aligned}\mathbb{E}\left[(\mathcal{L}(t+1) - \mathcal{L}(t) \mid \overline{Q}^2(t), \mathcal{L}(t) > 0, \overline{Q}_2^2(t) > 0\right] &\leq -2\overline{Q}_2^2(t)(\lambda_3 - \lambda_2 + \lambda_1) + \lambda_1 + \lambda_2 + \lambda_3 \\ &\leq -2\epsilon\overline{Q}_2^2(t) + 1\,,\end{aligned}$$

where we used the fact that $\lambda_3 - \lambda_2 + \lambda_1 \geq \epsilon$ and $\lambda_1 + \lambda_2 + \lambda_3 \leq 1$.

B.3 Proof of Claim 2

Under $\mathcal{E}_1(t)$, any arriving agent with types 1 or 3 increases $\overline{Q}_1^3(t)$ or $\overline{Q}_3^3(t)$ by 1, respectively. If the arriving agent is of type 2, $\overline{Q}_1^3(t)$ decreases by 1, and an arriving agent of type 4 does not affect the queue-lengths since $\overline{Q}_3^3(t) = 0$. Thus, we have

$$\begin{aligned}&\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid \overline{Q}^3(t), \mathcal{L}(t) > 0, \mathcal{E}_1(t)] \\ &\leq (-2\beta_1(\lambda_2 - \lambda_1)\overline{Q}_1^3(t) + \beta_1) + (-2\beta_2(\lambda_2 - \lambda_1)\overline{Q}_1^3(t) + \beta_2) + (2(\lambda_3 - \lambda_2 + \lambda_1)\overline{Q}_1^3(t) + 1) \\ &= (-2(\beta_1 + \beta_2)(\lambda_2 - \lambda_1)\overline{Q}_1^3(t) + \beta_1 + \beta_2) + (2(\lambda_3 - \lambda_2 + \lambda_1)\overline{Q}_1^3(t) + 1) \\ &= -2\left[(\beta_1 + \beta_2)(\lambda_2 - \lambda_1) - (\lambda_3 - \lambda_2 + \lambda_1)\right]|\overline{Q}_3^3(t) - \overline{Q}_2^3(t) + \overline{Q}_1^3(t)| + \beta_1 + \beta_2 + 1\,,\end{aligned}$$

where the last equality holds because under $\mathcal{E}_1(t)$, we have $\overline{Q}_3^3(t) = \overline{Q}_2^3(t) = 0$. Under $\mathcal{E}_2(t)$, any arriving agent of type 2 increases $\overline{Q}_2^3(t)$ by 1, and any arriving agent of types 1 or 3 decreases $\overline{Q}_2^3(t)$ by 1, while $\overline{Q}_1^3(t+1) = \overline{Q}_3^3(t+1) = 0$. An arriving agent of type 4 does not affect the queue-lengths. Thus, we have

$$\begin{aligned}\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid \overline{Q}^3(t), \mathcal{L}(t) > 0, \mathcal{E}_2(t)] &\leq -2(\beta_2+1)(\lambda_3 - \lambda_2 + \lambda_1)\overline{Q}_2^3(t) + \beta_2 + 1 \\ &\leq -2(\beta_2+1)\epsilon|\overline{Q}_3^3(t) - \overline{Q}_2^3(t) + \overline{Q}_1^3(t)| + \beta_2 + 1 \\ &\leq -2\epsilon|\overline{Q}_3^3(t) - \overline{Q}_2^3(t) + \overline{Q}_1^3(t)| + \beta_2 + 1.\end{aligned}$$

Under $\mathcal{E}_3(t)$, any arriving agent of type 1 increases $\overline{Q}_1^3(t)$ by 1, any arriving agent of type 2 decreases $\overline{Q}_3^3(t)$ by 1, any arriving agent of type 3 increases $\overline{Q}_3^3(t)$ by 1, and any arriving agent of type 4 decreases $\overline{Q}_3^3(t)$ by 1. Thus, we have

$$\begin{aligned}\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid \overline{Q}^3(t), \mathcal{L}(t) > 0, \mathcal{E}_3(t)] &\leq 2(\beta_1+\beta_2)\lambda_1 - 2(\lambda_4 - \lambda_3 + \lambda_2 - \lambda_1)\overline{Q}_3^3(t) + 1 \\ &\leq -2\epsilon|\overline{Q}_3^3(t) - \overline{Q}_2^3(t) + \overline{Q}_1^3(t)| + 2(\beta_1+\beta_2) + 1.\end{aligned}$$

Finally, under $\mathcal{E}_4(t)$, any arriving agent of types 1 or 3 increases $\overline{Q}_1^3(t)$ or $\overline{Q}_3^3(t)$ by 1, respectively. Any arriving agent of type 2 decreases $\overline{Q}_1^3(t)$ by 1, and any arriving agent of type 4 decreases $\overline{Q}_3^3(t)$ by 1. Thus, we have

$$\begin{aligned}&\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid \overline{Q}^3(t), \mathcal{L}(t) > 0, \mathcal{E}_4(t)] \\ &\leq -(2(\beta_1+1)(\lambda_2 - \lambda_1)\overline{Q}_1^3(t) + \beta_1 + 1) - (2(\beta_2+1)(\lambda_2 - \lambda_1)\overline{Q}_1^3(t) + \beta_2 + 1) \\ &- 2(\lambda_4 - \lambda_3 + \lambda_2 - \lambda_1)(\overline{Q}_1^3(t) + \overline{Q}_3^3(t)) + 1) \\ &\leq -2(\lambda_4 - \lambda_3 + \lambda_2 - \lambda_1)(\overline{Q}_1^3(t) + \overline{Q}_3^3(t)) + \beta_1 + \beta_2 + 3 \\ &\leq -2\epsilon|\overline{Q}_3^3(t) - \overline{Q}_2^3(t) + \overline{Q}_1^3(t)| + \beta_1 + \beta_2 + 3,\end{aligned}$$

concluding the proof.

B.4 Proof of Lemma 7

The upper bound straightforwardly holds for $i \in \{r\} \cup \mathcal{C}(r)$. Assume by induction that the upper bound holds for all $j \in \mathcal{P}(i)$, and we show that it also holds for i. By (10),

$$\begin{aligned}\alpha_i &= 1 + \frac{1}{w_i}\sum_{j\in\mathcal{P}(i)} \alpha_j(\lambda_j - w_j) \leq 1 + \frac{1}{\epsilon}\sum_{j\in\mathcal{P}(i)} \alpha_j \leq 1 + \frac{1}{\epsilon}\sum_{j\in\mathcal{P}(i)} \left(1 + \frac{1}{\epsilon}\right)^{\lfloor d(r,j)/2\rfloor} \\ &= 1 + \frac{1}{\epsilon}\sum_{t=0}^{\lfloor d(r,i)/2\rfloor - 1} \left(1 + \frac{1}{\epsilon}\right)^t = \left(1 + \frac{1}{\epsilon}\right)^{\lfloor d(r,i)/2\rfloor},\end{aligned}$$

where the first inequality holds since $w_i \geq \epsilon$ and $0 \leq w_j \leq \lambda_j \leq 1$ for every $j \in \mathcal{P}(i)$, and the second inequality holds by the inductive hypothesis. Hence, the upper bound also holds for i, concluding the proof.

B.5 Proof of Lemma 8

Recall that $Q(t+1) = Q(t) + \Delta A(t+1) - M\Delta D(t+1)$. By (12),

$$f_i(Q(t+1)) - f_i(Q(t)) = \sum_{j\in\mathcal{T}^-(i)} (-1)^{d(i,j)+1}(\Delta A(t+1) - M\Delta D(t+1))_j.$$

Since D is a greedy policy, in view of (P1) and (P2), there is precisely one non-zero entry in $\Delta A(t+1) - M\Delta D(t+1)$, which must be either 1 or -1. Hence,

$$|f_i(Q(t+1)) - f_i(Q(t))| \leq 1\,. \tag{18}$$

By (11),

$$\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid Q(t)] = \sum_{i \in \mathcal{A}^-} \alpha_i \cdot \mathbb{E}[(f_i(Q(t+1))^+)^2 - (f_i(Q(t))^+)^2 \mid Q(t)].$$

For each $i \in \mathcal{A}^-$ with $f_i(Q(t)) < 0$, we have $f_i(Q(t+1)) \leq f_i(Q(t)) + 1 \leq 0$, and hence $f_i(Q(t))^+ = f_i(Q(t+1))^+ = 0$. Also, for each $i \in \mathcal{A}^-$ with $f_i(Q(t)) = 0$, we have $f_i(Q(t+1)) \leq f_i(Q(t)) + 1 \leq 1$, and hence $(f_i(Q(t+1))^+)^2 - (f_i(Q(t))^+)^2 \leq 1$. Recall that $\mathcal{E}_1 := \{i \in \mathcal{A}^- \mid f_i(Q(t)) > 0\}$ denotes the set of nodes i with a strictly positive $f_i(Q(t))$. As a result,

$$\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid Q(t)] \leq \sum_{i \in \mathcal{E}_1} \alpha_i \cdot \mathbb{E}[(f_i(Q(t+1))^+)^2 - (f_i(Q(t))^+)^2 \mid Q(t)] + \sum_{i \in \mathcal{A}^- \setminus \mathcal{E}_1} \alpha_i\,. \tag{19}$$

Fix $i \in \mathcal{E}_1$, which implies $f_i(Q(t+1)) \geq f_i(Q(t)) - 1 \geq 0$. It holds that

$$\begin{aligned}
(f_i(Q(t+1))^+)^2 - (f_i(Q(t))^+)^2 &= f_i(Q(t+1))^2 - f_i(Q(t))^2 \\
&= (f_i(Q(t+1)) - f_i(Q(t)))^2 - 2f_i(Q(t))\,(f_i(Q(t)) - f_i(Q(t+1))) \\
&\leq 1 + 2f_i(Q(t))\,(f_i(Q(t+1)) - f_i(Q(t))) \\
&= 2f_i(Q(t)) \cdot f_i(\Delta A(t+1) - M\Delta D(t+1)) + 1\,,
\end{aligned}$$

where the inequality holds by (18), and the last equality holds by (12). Then, we get

$$\begin{aligned}
\mathbb{E}[(f_i(Q(t+1))^+)^2 - (f_i(Q(t))^+)^2 \mid Q(t)] &\leq \mathbb{E}[2f_i(Q(t)) \cdot f_i(\Delta A(t+1) - M\Delta D(t+1)) + 1 \mid Q(t)] \\
&= 2f_i(Q(t)) \cdot f_i(\lambda - Mx) + 1.
\end{aligned}$$

Combining the above inequality and (19), we obtain

$$\begin{aligned}
\mathbb{E}[\mathcal{L}(t+1) - \mathcal{L}(t) \mid Q(t)] &\leq \sum_{i \in \mathcal{E}_1} \alpha_i (2f_i(Q(t)) \cdot f_i(\lambda - Mx) + 1) + \sum_{i \in \mathcal{A}^- \setminus \mathcal{E}_1} \alpha_i \\
&\leq 2 \sum_{i \in \mathcal{E}_1} \alpha_i \cdot f_i(Q(t)) \cdot f_i(\lambda - Mx) + n\left(1 + \frac{1}{\epsilon}\right)^{\lfloor (d_r - 1)/2 \rfloor},
\end{aligned}$$

where the second inequality holds by Lemma 7 and the fact that $d(r,i) \leq d_r - 1$ for every $i \in \mathcal{A}^-$.

B.6 Proof of Lemma 9

By (12),

$$\begin{aligned}
f_i(\lambda - Mx) &= \sum_{j \in \mathcal{T}^-(i)} (-1)^{d(i,j)+1} (\lambda - Mx)_j \\
&= \lambda_i - w_i - \sum_{j \in \mathcal{T}^-(i)} (-1)^{d(i,j)+1} (Mx)_j\,,
\end{aligned} \tag{20}$$

where the second equality holds by following lemma:

Lemma 15 (Theorem 4.1 in [KAG23]). *Suppose that $\mathcal{G}$ satisfies the GPG condition, and the graph $(\mathcal{A}, \mathcal{M})$ forms a tree. Then, for every $i \in \mathcal{A}^-$,*

$$w_i = \sum_{j \in \mathcal{T}(i)} (-1)^{d(i,j)} \lambda_j .$$

Hence, to upper bound $f_i(\lambda - Mx)$ given by (20), it suffices to lower bound

$$(\lambda_i - w_i) - f_i(\lambda - Mx) = \sum_{j \in \mathcal{T}^-(i)} (-1)^{d(i,j)+1} (Mx)_j = \mathbb{E}\left[\sum_{j \in \mathcal{T}^-(i)} (-1)^{d(i,j)+1} (M \Delta D(t+1))_j \;\middle|\; Q(t) \right].$$

Conditioning on $Q(t)$, define random variable $W_i := \sum_{j \in \mathcal{T}^-(i)} (-1)^{d(i,j)+1} (M \Delta D(t+1))_j$, whose randomness comes from the arrival at time $t+1$. Depending on the matches performed at time $t+1$, $W_i = 0$ happens in the following three cases:

1. No match is performed, which implies $\Delta D(t+1) = \mathbf{0}$.
2. The performed match (ℓ_1, ℓ_2) satisfies $\ell_1, \ell_2 \in \mathcal{A} \setminus \mathcal{T}^-(i)$, which implies $(M\Delta D(t+1))_j = 0$ for every $j \in \mathcal{T}^-(i)$.
3. The performed match (ℓ_1, ℓ_2) satisfies $\ell_1, \ell_2 \in \mathcal{T}^-(i)$. In this case, $(M\Delta D(t+1))_j = 0$ for every $j \in \mathcal{T}^-(i) \setminus \{\ell_1, \ell_2\}$, and

   $$(-1)^{d(i,\ell_1)+1} (M\Delta D(t+1))_{\ell_1} + (-1)^{d(i,\ell_2)+1} (M\Delta D(t+1))_{\ell_2} = (-1)^{d(i,\ell_1)+1} + (-1)^{d(i,\ell_2)+1} = 0,$$

   where the last equality holds since ℓ_1 and ℓ_2 are adjacent.

As the only remaining case, suppose that the performed match (ℓ_1, ℓ_2) satisfies $\ell_1 = i$ and $\ell_2 \in \mathcal{C}(i)$. In this case, $(M\Delta D(t+1))_j = 0$ for every $j \in \mathcal{T}^-(i) \setminus \{\ell_2\}$ and $(M\Delta D(t+1))_{\ell_2} = 1$, implying that $W_i = 1$. Hence, we always have $W_i \geq 0$, which gives $f_i(\lambda - Mx) = \lambda_i - w_i - \mathbb{E}[W_i] \leq \lambda_i - w_i$. This concludes the second part of Lemma 9. Furthermore, if $i \in \mathcal{E}_2$, by the matching rule of **SP**, $W_i = 1$ holds if and only if an agent of type i arrives at time $t+1$, which happens with probability λ_i. Therefore, $\mathbb{E}[W_i \mid Q(t)] = \lambda_i$ for $i \in \mathcal{E}_2$, concluding the first part of Lemma 9.

B.7 Proof of Lemma 10

Recall that $\mathcal{E}_2 := \{i \in \mathcal{A}^- \mid \left\|Q_{\mathcal{C}(i)}(t)\right\|_1 > 0\}$. Fix $i \in \mathcal{A}^- \setminus \mathcal{E}_2$. Recall that $\mathcal{T}^-(j) \cap \mathcal{T}^-(k) = \emptyset$ for all $j, k \in \mathcal{H}(i)$. Define $\mathcal{U} := \bigcup_{j \in \mathcal{H}(i)} \mathcal{T}^-(j)$. Since $d(i,j) \equiv 0 \pmod 2$ for every $j \in \mathcal{H}(i)$, we have

$$\sum_{k \in \mathcal{U}} (-1)^{d(i,k)+1} Q_k(t) = \sum_{j \in \mathcal{H}(i)} \sum_{k \in \mathcal{T}^-(j)} (-1)^{d(j,k)+1} Q_k(t) = \sum_{j \in \mathcal{H}(i)} f_j(Q(t)).$$

Moreover, by the definition of $\mathcal{H}(i)$, $j \notin \mathcal{E}_2$ for every $j \in \mathcal{T}^-(i) \setminus (\mathcal{U} \cup \mathcal{H}(i))$ such that $d(i,j) \equiv 0 \pmod 2$, which indicates that $Q_j(t) = 0$ for every $j \in \mathcal{T}^-(i) \setminus \mathcal{U}$ such that $d(i,j) \equiv 1 \pmod 2$. Hence,

$$\sum_{j \in \mathcal{T}^-(i) \setminus \mathcal{U}} (-1)^{d(i,j)+1} Q_j(t) = \sum_{j \in \mathcal{T}^-(i) \setminus \mathcal{U}} \mathbf{1}_{\{d(i,j) \equiv 0 \pmod 2\}} \cdot (-1)^{d(i,j)+1} Q_j(t) \leq 0.$$

Combining both of the above displayed equations,

$$\begin{aligned}
f_i(Q(t)) &= \sum_{j\in\mathcal{T}^-(i)} (-1)^{d(i,j)+1} Q_j(t) \\
&= \sum_{j\in\mathcal{U}} (-1)^{d(i,j)+1} Q_j(t) + \sum_{j\in\mathcal{T}^-(i)\setminus\mathcal{U}} (-1)^{d(i,j)+1} Q_j(t) \leq \sum_{j\in\mathcal{H}(i)} f_j(Q(t)),
\end{aligned}$$

concluding the proof.

B.8 Proof of Lemma 11

We will prove a stronger statement that

$$\frac{1}{2^{d_i}} \sum_{j\in\mathcal{T}^-(i)} q_j \leq \sum_{j\in\mathcal{T}(i)\cap\mathcal{E}_1\cap\mathcal{E}_2} f_j(q) \tag{21}$$

for every $i \in \mathcal{A}$, and Lemma 11 follows by setting $i = r$.

For those $i \in \mathcal{A}$ with $d_i = 0$, i.e., i is a leaf node, (21) straightforwardly holds since both sides are equal to 0. Assume by induction that (21) holds for all $i \in \mathcal{A}$ with $d_i < k$ such that $k \in [d_r]$, and we show that (21) holds for all $i \in \mathcal{A}$ with $d_i = k$. Fix $i \in \mathcal{A}$ with $d_i = k$. Observe that

$$\begin{aligned}
\sum_{j\in\mathcal{T}^-(i)\cap\mathcal{E}_1\cap\mathcal{E}_2} f_j(q) &= \sum_{j\in\mathcal{C}(i)} \sum_{k\in\mathcal{T}(j)\cap\mathcal{E}_1\cap\mathcal{E}_2} f_k(q) \\
&\geq \sum_{j\in\mathcal{C}(i)} \frac{1}{2^{d_j}} \sum_{k\in\mathcal{T}^-(j)} q_k \geq \frac{1}{2^{d_i-1}} \sum_{j\in\mathcal{C}(i)} \sum_{k\in\mathcal{T}^-(j)} q_k,
\end{aligned}$$

where the first inequality holds by the inductive hypothesis. Hence,

$$\sum_{j\in\mathcal{T}(i)\cap\mathcal{E}_1\cap\mathcal{E}_2} f_j(q) \geq \mathbf{1}_{\{i\in\mathcal{E}_1\cap\mathcal{E}_2\}} \cdot f_i(q) + \frac{1}{2^{d_i-1}} \sum_{j\in\mathcal{C}(i)} \sum_{k\in\mathcal{T}^-(j)} q_k,$$

and it suffices to show that

$$\mathbf{1}_{\{i\in\mathcal{E}_1\cap\mathcal{E}_2\}} \cdot f_i(q) + \frac{1}{2^{d_i-1}} \sum_{j\in\mathcal{C}(i)} \sum_{k\in\mathcal{T}^-(j)} q_k \geq \frac{1}{2^{d_i}} \sum_{j\in\mathcal{T}^-(i)} q_j,$$

which is equivalent to

$$\mathbf{1}_{\{i\in\mathcal{E}_1\cap\mathcal{E}_2\}} \cdot f_i(q) \geq \frac{1}{2^{d_i}} \sum_{j\in\mathcal{C}(i)} \left(q_j - \sum_{k\in\mathcal{T}^-(j)} q_k \right). \tag{22}$$

We show that (22) holds under three different cases. Firstly, if $i \notin \mathcal{E}_1$, i.e., $f_i(q) \leq 0$, then $\mathbf{1}_{\{i\in\mathcal{E}_1\cap\mathcal{E}_2\}} \cdot f_i(q) = 0$, and

$$\sum_{j\in\mathcal{C}(i)} \left(q_j - \sum_{k\in\mathcal{T}^-(j)} q_k \right) \leq \sum_{j\in\mathcal{T}^-(i)} (-1)^{d(i,j)+1} q_j = f_i(q) \leq 0,$$

implying that (22) holds. Next, if $i \notin \mathcal{E}_2$, i.e., $\left\|q_{\mathcal{C}(i)}\right\|_1 = 0$, then $\mathbf{1}_{\{i\in\mathcal{E}_1\cap\mathcal{E}_2\}} \cdot f_i(q) = 0$, and

$$\sum_{j\in\mathcal{C}(i)} \left(q_j - \sum_{k\in\mathcal{T}^-(j)} q_k \right) = - \sum_{j\in\mathcal{C}(i)} \sum_{k\in\mathcal{T}^-(j)} q_k \leq 0,$$

implying that (22) holds. Finally, if $i \in \mathcal{E}_1 \cap \mathcal{E}_2$, which implies $f_i(q) > 0$ and $\mathbf{1}_{\{i\in\mathcal{E}_1\cap\mathcal{E}_2\}} = 1$. Then, by (12),

$$\begin{aligned}
f_i(q) - \frac{1}{2^{d_i}} \sum_{j\in\mathcal{C}(i)} \left(q_j - \sum_{k\in\mathcal{T}^-(j)} q_k \right) &= \sum_{j\in\mathcal{T}^-(i)} (-1)^{d(i,j)+1} q_j - \frac{1}{2^{d_i}} \sum_{j\in\mathcal{C}(i)} \left(q_j - \sum_{k\in\mathcal{T}^-(j)} q_k \right) \\
&= \sum_{j\in\mathcal{C}(i)} \left(\left(1 - \frac{1}{2^{d_i}}\right) q_j + \sum_{k\in\mathcal{T}^-(j)} \left((-1)^{d(i,k)+1} + \frac{1}{2^{d_i}} \right) q_k \right) \\
&\geq \sum_{j\in\mathcal{T}^-(i)} \left(1 - \frac{1}{2^{d_i}}\right) (-1)^{d(i,j)+1} q_j \\
&= \left(1 - \frac{1}{2^{d_i}}\right) f_i(q) > 0,
\end{aligned}$$

where the inequality holds since $q_j \geq 0$ for every $j \in \mathcal{A}$. This concludes the proof.

C Postponed proofs in Section 6

C.1 Proof of Proposition 4

Fix valid initial states $Q(0)$ and $Q'(0)$, and let $i \in \mathcal{A}$ be the type of the arriving agent at time 1. To establish the consistency of Π, we prove a stronger statement that there exists a coupling P between $Q(1)$ and $Q'(1)$ such that

$$\Pr_{(Q(1),Q'(1))\sim P} \left[\left\|Q(1) - Q'(1)\right\|_1 \leq \left\|Q(0) - Q'(0)\right\|_1 \right] = 1. \tag{23}$$

For the ease of presentation, we will use a match being performed in Q (resp. Q') to denote the match being performed at time 1 with initial state $Q(0)$ (resp. $Q'(0)$.)

Firstly, if $Q_{\mathcal{N}(i)}(0) = Q'_{\mathcal{N}(i)}(0) = \mathbf{0}$, then Π will not perform any matches in both Q and Q'. Hence, $Q(1) = Q(0) + \mathbf{1}_{\{i\in\mathcal{A}'\}} \cdot \mathbf{e}_i$ and $Q'(1) = Q'(0) + \mathbf{1}_{\{i\in\mathcal{A}'\}} \cdot \mathbf{e}_i$, which implies

$$\left\|Q(1) - Q'(1)\right\|_1 = \left\|Q(0) - Q'(0)\right\|_1,$$

concluding (23).

Next, if $Q_{\mathcal{N}(i)}(0) = \mathbf{0}$ and $Q'_{\mathcal{N}(i)}(0) \neq \mathbf{0}$, then Π will not perform any matches in Q and will perform a (random) match, denoted as $m(i,j)$ for $j \in \mathcal{N}(i)$, in Q', which implies $Q(1) = Q(0) + \mathbf{1}_{\{i\in\mathcal{A}'\}} \cdot \mathbf{e}_i$ and $Q'(1) = Q'(0) - \mathbf{e}_j$. Since we must have $Q_j(0) = Q'_i(0) = 0$ and $Q'_j(0) > 0$, it follows that

$$\begin{aligned}
\left\|Q(1) - Q'(1)\right\|_1 &= \left\|Q(0) + \mathbf{1}_{\{i\in\mathcal{A}'\}} \cdot \mathbf{e}_i - Q'(0) + \mathbf{e}_j\right\|_1 \\
&= \sum_{k\in\mathcal{A}\setminus\{i,j\}} |Q_k(0) - Q'_k(0)| + |Q_i(0) - Q'_i(0) + \mathbf{1}_{\{i\in\mathcal{A}'\}}| + |Q_j(0) - Q'_j(0) + 1| \\
&= \left\|Q(0) - Q'(0)\right\|_1 + \mathbf{1}_{\{i\in\mathcal{A}'\}} - 1 \\
&\leq \left\|Q(0) - Q'(0)\right\|_1,
\end{aligned}$$

implying that the naive coupling between $Q(1)$ and $Q'(1)$ (observe that $Q(1)$ is deterministic) satisfies (23). The case where $Q_{\mathcal{N}(i)}(0) \neq \mathbf{0}$ and $Q'_{\mathcal{N}(i)}(0) = \mathbf{0}$ can be handled analogously.

It remains to consider the case in which $Q_{\mathcal{N}(i)}(0) \neq \mathbf{0}$ and $Q'_{\mathcal{N}(i)}(0) \neq \mathbf{0}$, where the consistency of Π now comes into play. We slightly abuse notations and define $x, x' \in \mathbb{R}^{\mathcal{N}(i)}$ such that for every $j \in \mathcal{N}(i)$, $x_j = x^{\Pi}_{m(i,j)}(Q(0), i)$ and $x'_j = x^{\Pi}_{m(i,j)}(Q'(0), i)$. Note that $\sum_{j\in\mathcal{N}(i)} x_j = \sum_{j\in\mathcal{N}(i)} x'_j = 1$, i.e., x and x' are distributions over $\mathcal{N}(i)$. To couple $Q(1)$ and $Q'(1)$, it is equivalent to couple $j \sim x$ and $j' \sim x'$ since j and j' uniquely determine $Q(1)$ and $Q'(1)$, respectively. Define $\mathcal{N}_< := \{j \in \mathcal{N}(i) \mid Q_j(0) < Q'_j(0)\}$, $\mathcal{N}_= := \{j \in \mathcal{N}(i) \mid Q_j(0) = Q'_j(0)\}$, and $\mathcal{N}_> := \{j \in \mathcal{N}(i) \mid Q_j(0) > Q'_j(0)\}$. We first characterize the desired coupling structure in the following lemma.

Lemma 16. *If $j = j'$, $j \in \mathcal{N}_>$, or $j' \in \mathcal{N}_<$, then the states $Q(1)$ and $Q'(1)$ respectively induced by j and j' satisfy*

$$\left\|Q(1) - Q'(1)\right\|_1 \leq \left\|Q(0) - Q'(0)\right\|_1 .$$

Proof. Note that $Q(1) = Q(0) - \mathbf{e}_j$ and $Q'(1) = Q'(0) - \mathbf{e}_{j'}$. If $j = j'$, then

$$\left\|Q(1) - Q'(1)\right\|_1 = \left\|Q(0) - \mathbf{e}_j - Q'(0) + \mathbf{e}_j\right\|_1 = \left\|Q(0) - Q'(0)\right\|_1 .$$

Next, if $j \neq j'$ and $j \in \mathcal{N}_>$, then

$$\begin{aligned}\left\|Q(1) - Q'(1)\right\|_1 &= \left\|Q(0) - \mathbf{e}_j - Q'(0) + \mathbf{e}_{j'}\right\| \\ &= \sum_{k\in\mathcal{N}(i)\setminus\{j,j'\}} |Q_k(0) - Q'_k(0)| + |Q_j(0) - Q'_j(0) - 1| + |Q_{j'}(0) - Q'_{j'}(0) + 1| \\ &\leq \left\|Q(0) - Q'(0)\right\|_1 ,\end{aligned}$$

where the last inequality holds since $Q_j(0) > Q'_j(0)$ and $|Q_{j'}(0) - Q'_{j'}(0) + 1| \leq |Q_{j'}(0) - Q'_{j'}(0)| + 1$. Finally, the case where $j' \in \mathcal{N}_<$ can be handled analogously. □

Now, it suffices to give a coupling P between j and j' such that with probability 1, at least one condition in Lemma 16 holds for the pair (j, j') sampled from P. To start with, we couple elements in $\mathcal{N}_=$ maximally, i.e., for every $k \in \mathcal{N}_=$,

$$\Pr_{(j,j')\sim P}[j = j' = k] = \min\{x_k, x'_k\}.$$

Then, we maximally couple the case $j' \in \mathcal{N}_<$ with the case $j \in \mathcal{N}_<$ and the remaining probability mass in the case $j \in \mathcal{N}_=$, i.e.,

$$\begin{aligned}\Pr_{(j,j')\sim P}[j \in \mathcal{N}_< \cup \mathcal{N}_=, j' \in \mathcal{N}_<] &= \min\left\{\sum_{k\in\mathcal{N}_<} x_k + \sum_{k\in\mathcal{N}_=} (x_k - x'_k)^+, \sum_{k\in\mathcal{N}_<} x'_k\right\} \\ &= \sum_{k\in\mathcal{N}_<} x_k + \sum_{k\in\mathcal{N}_=} (x_k - x'_k)^+,\end{aligned}$$

where the last equality holds by (17). So far, we have consumed all probability mass for the case $j \in \mathcal{N}_< \cup \mathcal{N}_=$. Finally, we arbitrarily couple the case $j \in \mathcal{N}_>$ with the remaining probability mass for j'. The proof is completed by observing

$$\Pr_{(j,j')\sim P}[(j = j') \vee (j \in \mathcal{N}_>) \vee (j' \in \mathcal{N}_<)] = 1$$

and by Lemma 16.

C.2 Proof of Corollary 1

Fix an arbitrary static priority policy Π with priority order $\succ$. Fix valid states q and q', and fix an agent type $i \in \mathcal{A}$ satisfying $q_{\mathcal{N}(i)} \neq \mathbf{0}$ and $q'_{\mathcal{N}(i)} \neq \mathbf{0}$. Denote $x := x^{\Pi}(q, i)$ and $x' := x^{\Pi}(q', i)$. Define $\mathcal{N}_{=} := \{j \in \mathcal{N}(i) \mid q_j = q'_j\}$, $\mathcal{N}_{<} := \{j \in \mathcal{N}(i) \mid q_j < q'_j\}$, and $\mathcal{N}_{>} := \{j \in \mathcal{N}(i) \mid q_j > q'_j\}$. Our goal is to show that (17) holds. Since Π is deterministic, there exist $j, j' \in \mathcal{N}(i)$ such that $x = \mathbf{e}_{m(i,j)}$ and $x' = \mathbf{e}_{m(i,j')}$. It is easy to verify that (17) holds when $j = j'$, $j \in \mathcal{N}_{>}$, or $j' \in \mathcal{N}_{<}$. Now, assume that $j \neq j'$, $j \notin \mathcal{N}_{>}$, and $j' \notin \mathcal{N}_{<}$. On one hand, if $j' \succ j$, then we must have $q_{j'} = 0$ by the matching rule of Π, contradicting the assumption that $j' \notin \mathcal{N}_{<}$. On the other hand, if $j \succ j'$, then we must have $q'_j = 0$, contradicting the assumption that $j \notin \mathcal{N}_{>}$.

The proof for the longest-queue policy, denoted as **LQ**, follows a similar argument. Fix valid states q and q', and fix an agent type $i \in \mathcal{A}$ satisfying $q_{\mathcal{N}(i)} \neq \mathbf{0}$ and $q'_{\mathcal{N}(i)} \neq \mathbf{0}$. Denote $x := x^{\mathbf{LQ}}(q, i)$ and $x' := x^{\mathbf{LQ}}(q', i)$. Define $\mathcal{N}_{=}$, $\mathcal{N}_{<}$, and $\mathcal{N}_{>}$ similarly. Our goal is to show that (17) holds. Since **LQ** is deterministic, there exist $j, j' \in \mathcal{N}(i)$ such that $x = \mathbf{e}_{m(i,j)}$ and $x' = \mathbf{e}_{m(i,j')}$. It is easy to verify that (17) holds when $j = j'$, $j \in \mathcal{N}_{>}$, or $j' \in \mathcal{N}_{<}$. Now, assume that $j \neq j'$, $j \notin \mathcal{N}_{>}$, and $j' \notin \mathcal{N}_{<}$. On one hand, if $j \in \mathcal{N}_{=}$, then we must have $q_{j'} < q'_{j'}$ by the matching rule of **LQ**, contradicting the assumption that $j' \notin \mathcal{N}_{<}$. On the other hand, if $j \in \mathcal{N}_{<}$, then $q'_{j'} \geq q'_j > q_j \geq q_{j'}$, contradicting the assumption that $j' \notin \mathcal{N}_{<}$ as well.

D Extension to non-empty initial state

In this section, we discuss how to modify our results when we start from a non-empty initial state, i.e., $Q(0) = q \neq \mathbf{0}$, where we assume that q is finite and does not depend on the time horizon T. In this case, the upper bound for $\mathcal{R}^*(t)$ given in (2) becomes

$$\mathcal{R}^*(t) = \mathbb{E}\left[\begin{array}{ll} \max & r^T y \\ \text{s.t.} & My \leq q + A(t) \\ & y \in \mathbb{Z}^d_{\geq 0} \end{array}\right] \leq \begin{array}{ll} \max & r^T x \\ \text{s.t.} & Mx \leq q + t\lambda. \\ & x \in \mathbb{R}^d_{\geq 0} \end{array}$$

By the Lipschitz continuity of LP with respect to the RHS of the constraints [MS87], the value of the upper bound only increases by a constant that depends on q, r, M. Hence, given an arbitrary (not necessarily valid) initial state, we modify a greedy policy as follows. At the beginning, if the initial state is invalid, i.e., there are two adjacent non-empty queues, we keep performing matches in an arbitrary order until the state becomes valid. Then, we run the greedy policy starting from the resulting valid state. By doing so, the regret of the policy will only increase by a constant depending on q, r, M but, importantly, not depending on ϵ.