Dynamic Matching:
Characterizing and Achieving Constant Regret

Süleyman Kerimov, Itai Ashlagi, Itai Gurvich*

April 15, 2023

Abstract

We study how to optimally match agents in a dynamic matching market with heterogeneous match cardinalities and values. A network topology determines the feasible matches in the market. In general, a fundamental trade-off exists between short-term value—which calls for performing matches frequently—and long-term value—which calls, sometimes, for delaying match decisions in order to perform better matches.

We find that in networks that satisfy a general position condition, the tension between short- and long-term value is limited, and a simple periodic clearing policy (nearly) maximizes the total match value simultaneously at all times. Central to our results is the general position *gap* ϵ; a proxy for capacity slack in the market. With the exception of trivial cases, no policy can achieve an all-time regret that is smaller, in terms of order, than ϵ^{-1}. We achieve this lower bound with a policy, which periodically resolves a natural matching integer linear program, provided that the delay between resolving periods is of the order of ϵ^{-1}. Examples illustrate the necessity of some delay to alleviate the tension between short- and long-term value.

1 Introduction

We study a centralized dynamic matching market, in which agents arrive stochastically over time, matches can be multilateral, and match values are heterogeneous. Uncertainty in agents' arrivals creates an inherent trade-off between short- and long-term allocative efficiency; being overly greedy may compromise opportunities to perform valuable matches in the future.

*Kerimov: Jones Graduate School of Business, Rice University, kerimov@rice.edu. Ashlagi: Department of Management Science and Engineering, Stanford University, iashlagi@stanford.edu. Gurvich: Kellogg School of Management, Northwestern University, i-gurvich@kellogg.northwestern.edu. This work is supported by the Department of Defense grant STTR A18B-T007 and by NSF grant CMM-2010940.

Carpooling platforms delay match decisions to better pool passengers with each other, yet passengers may wait longer to be served. Kidney exchange platforms, which arrange exchanges between incompatible patient-donor pairs, can form a match as soon as it becomes feasible, or wait for more pairs in order to generate exchanges that yield more life years from transplants.[1] Programs in the Netherlands, the United Kingdom, Canada and Australia form matches every 3 or 4 months (Ferrari et al., 2014; Malik and Cole, 2014; Johnson et al., 2008). In contrast, programs in the United States have gradually moved towards daily matching; this practice raised concerns that matching frequently may harm efficiency (Gentry and Segev, 2015).

To better understand this tension between short and long-term objectives, and to speak to the reality described above, we seek to address the following questions: (i) How do we formally measure this tension, and how does it depend on the market primitives? (ii) How should a planner match agents dynamically in order to achieve the best possible balance between short- and long-term objectives? (iii) If a periodic matching policy is applied, what is the right delay between consecutive match decisions?

We introduce a queueing perspective to study these questions and model the market as a network of matching queues. In our model agents arrive sequentially to the market, and the type of an arrival is drawn from a known distribution over finitely many types. A given network topology determines the set of feasible matches. Matches include two or more agent types, and match values are heterogeneous (see, e.g., Figure 1). We impose no a priori assumptions on the underlying network topology; it may be acyclic, or it may include cycles. A matching policy determines when and which matches to perform, and agents leave the market once they are matched.

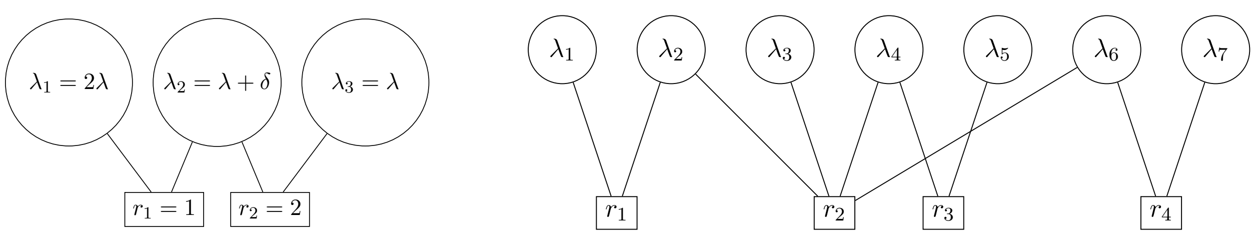

Figure 1: Matching network graphs. Circles and rectangles represent agent types and matches, respectively. Agents arrive sequentially, and an arrival is of type$-i$ with probability λ_i. When match$-m$ is performed once, a value of r_m is collected. (LEFT) A network with 3 agent types and 2 (two-way) matches. The leftmost match includes one agent of each of types 1 and 2, and generates a value of r_1. (RIGHT) A network with 7 agent types and 4 matches. The (multi-way) match yields a value of r_2 and includes one agent from each of 4 different agent types.

To study the trade-off between short- and long-term allocative efficiency, we use a notion of

[1]For example, fewer tissue type mismatches or better age matches may increase life years from transplants.

all-time regret. Given a fixed horizon of length t, the maximum allocative efficiency is achieved by waiting until time t, and only then forming an optimal set of matches. The *static planning problem* is a deterministic counterpart of this upper bound where the arrivals are replaced by their means. For the network in Figure 1(LEFT) with $0 < \delta < \lambda$, the deterministic counterpart performs δ many match-1 and λ many match-2 per time unit; and it collects a match value of $r_1\delta + r_2\lambda$ per time unit. The *regret* of a matching policy at a *fixed* time t, measures the difference between this upper bound and the value generated by the matching policy by time t; the all-time regret measures the supremum over all times t. In general, a smaller regret in short-term may yield larger regret in long-term; in that case the all-time regret will be large. If it is possible to have a small regret *simultaneously* at all times, then the tension between short-term and long-term is moot.

We prove that this is indeed possible for matching networks that satisfy a general position condition. General position is nothing but the requirement that the deterministic counterpart has a non-degenerate optimal solution. In a matching network, the non-degeneracy implies (loosely speaking) some "imbalance" in the market.

Before describing the main results, it will be helpful to discuss a couple of examples. Consider the network in Figure 1(LEFT), where $0 < \delta < \lambda$. Because $r_2 > r_1$, the deterministic counterpart matches λ many type-2 agents with type-3 agents, and matches the remaining $\delta > 0$ many with type-1 agents. Now consider the dynamic (stochastic) market, where the planner adopts a periodic clearing policy: every τ time periods, the planner solves a static matching problem given the number of agents in each queue. In *expectation*, there are $\delta\tau$ more arrivals of agent type-2 than those of type-3. But the smaller the δ, the greater the *probability* that the number of type-2 arrivals will not suffice to match all type-3 arrivals during the period of length τ. Conversely, the greater the δ, the greater the probability that we will be able to match all arriving type-3 agents, in alignment with the deterministic upper bound. If $\delta = 0$ (in violation of the general position condition), regret inevitably—regardless of τ—grows over time; see Section 2.

For fixed δ the greater the τ, the greater the probability that the number of type-2 arrivals over the inter-action delay τ exceeds that of type-3. This τ is a design choice and, in some networks, this choice matters. Consider the network in Figure 2, and assume that the planner is using a periodic clearing policy with an inter-action delay τ. When $\delta = 0.05$, we note that the regret grows when $\tau = 5$, but it is bounded when $\tau = 20$. When $\delta = 0.01$, the period length $\tau = 20$ no longer maintains a bounded regret, but $\tau = 100$ does. To maintain a bounded all-time regret, τ cannot be too small. Picking τ to be too large is also a problem, since we might be unnecessarily giving up on short-term value.

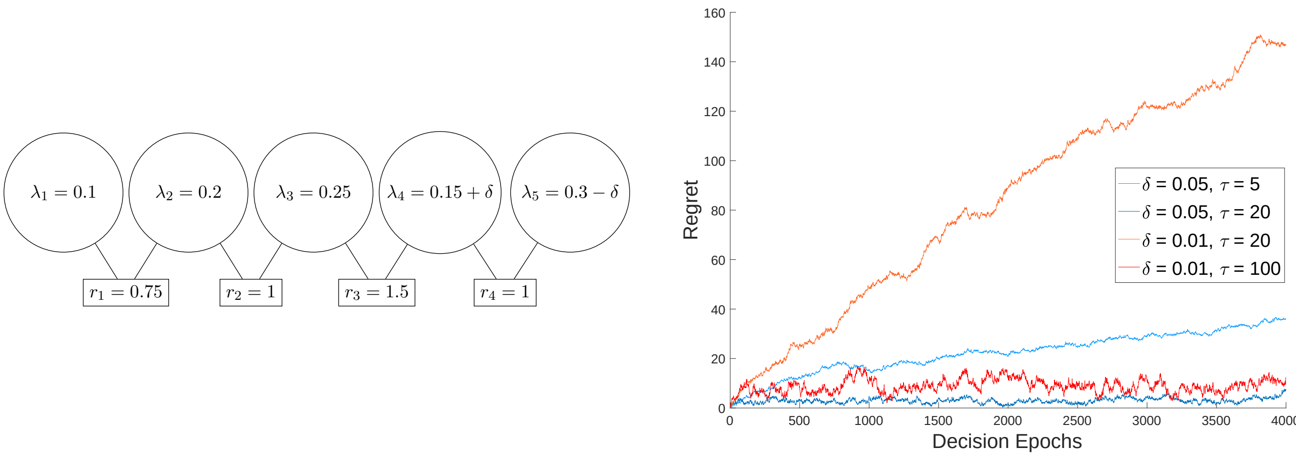

Figure 2: The regret of our proposed periodic clearing policy applied to the network on the left. Both the period length τ and the parameter δ are varied. For any $\delta \leq 0.05$, the optimal static solution is $(0.1, 0.1, 0.15, \delta)$ for the four matches, respectively. The plotted regret is based on 10 replications. Note that since the $x-$axis corresponds to decision epochs, the time horizon is $4000 \cdot \tau$.

Main contributions. First, we introduce the *general position gap*, denoted by ϵ, that quantifies the (in)stability of the network, and it is characterized explicitly in terms of the network primitives. Loosely speaking, this quantity captures the "inherent thickness" in the market via the imbalance in the arrival probabilities. Mathematically, the general position gap is the minimum over sizes of matches and unmatched agents in each queue based on the optimal static solution. For the network in Figure 1(LEFT), $\epsilon = \min\{\delta, \lambda, 2\lambda - \delta\}$; in Figure 2, $\epsilon = \min\{0.1, 0.1, 0.15, \delta, 0.3 - 2\delta\}$.

Second, we show that with the exception of trivial cases, no matching policy (periodic clearing or not) can achieve an all-time regret that is smaller, in terms of order, than ϵ^{-1}. We introduce a periodic resolving policy that achieves this lower bound and therefore, not only maintains the regret uniformly bounded simultaneously at all times, but also achieves the optimal scaling for the all-time regret. At each clearing period, one resolves a simple integer linear program that maximizes the total match value given the state of the market (the number of agents in each queue). The lower bound is attained by this policy, provided that the inter-action delay, i.e., the length between two consecutive resolving periods is of the order of ϵ^{-1}. In other words, under a carefully designed resolving policy, the market is just thick enough at each clearing period (without unnecessary waiting) to achieve high allocative efficiency at all times. Overall, the general position gap prescribes a precise operational measure for market "thickness"; it is inversely proportional to the attainable regret and the ideal clearing period length.

Delaying actions, we show, is generally necessary to maintain bounded regret at all times. Consider, for example, the network in Figure 1(RIGHT), and suppose that match-2 is a high-value match. This introduces a complementarity that prevents greedy-like policies to perform well; acting greedily (over) utilizes other matches abundantly at the expense of match-2; see Example 3.2.

Finally, we prove that in acyclic matching networks, the general position gap ϵ can be formalized as a measure of capacity slack (the excess of capacity above demand) akin to similar notions in standard queueing networks. In these networks, the optimal static solution effectively "labels" a subset of agent types as servers (and their total arrival rate as capacity) and the remaining set of agent types as customers (and their total arrival rate as demand).

1.1 Related literature

Value maximization, as well as the tension between value and delay, have received significant attention in the matching literature. At the risk of being a bit coarse, we divide the related literature into two streams characterized by their modeling language.

The first stream is based on random graphs, where agents arrive over time and form an edge with existing agents with some exogenous probabilities. A large subset of this stream, motivated by kidney exchange, is concerned with dynamic matching under homogeneous values—maximizing the total match value is the same, in this case, as maximizing the total number of matched agents. Anderson et al. (2017); Ashlagi et al. (2019b) focus on the average waiting time of agents and show that greedy policies achieve near optimality as the exogenous match probability tends to zero, which suggests that waiting to thicken the market is not beneficial. Ashlagi et al. (2018); Akbarpour et al. (2020) explicitly model agents' departures (abandonments) and find that greedy policies maximize the total number of matches in large markets. If departure times (agents' patience levels) are observed, matching just before departures yields an improvement over greedy matching (Akbarpour et al., 2020).

A growing literature considers dynamic matching under heterogeneous match values. Blanchet et al. (2020) studies a two-sided market model with departures, in which the value from matching a single buyer to a single seller (a two-way match) is drawn from a given distribution. The optimal frequency of match decisions depends on the tail of the value distribution, where the policies that are studied include population and utility threshold policies. In our model, there is a finite number of match types (rather than a continuum), and the feasibility of matches is determined, instead, by a given network topology. In addition, our model allows for matches to include more than two agent types (multi-way matches). Ashlagi et al. (2019a); Collina et al. (2020) also identify the need of delaying actions in a model with departures. Dynamic policies based on heuristics for continuation values were studied in the context of kidney exchange (Dickerson et al., 2016; Li et al., 2019).

Other papers in this stream consider incentives and decentralized decisions Leshno (2011); Arnosti and Shi (2019); Baccara et al. (2020). Our model is of a central decision maker, and in that sense we are closer to Dickerson et al. (2012), which develop a heuristic to approximate the full dynamic program and overcomes the "curse of dimensionality", and to Karp et al. (1990); Goel and Mehta (2008); Feldman et al. (2009); Manshadi et al. (2012), which benchmark against an offline upper bound.

Our work uses the modeling language of queueing networks rather than that of random graphs. It considers environments, in which *match values are not binary*, and the number of agent and match types are finite.

Within the queueing literature, a subset of papers focuses on performance evaluation of specific important policies; e.g., see Caldentey et al. (2009); Adan et al. (2018); Afeche et al. (2021) and the references therein. Several recent papers succeeded in reducing the control problem's complexity by relying on heavy-traffic approximations (Gurvich and Ward, 2014; Bušic and Meyn, 2014; Nazari and Stolyar, 2019). Gurvich and Ward (2014); Bušic and Meyn (2014) study the minimization of heterogeneous delay costs. For homogeneous delay costs, Ünver (2010) establishes the optimality of a greedy policy, if all matches are two-way (involving one donor and one recipient, in the context of kidney exchange); it also underscores the value of delaying match decisions in networks with multi-way matches. Nazari and Stolyar (2019), like us, study value maximization, but focuses on the long-run average value. Our main focus is on finite horizon optimization and on the trade-off between short- and long-term value. The policy we devise is, in particular, long-run average optimal.

Aouad and Saritac (2020) study matching networks when agent departures are allowed. These departures make the problem more difficult, as *any* delay between actions may sacrifice value when agents are sufficiently impatient. The authors introduce algorithms that achieve, in the long-run, a constant percent of the upper bound (the optimality gap, then, grows with the horizon). By considering a more limited family of networks and assuming that agents are patient, we make headway in the refined understanding of matching networks that, we believe, can subsequently inform the design of algorithms for networks with departures; we revisit this in the concluding remarks.

This paper is also related to recent work on achieving constant regret in dynamic resource allocation problems; e.g., see Bumpensanti and Wang (2020); Vera et al. (2020); Vera and Banerjee (2021). In these papers, it is proved that policies, which resolve at *each arrival* an intuitive linear program, can achieve constant regret in the online packing context, where an initial supply of inventory is depleted over a finite horizon by arriving requests. Requests must be accepted or rejected on the spot (there is no queue), and the criterion is to maximize the value collected by the end of the horizon. Of conceptual importance is Jasin and Kumar (2012), where a non-degeneracy assumption supports the optimality of such greedy resolving policies in the packing setting. While the differences are significant, both dynamic matching and online packing problems can be conceptually framed as specific instances of online linear programming; e.g., see Li and Ye (2020) and the references therein.

Notation. For real numbers x and y, we use $x \wedge y := \min\{x, y\}$, $(x)^+ := \max\{0, x\}$ and $(x)^- := \max\{0, -x\}$. We follow the accepted meaning of little o, big $\mathcal{O}$ and big Ω. For example $a_t = \Omega(b_t)$ for all $t > 0$ (for non-negative a_t, b_t) means that $\liminf_{t\to\infty} a_t/b_t > 0$. We write $[1, n]$ to denote the set of positive integers $\{1, 2, \ldots, n\}$.

2 Model

Matching network and dynamics. There is a finite set of *agent types* $\mathcal{A} = \{1, 2, \ldots, n\}$ and a finite set of *matches* $\mathcal{M} = \{1, \ldots, d\}$. Each match $m \in \mathcal{M}$ corresponds to a subset of at least two agent types. We denote by $\mathcal{A}(m)$ the set of agent types participating in match$-m$. The *network topology* is given by a *matching matrix* $M \in \{0, 1\}^{n\times d}$, where $M_{im} = 1$ if and only if $i \in \mathcal{A}(m)$. We assume that each agent type is participating in at least one match.

Agents arrive in discrete time following a multinomial distribution: at each time $t \in \mathbb{N}$, an arrival is of type$-i$ with probability $\lambda_i > 0$, where $\sum_{i\in\mathcal{A}} \lambda_i = 1$. Match$-m$ is *feasible* at time t, if there is at least one agent type$-i$ present in the market at time t, for all $i \in \mathcal{A}(m)$. When match$-m$ is performed once, it includes one agent of each type in $\mathcal{A}(m)$, and generates a *value* of $r_m > 0$. We refer to the tuple $\mathcal{G} := (M, \lambda, r)$ as the *matching network*.

To track the state of the market, we maintain a queue for each agent type, and agents join their type-dedicated queues upon arrival. All queues are empty at $t = 0$, and we denote by A_i^t the number of arrivals to queue$-i$ by time t. Matches are performed instantaneously (after which the matched agents leave the market), and we denote the pre-match queue-length vector at time t by Q^t. Note that at most $\min_{i\in\mathcal{A}(m)} Q_i^t$ many matches of $m \in \mathcal{M}$ can be performed at time t.

Matching network graph. The network topology is a hypergraph, where each agent type is a vertex and each match is a collection of vertices—which are the agent types that participate in the match. We represent this hypergraph by a simple bipartite graph, where agent types and matches are the vertices, and there is an edge between agent type$-i$ and match$-m$ if and only if $i \in \mathcal{A}(m)$. We refer to this bipartite graph as the *matching network graph*, and we denote this graph by $\mathcal{G}$ as a slight abuse of notation. Figure 1 is the first instance of multiple matching network graphs that we will use throughout the paper. In the figures, circles and rectangles represent agent types and matches, respectively, and we indicate the arrival probabilities and match values in their corresponding shapes.

Performance measure. A *matching policy* maps histories of arrivals and performed matches to a (possibly empty) set of matches and determines how many times each of these matches will be performed at each time t. Such a policy can be represented by a right-continuous with left limits non-anticipative increasing process $D_m := (D_m^t, t \geq 0)$, where D_m^t is the total number of times match$-m$ is performed by time t; $\Delta D_m^t := D_m^t - D_m^{t^-}$ is then the number of times match$-m$ is performed at time t. An *admissible* matching policy D must satisfy the following:

$$Q^t = A^t - MD^{t^-} \text{ for all } t > 0. \tag{1}$$

Denote by Π the set of all admissible matching policies. We add the superscript D on expectations to make explicit the dependence on the policy. We use Q^{t^+} to denote the post-match queue-length vector at time t, i.e., $Q^{t^+} = Q^t - M\Delta D_m^t$.

The expected *total value* collected by time t, under a matching policy D, is given by

$$\mathcal{R}^{D,t} := \mathbb{E}^D[r \cdot D^t].$$

The optimal value *for fixed* t, $\mathcal{R}^{*,t} := \max_{D\in\Pi} \mathcal{R}^{D,t}$, is trivially attained by the *ultimate batching policy*, which takes no action until time t, and performs matches according to an optimal solution of the (static) weighted matching problem at time t. The optimal value $\mathcal{R}^{*,t}$ is then the expectation of the following static problem:

$$\mathcal{R}^{*,t} = \mathbb{E}\left[\begin{array}{ll} \max & r \cdot y \\ \text{s.t.} & My \leq A^t \\ & y \in \mathbb{Z}^d_{\geq 0} \end{array}\right].$$

Conceptually, it is useful to think of $\mathcal{R}^{*,\cdot}$ as tracking the total collected value of a decision maker that makes decisions continuously, but the decision maker is allowed, at all times, to correct past decisions (unmatch some agents and match new ones); this is a *hindsight upper bound.* A matching policy is hindsight optimal if it is, at all t, *almost* as good as this upper bound.

Definition 2.1 (hindsight optimality)**.** *A matching policy D is* hindsight optimal *if*

$$\mathcal{R}^{*,t} - \mathcal{R}^{D,t} = \mathcal{O}(1) \text{ for all } t > 0,$$

which implies, in particular, $\mathcal{R}^{D,t}/\mathcal{R}^{*,t} = 1 - \mathcal{O}(1/t)$ for all $t > 0$.

This notion of optimality—with its focus on the total collected value at all times—allows us to concentrate on the tension between short- and long- term value; whether *it is possible* to act frequently and remain near-optimal at all times. Explicit delay penalties naturally encourage taking frequent actions. We explicitly model delay penalties/holding costs in §6 and show that our proposed matching policies achieve near-optimality in that case as well.

Remark 2.1. Hindsight optimality implies optimality under other criteria. For instance, given a *finite horizon T*, a hindsight optimal matching policy makes a constant number of "mistakes" *that does not grow with the horizon*, i.e., $\mathcal{R}^{*,T} - \mathcal{R}^{D,T} = \mathcal{O}(1)$. In particular, the policy is optimal in the *long-run average* sense, since

$$\frac{\mathcal{R}^{*,T} - \mathcal{R}^{D,T}}{\mathcal{R}^{*,T}} = \mathcal{O}(1/T) \to 0 \text{ as } T \to \infty,$$

with a convergence rate of $1/T$.

Another instance is a discounted infinite horizon model, where the *discounted collected value* with a discount factor $\beta \in (0, 1)$ under a matching policy D is defined as

$$\mathcal{R}^D_\beta := \mathbb{E}^D\left[\sum_{t=0}^{\infty} \beta^t (r \cdot \Delta D^t)\right].$$

Let $\mathcal{R}_\beta^* := \max_{D\in\Pi} \mathcal{R}_\beta^D$ and $\mathcal{R}_\beta^U := (1-\beta)\sum_{t=0}^{\infty} \beta^t \mathcal{R}^{*,t}$. Then for any matching policy D, we have $\mathcal{R}_\beta^U \geq \mathcal{R}_\beta^* \geq \mathcal{R}_\beta^D$. A hindsight optimal matching policy D satisfies $\mathcal{R}_\beta^U - \mathcal{R}_\beta^D = \mathcal{O}(1)$, and in particular, $\mathcal{R}_\beta^* - \mathcal{R}_\beta^D = \mathcal{O}(1)$. Since $\mathcal{R}_\beta^* = \Omega(1/(1-\beta))$, the relative error satisfies

$$\frac{\mathcal{R}_\beta^* - \mathcal{R}_\beta^D}{\mathcal{R}_\beta^*} = \mathcal{O}(1-\beta),$$

and shrinks as the effective horizon becomes longer (as $\beta \uparrow 1$). ■

2.1 The static-planning problem (SPP) and the general position condition

A natural upper bound for the optimal value $\mathcal{R}^{*,t}$ is given by the following optimization problem, where stochastic arrivals are replaced by their rates:

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

An optimal solution x_m^* of the problem on the right-hand side of (2) provides a first-order proxy for optimal match rate of match$-m$. The inequality in (2) simply follows from relaxing the integrality constraints and applying Jensen's inequality. With the change of variables $z = x/t$, we arrive at a deterministic relaxation, which we write in standard form as

$$\begin{array}{ll} \max & r\cdot z \\ \text{s.t.} & Mz + s = \lambda \\ & z \in \mathbb{R}^d_{\geq 0}, s \in \mathbb{R}^n_{\geq 0}. \end{array} \tag{SPP}$$

We refer to this formulation as the static-planning problem (SPP). Given an optimal solution (z^*, s^*) of (SPP), z_m^* is the (per period) number of times match$-m$ is performed under the optimal solution, whereas s_j^* corresponds to the leftovers (slack) added to queue$-j$ per period. We partition the set of matches and queues as follows:

$$\mathcal{M}_+ := \{m \in \mathcal{M} : z_m^* > 0\},\ \mathcal{M}_0 := \mathcal{M}\backslash\mathcal{M}_+,\ \mathcal{Q}_+ := \{j \in \mathcal{A} : s_j^* > 0\} \text{ and } \mathcal{Q}_0 := \mathcal{A}\backslash\mathcal{Q}_+,$$

where $\mathcal{M}_+$ is the set of *active* matches, $\mathcal{M}_0$ is the set of *redundant* matches, $\mathcal{Q}_+$ is the set of *under-demanded (non-empty)* queues, and $\mathcal{Q}_0$ is the set of *over-demanded (empty)* queues.

We expect "good" policies to be consistent with this partition. It should perform those matches with $z^*_m > 0$, but avoid performing the redundant matches. Similarly, over-demanded/empty queues should be as empty as possible, while those queues with $s^*_j > 0$ should grow with time. We formalize this intuition in §4.

A simple property of the optimal solution of (SPP) determines, as we will prove, whether it is possible to achieve hindsight optimality.

Definition 2.2 (general position)**.** *A matching network $\mathcal{G}$ satisfies the* general position *condition* (**GP**) *if* (SPP) *has a unique non-degenerate optimal solution* (z^*, s^*)*, i.e., all n basic variables in this solution are strictly positive.*

GP is straightforward to verify. Non-degeneracy means that $|\mathcal{M}_+| + |\mathcal{Q}_+| = n$ and is, thus, easy to verify by inspection. As to uniqueness, if the dual of (SPP) has a non-degenerate optimal solution, then the primal has a unique optimal solution by complementary slackness.

Uniqueness is mathematically useful and comes at no practical restriction. When there are multiple solutions, a small perturbation of the match value vector $r \leftarrow r + \mathcal{O}(1/T)$—where T is the horizon length in consideration—guarantees uniqueness. This does not affect hindsight optimality because this perturbation, for any $t \le T$, changes the benchmark $\mathcal{R}^{*,t}$ at most by a constant.

General position is in fact necessary to maintain a uniformly bounded regret. To see this, consider the network in Figure 3(LEFT). Observe that match-2 is used by the ultimate batching policy (that achieves the optimal value) *for any fixed time* $t > 0$ only if $A^t_2 > A^t_1$. Since $\lambda_1 = \lambda_2$, whether $A^t_1 \ge A^t_2$ or $A^t_1 < A^t_2$ is discovered only late in the horizon. Thus, *any* optimal policy for a fixed t, *must* withhold performing match-2 until time t. This inevitably means suboptimality for subintervals $[0, s]$, for any $s > 0$ sufficiently smaller than t (say $s = t/2$). Therefore, a policy D that is optimal for $s = t/2$ must have $\mathcal{R}^{*,t} - \mathcal{R}^{D,t} = \Omega(\sqrt{t})$. Figure 3(RIGHT) illustrates this, and a formal proof appears in the appendix.

The growing regret in Figure 3 stems from having equal arrival probabilities of agent types 1 and 2. Consider some perturbation on λ_2 now. Intuitively, the larger the difference between λ_2 and λ_1, the earlier one can decide whether to perform match-2, and one should also expect a smaller regret. The general position gap, which is defined next, captures the inherent imbalance in the network, or the "distance" from degeneracy.

Definition 2.3 (general position gap)**.** *Suppose that the matching network $\mathcal{G}$ satisfies* **GP***. We*

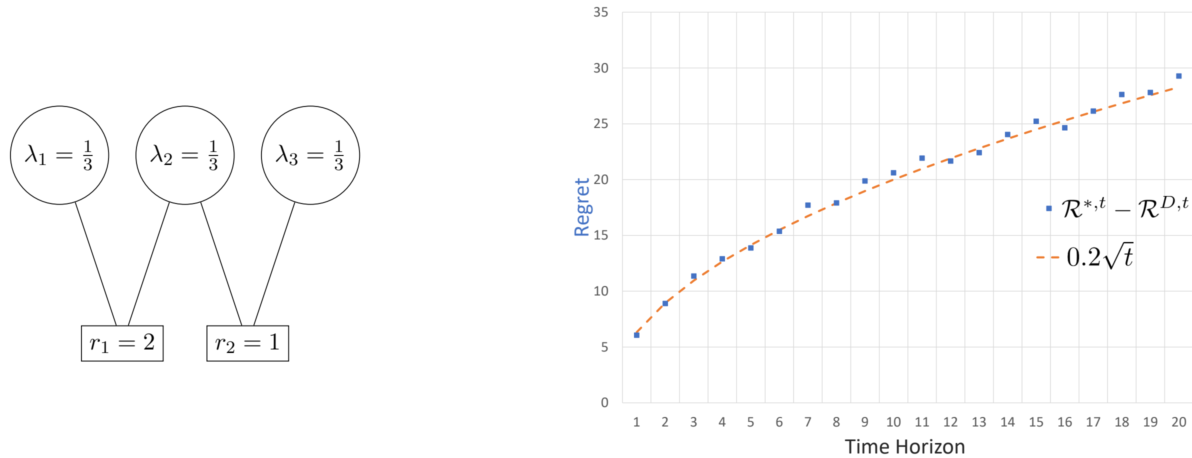

Figure 3: (LEFT) A network that violates **GP**. (RIGHT) The policy D performs one batched optimal solution at time $t/2$, and then another at time t. $\mathcal{R}^{*,t}$ is obtained by the ultimate batching policy at time t; we vary t (the time horizon is scaled down by 10^3). This captures a regret that is of the order of $\sqrt{t}$: optimizing total value at time $s < t$ necessitates a $\mathcal{O}(\sqrt{t})$ optimality gap at time t.

define the general position gap as

$$\epsilon = \min_{m \in \mathcal{M}_+} z_m^* \wedge \min_{j \in \mathcal{Q}_+} s_j^*.$$

The general position gap ϵ is, by definition, strictly positive, and since λ is a probability vector, $z_m^*, s_j^* < 1$ for all $m \in \mathcal{M}_+$ and $j \in \mathcal{Q}_+$ so that $0 < \epsilon < 1$. Mathematically, the general position captures the minimum entry among basic variables. For example in Figure 3, if one increases λ_2 by a sufficiently small constant $\delta > 0$ and decreases λ_1 by δ, then **GP** holds, where (SPP) has a unique optimal solution $z^* = (1/3 - \delta, 2\delta)$ and $s^* = (0, 0, 1/3 - 2\delta)$ with $\epsilon = 2\delta$.

For a large family of matching networks, ϵ can be thought of as a measure of capacity slack; see §5. Loosely speaking, the larger the general position gap ϵ, the larger the region of queue-lengths in the dynamic system that will enable performing "correct" matches by acting more frequently. As we will show later, the general position gap will be inversely proportional to the achievable regret and the desirable delay between decision epochs.

3 Main results

Our proposed matching policy—the *exhaustive resolving policy*—is a periodic clearing policy, where matches are performed at each decision epoch following an optimal solution of a natural linear integer program.

1. *Pre-processing and removal of redundant matches.* Solve (SPP) and identify the set $\mathcal{M}_0$. All redundant matches are *removed* from the network and *never* used ($D^t_m = 0$ for all $t > 0$ and $m \in \mathcal{M}_0$).[2] This decomposes the network into (possibly) multiple connected components, and the policy is applied to each component separately. Alternatively, the policy can be applied directly to the original network with an extra constraint that the matches in $\mathcal{M}_0$ are never used.

2. *Decision epochs.* Matches are performed only at decision epochs

$$t_k = k\tau,\ \ k \in \mathbb{N},$$

   where $\tau \in \mathbb{N}$ is the inter-action delay.

3. *Solving a linear (integer) program.* At each decision epoch t_k, perform $z^*_m(Q^{t_k})$ many matches for all $m \in \mathcal{M}_+$, where

$$\begin{aligned} z^*(Q^{t_k}) \in \quad & \arg\max \quad r \cdot z \\ & \text{s.t.} \qquad Mz \le Q^{t_k} \\ & \qquad\qquad z \in \mathbb{Z}^d_{\ge 0}, \end{aligned} \tag{3}$$

   where, we recall, $Q^{t_k}_i$ is the pre-match length of queue$-i$: the number of agents in queue$-i$ right before the matches are performed at time t_k.

Observe that immediately after a decision epoch t_k, no feasible matches remain to perform; otherwise one could increase the objective value in (3) by forming an additional match.

In our analysis, we will assume that immediately after the matches are performed, all remaining unmatched agents from queues $j \in \mathcal{Q}_+$ (under-demanded queues) are removed. This is done for mathematical exposition and without loss of generality; we will show that these removals are not necessary; see Proof of Theorem 3.1 in Appendix D. Arguably, removals are practically reasonable in order to prevent agents of these types from waiting indefinitely.

[2]That we remove the matches in $\mathcal{M}_0$ from the network is, in fact, necessary; see Remark 3.1.

Definition 3.1 (trivial networks)**.** *A matching network that satisfies* **GP** *is trivial if the general position gap equals the arrival probability of some agent type. That is, for some $i \in \mathcal{A}$,*

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

In trivial networks, as illustrated in Figure 4, it is possible to keep the regret small at all times (in particular, in terms of order, smaller than $\Omega(\epsilon^{-1})$).

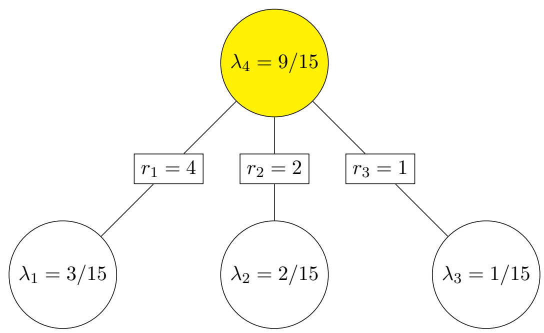

Figure 4: An example of a trivial network, where (SPP) has a unique optimal solution $z^* = (3/15, 2/15, 1/15)$ and $s^* = (0, 0, 0, 3/15)$ so that $\epsilon = z^*_3 = 1/15 = \lambda_3$. Since $\lambda_4 > \lambda_1 + \lambda_2 + \lambda_3$, queue$-4$ will grow with time regardless of the matching policy. After some initial time t_0, queue-4 will be non-empty with probability close to 1. In particular, we will be able to immediately match any arriving agents of type 1, 2, or 3. The regret is 0 at all large enough times t.

Theorem 3.1 (hindsight optimality)**.** *Assume that $\mathcal{G}$ satisfies* **GP***, and let ϵ be the* **GP** *gap. Then, there exists a matching policy D such that*

$$\mathcal{R}^{*,t} - \mathcal{R}^{D,t} \le \Gamma \epsilon^{-1} \text{ for all } t > 0, \qquad \text{(Upper bound)}$$

where $\Gamma > 0$ is a constant that may depend on n, d, M, and r (but not λ or ϵ). This performance is achieved by the exhaustive resolving policy with an inter-action delay $\tau = \lceil \kappa \epsilon^{-1} \rceil = \Theta(\epsilon^{-1})$, where $\kappa > 0$ is some constant that does not depend on ϵ.

If the network is non-trivial, any matching policy D has

$$\sup_{t > 0} (\mathcal{R}^{*,t} - \mathcal{R}^{D,t}) \ge \gamma \epsilon^{-1}, \qquad \text{(Lower bound)}$$

where $\gamma > 0$ is a constant that may depend on n, d, M, and r (but not λ or ϵ).

Our main theorem states that an inter-action delay proportional to ϵ^{-1} is sufficient to achieve the optimal regret scaling. By the lower bound result, a smaller τ cannot improve this achieved regret scaling. It can, however, make it worse; see Example 3.1. Picking τ larger, in terms of order, compromises the regret; for example with $\tau = \Theta(\epsilon^{-2})$, the regret scales with $\epsilon^{-2} \gg \epsilon^{-1}$. This is because just before a decision epoch, there are (of the order of) ϵ^{-2} unmatched agents waiting in queues. Thus, at that point in time the regret is of the order of ϵ^{-2}.

Queueing-intuition for the lower bound. The proof of the lower bound appears in Appendix E. We provide here some intuition using a simple example. Consider the network in Figure 5. Let us pretend that upon arrival, an agent type−2 is lost if it is not used to form a match with queue−1, and match−1 is performed otherwise. Then queue−1 behaves like a single-server queue with arrival rate λ_1, and service rate $\lambda_2 = \lambda_1 + \epsilon$; the utilization is $\rho = \lambda_1/(\lambda_1 + \epsilon)$. Then the stationary mean queue-length of queue−1 is given by

$$\frac{\rho}{1-\rho} = \frac{\lambda_1}{\epsilon} \sim \frac{1}{\epsilon}.$$

Thus, while the upper bound (SPP) makes queue−1 empty at all times, we will, in the stochastic system, have of the order of ϵ^{-1} unmatched type−1 agents, which will constitute an unrealized value of $\sim r_1/\epsilon$. The main challenge in formalizing this intuition is that not only the arrivals to queue−2 are not "lost" if not immediately matched, but also that we must allow the matching policy to be arbitrary.

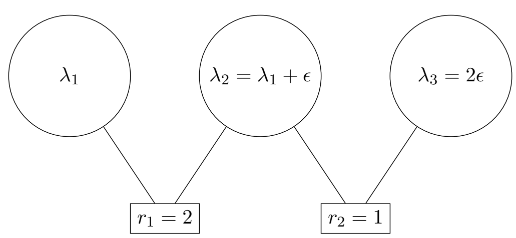

Figure 5: A simple network for the lower bound intuition.

3.1 Discussion

On the policy ingredients. The exhaustive resolving policy utilizes (SPP) to identify which matches to avoid and what delay to impose between decision epochs. In particular, our results require the knowledge of the parameters λ and r. Next, we discuss the importance of these ingredients under our resolving policy.

Remark 3.1 (pre-removal of redundant matches)**.** Avoiding matches in $\mathcal{M}_0$ is necessary for the resolving policy to achieve hindsight optimality. To see this consider the network in Figure 6.

Independent of the size of τ, the figure showcases the linear growth (in t) of the regret $\mathcal{R}^{*,t} - \mathcal{R}^{D,t}$. In this example, (SPP) has $z_4^* = 0$, but the static problem (3) uses it occasionally (even if not frequently). Regardless of the fixed τ, there is a positive probability (that decreases with τ, but is constant once τ is fixed) that both queues 4 and 5 will be non-empty at a decision epoch, where queues 3 and 6 will be empty. In such a case, our *exhaustive* resolving policy will perform match-4. This is a "mistake", and it will be repeated at a fixed frequency. ■

The next two examples illustrate the necessity of some delay between decision epochs under our resolving policy (regardless of how ties are broken).

Example 3.1 (the frequency of resolving in two-way networks)**.** As briefly discussed in the introduction, Figure 2 considers our resolving policy for a two-way network, and captures the regret for multiple values of the "batching" parameter $\tau \in \{5, 20, 100\}$. Even in this simple (two-way) network, τ cannot be too small; if it is too small, the performance of the resolving policy is suboptimal. ■

Example 3.2 (the necessity of some delay in multi-way networks)**.** In Figure 7, the tuple $\mathcal{G} = (M, \lambda, r)$ satisfies **GP**. Since match-1 has a relatively high value, it is important to utilize agent types 1, 2, 4, and 6 towards performing this match. Any greedy policy "fails", since agents of types 2, 4, and 6 (required to perform match-1) "disappear" before they can be used to perform match-1. For instance, since $\lambda_7 = 64\lambda \gg \lambda_6 = 32\lambda$, there will be (after some initial transient horizon) available agents waiting to be matched in queue-7, with high probability. Under any greedy policy, any arriving type-6 agent will then immediately be matched to an agent of type 7, and disappear. Our resolving policy with a suitable inter-action delay prevents this, and performs match-1 sufficiently many; see its constant regret in Figure 7(BOTTOM LEFT). In Figure 7(BOTTOM RIGHT), we can see that resolving too frequently results in a large regret. ■

We do not offer a precise recipe to pick τ. However, an initial pre-processing step based on simulations can help to fine tune this parameter; a simple heuristic would be to initialize τ to ϵ^{-1} and keep increasing it "slightly" as long as the regret grows. Note that such simulations, like the exhaustive resolving policy, rely on knowing the arrival probabilities and match values.

Further comments. In some applications, the main objective is to maximize the total number of matched agents, i.e., the value of a match equals the number of agent types participating in the match. Similar arguments to those in Example 3.2 imply that in multi-way networks, even such

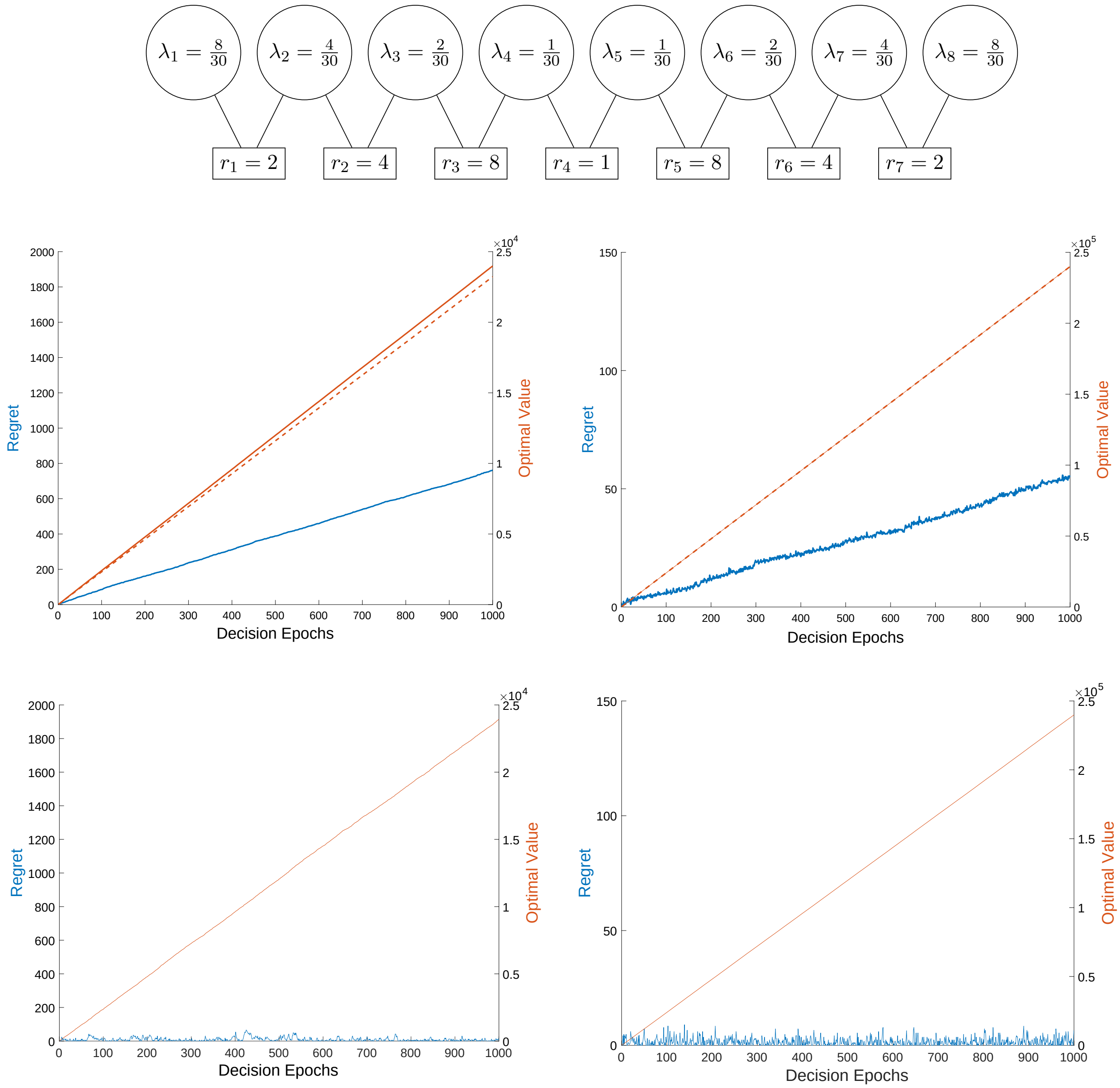

Figure 6: Resolving without removing all matches in $\mathcal{M}_0$ does not achieve hindsight optimality. The network in this figure exhibits a regret that grows linearly with time. (TOP LEFT) The performance of the exhaustive resolving policy without removing match-4 with $\tau = 20$. The solid line represents the optimal value of (SPP) (where the arrivals are replaced with their expectations) scaled with t, and the dashed line represents the optimal match value given the actual arrival realizations (not in expectation). (TOP RIGHT) The performance without removing match-4 with $\tau = 200$. The regret grows slower, but it nevertheless grows. (BOTTOM LEFT) The performance with removing match-4 and $\tau = 20$. (BOTTOM RIGHT) The performance with removing match-4 and $\tau = 200$.

a simple cardinality maximizing objective requires delaying match decisions to achieve hindsight optimality; this can be illustrated by extending the network in Figure 7 by adding new agent types with relatively large arrival probabilities to align match values with their cardinalities. Finally, in §5, we identify an alternative periodic clearing policy, which is also hindsight optimal for a large

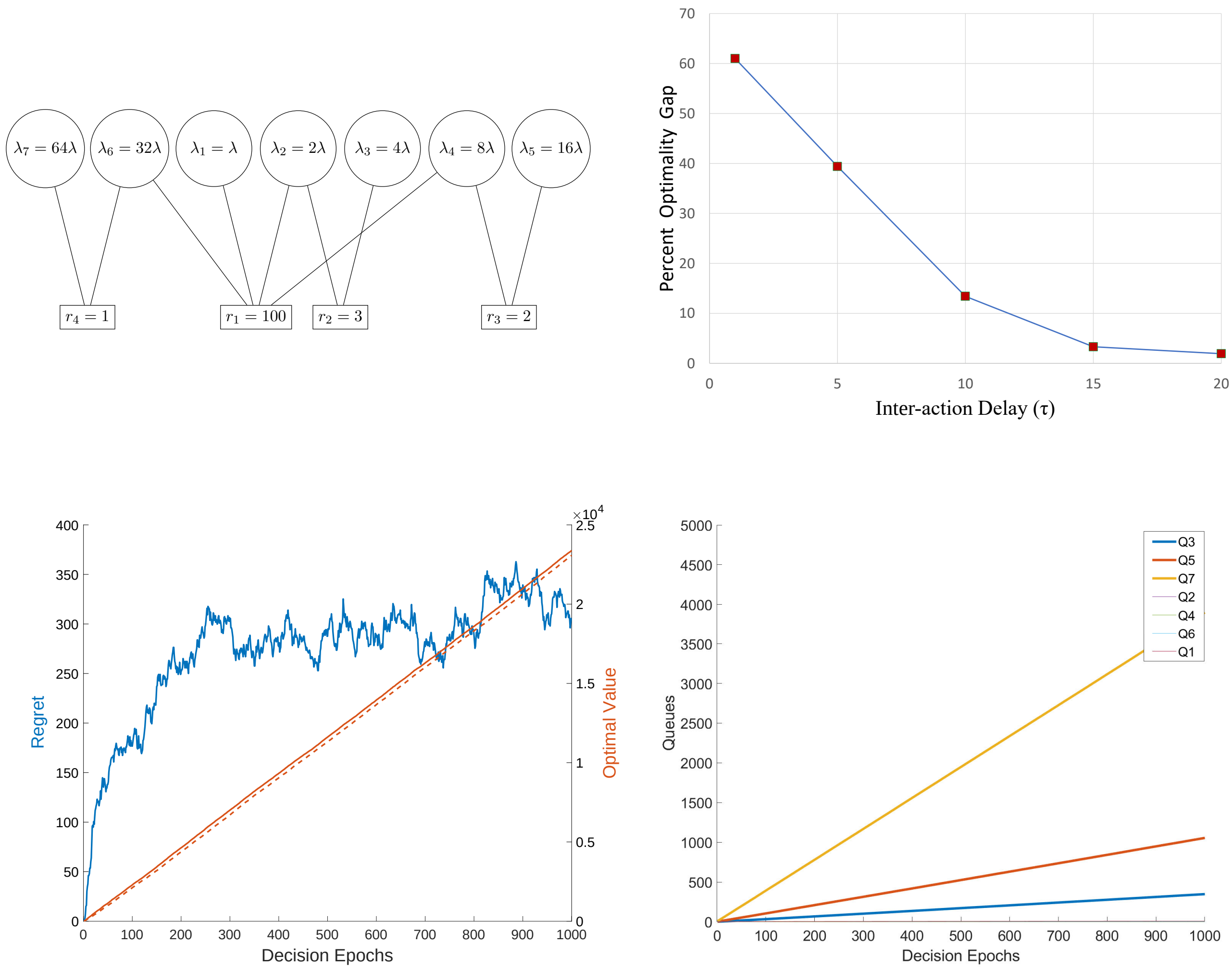

Figure 7: (TOP LEFT) A (multi-way) network, where λ is chosen so that $\sum_{i \in \mathcal{A}} \lambda_i = 1$. (TOP RIGHT) The percent optimality gap (regret) as a function of the inter-action delay τ. For each τ, the reported gap is an average of 1000 replications. With $\tau = 1$ (acting every period), the gap is as high as 60%; it decreases to less than 1.5% with a delay of $\tau = 20$. (BOTTOM LEFT) Hindsight optimality: the regret as a function of decision epochs with $\tau = 20$. Note that a regret of 300 corresponds to not performing match-1 three times *throughout* the horizon. (BOTTOM RIGHT) The queues of type $i \in \mathcal{Q}_+ = \{3, 5, 7\}$ grow linearly with time. All the queues in $\mathcal{Q}_0$ remain bounded in expectation, and these queues are not visible in this scale.

family of networks.

4 Upper bound: The regret of exhaustive resolving

In this section, we prove the first part of our main result Theorem 3.1, that is the exhaustive resolving policy achieves the desired regret $\mathcal{O}(\epsilon^{-1})$. We first present in Lemma 4.1 a sufficient condition for a matching policy to be hindsight optimal. Next, we present structural properties of the optimal solution of (SPP), which will be useful to analyze the dynamic system including proving Lemma 4.1. Finally, the proof uses Lyapunov arguments to establish that the conditions of Lemma 4.1 hold.

4.1 Optimality test

The following lemma provides a sufficient condition for hindsight optimality. Essentially, the non-degeneracy provided by **GP** guarantees that any matching policy, whose set of bounded queues coincides with the set of over-demanded queues (the set $\mathcal{Q}_0$) is hindsight optimal.

Lemma 4.1 (optimality test)**.** *Suppose that* **GP** *holds. Let* (z^*, s^*) *be the unique non-degenerate optimal solution of* (SPP)*. Then a matching policy* D *that*

(i) does not reject any agents of type $i \in \mathcal{Q}_0$*,*

(ii) does not perform any matches in $\mathcal{M}_0$*, i.e.,* $D^t_m = 0$ *for all* $m \in \mathcal{M}_0$ *and for all* $t > 0$*, and*

(iii) has $\mathbb{E}^D[Q^t_i] = \mathcal{O}(\epsilon^{-1})$ *for all* $i \in \mathcal{Q}_0$ *and for all* $t > 0$*,*

is hindsight optimal, and $\mathcal{R}^{*,t} - \mathcal{R}^{D,t} = \mathcal{O}(\epsilon^{-1})$ *for all* $t > 0$*.*

Lemma 4.1 translates Theorem 3.1 to the constancy—uniformly in t—of the queues in the set $\mathcal{Q}_0$. Indeed, if the policy avoids redundant matches and keeps the expected lengths of over-demanded queues sufficiently "small" at all times, then hindsight optimality is achieved.

4.2 The structure of the optimal solution of (SPP).

The optimality test utilizes properties of the optimal solution of (SPP), which will be key to our analysis for the dynamic system. Without loss of generality, assume that $\mathcal{M}_+ = \{1, 2, ..., d - \varrho\}$ and $\mathcal{Q}_+ = \{d - \varrho + 1, d - \varrho + 2, ..., n\}$, where we let $\varrho := |\mathcal{M}_0|$. Then the optimal basis matrix takes the form

$$\mathcal{B} = \begin{bmatrix} M^0 & \mathbf{0} \\ M^+ & I \end{bmatrix},$$

where M^0 has the rows of M corresponding to the queues in $\mathcal{Q}_0$, M^+ has the remaining $n-d+\varrho$ rows, and $\mathcal{B}$ has the columns corresponding to $\mathcal{M}_+$ and $\mathcal{Q}_+$ in order; I is an $(n-d+\varrho)\times(n-d+\varrho)$ identity matrix, and $\mathbf{0}$ is a $(d-\varrho)\times(n-d+\varrho)$ zero matrix. Being the basis matrix, $\mathcal{B}$ is invertible, and $Y=\mathcal{B}^{-1}$ has the following form

$$\mathcal{B}^{-1}=Y:=\begin{bmatrix} Y^0 & \mathbf{0} \\ Y^+ & I \end{bmatrix},$$

where $[Y^0,\ \mathbf{0}]$ is a $(d-\varrho)\times n$ matrix and $[Y^+,\ I]$ is an $(n-d+\varrho)\times n$ matrix, where

1. m^{th} row of $[Y^0,\ 0]$ is y^m for each $m\in\mathcal{M}_+$, and
2. j^{th} row of $[Y^+,\ I]$ is $y^{d-\varrho+j}$ for each $d-\varrho+j\in\mathcal{Q}_+$.

In turn, the optimal solution of (SPP) can be written as

$$\begin{bmatrix} z^*_{\mathcal{M}_+} \\ s^*_{\mathcal{Q}_+} \end{bmatrix}=\mathcal{B}^{-1}\lambda=Y\lambda,$$

which implies

$$z^*_m=y^m\lambda>0 \text{ for all } m\in\mathcal{M}_+, \text{ and } s^*_j=y^j\lambda>0 \text{ for all } j\in\mathcal{Q}_+, \tag{4}$$

where strict inequalities follow from the non-degeneracy of (z^*,s^*) under **GP**. Finally, since $\mathcal{G}$ is a finite matching network, i.e., $n<\infty$, we must have $\max_{i,j\in[1,n]}|Y_{i,j}|\le\omega$, for some constant $\omega>0$, where ω may depend on n and M. The matrix Y (and in turn, the vectors y^m's and y^j's) can be explicitly constructed for a special family of networks; see §5.

Non-degeneracy implies (e.g., see (Bertsimas and Tsitsiklis, 1997, Section 5.1)), that the same basis remains optimal for any $\tilde{\lambda}>0$ such that $\tilde{\lambda}=\lambda+\zeta$, where $\|\zeta\|_\infty\le\zeta_0$ for all sufficiently small $\zeta_0>0$. The dual of (SPP) will also be useful in what follows. It readily follows that under **GP**, $\theta_i:=(\sum_{m\in\mathcal{M}_+}r_my^m)_i\ge 0,\ i\in\mathcal{A}$, are the corresponding optimal dual variables. In particular, uniqueness of (z^*,s^*) implies $\theta_i>0$ for all $i\in\mathcal{Q}_0$.

4.3 Lyapunov arguments for analyzing the exhaustive resolving policy

Since the first two conditions of Lemma 4.1 are clearly satisfied under the exhaustive resolving policy, our main focus in this section to provide tools to analyze the third condition. Intuitively, we want to show that whenever the queue-length of an over-demanded queue hits a certain threshold, the exhaustive resolving policy is able to "pull back" the length below the threshold in the next decision epoch, as the non-degeneracy provided by **GP** allows the exhaustive resolving policy to approximately "mimic" the optimal solution of (SPP).

Drift arguments, as the one we are going to use, are common in the study of stochastic networks and queues. The following result (e.g., see (Glynn and Zeevi, 2008, Corollary 4)) is useful to bound stationary expectations of Markov processes.

Lemma 4.2. *Let $X = (X^t : t \geq 0)$ be a discrete-time $\mathcal{S}$-valued Markov chain with transition kernel P, and suppose $f : \mathcal{S} \to \mathbb{R}$ is non-negative. If there exists a non-negative function $g : \mathcal{S} \to \mathbb{R}$ and a constant c for which*

$$\int_S P(x, dy) g(y) - g(x) \leq -f(x) + c \text{ for all } x \in \mathcal{S}, \tag{5}$$

then

$$\int_S \pi(dx) f(x) \leq c, \tag{6}$$

for any stationary distribution π of X.

The challenge lies in identifying a suitable Lyapunov function g—a "norm" of the total process—that decreases when the queues in $\mathcal{Q}_0$ are large. This is non-trivial, and relies in subtle ways on the network structure and the detailed analysis of the optimal solution of (SPP). As we will formulate our Lyapunov function next, the construction is based on the dual of (SPP), in particular our Lyapunov function originates from a weighted sum of the queue-lengths, where weights are determined by the dual variables.

Minimal Markov chain notation is needed before we proceed. Under the exhaustive resolving policy, the process $(Q^{t_k},\ k \in \mathbb{N})$ is clearly a Markov chain. We let $\mathbb{P}_q\{\cdot\}$ be the probability law of this Markov chain initialized at $q \in \mathbb{Z}^n_{\geq 0}$, and we write $\mathbb{E}_q[\cdot]$ for the corresponding expectation.

Since the policy is applied separately to each connected component of the network (recall that all matches in $\mathcal{M}_0$ are removed from the network), without loss of generality, we assume that there is a single component, i.e., $\mathcal{M}_0 = \emptyset$. Recall that at each decision period $t_k = k\tau$, $k \in \mathbb{N}$, the exhaustive resolving policy solves the following linear integer program

$$\begin{aligned}
&\max\ r \cdot z \\
&\text{s.t. }\ Mz + s = Q^{t_k} \\
&\qquad z \in \mathbb{Z}^d_{\geq 0}, s \in \mathbb{Z}^n_{\geq 0},
\end{aligned}$$

where Q^{t_k} is the pre-match queue-length vector. Since Y is invertible and $y^j M = 0$ for all $j \in \mathcal{Q}_+$, this linear program can be rewritten as

$$\begin{aligned}
\max \quad & r \cdot z \\
\text{s.t.} \quad & y^m M z + y^m s = y^m Q^{t_k} \text{ for all } m \in \mathcal{M}_+ \\
& y^j s = y^j Q^{t_k} \text{ for all } j \in \mathcal{Q}_+ \\
& z \in \mathbb{Z}^d_{\geq 0}, s \in \mathbb{Z}^n_{\geq 0}.
\end{aligned}$$

Recalling that $y^m M z = z_m$ for all $m \in \mathcal{M}_+$, we have $z_m = y^m(Q^{t_k} - s)$ for all $m \in \mathcal{M}_+$. Hence, the linear program above can be rewritten as

$$\begin{aligned}
&\max \sum_{m \in \mathcal{M}_+} r_m y^m (Q^{t_k} - s) \\
&\text{s.t. }\ z_m + y^m s = y^m Q^{t_k} \text{ for all } m \in \mathcal{M}_+ \\
&\qquad y^j s = y^j Q^{t_k} \text{ for all } j \in \mathcal{Q}_+ \\
&\qquad z \in \mathbb{Z}^d_{\geq 0}, s \in \mathbb{Z}^n_{\geq 0}.
\end{aligned}$$

Finally, since $(y^m)_j = 0$ for all $m \in \mathcal{M}_+$ and for all $j \in \mathcal{Q}_+$, we obtain, with $u := Q^{t_k}$, the following equivalent problem (in terms of optimizers)

$$\begin{aligned}
h^*(u) := \quad \min \quad & \textstyle\sum_{i \in \mathcal{Q}_0} \sum_{m \in \mathcal{M}_+} (r_m y^m)_i s_i \\
\text{s.t.} \quad & z_m + y^m s = y^m u \text{ for all } m \in \mathcal{M}_+ \\
& y^j s = y^j u \text{ for all } j \in \mathcal{Q}_+ \\
& z \in \mathbb{Z}^d_{\geq 0}, s \in \mathbb{Z}^n_{\geq 0}.
\end{aligned} \tag{7}$$

For ease of exposition, without loss of generality, we initialize the pre-match queue-length vector at $q \in \mathbb{Z}_{\geq 0}^n$, and let $q^+ \in \mathbb{Z}_{\geq 0}^n$ be the post-match queue-length vector right after the exhaustive resolving policy is executed at time 0. Thus, with this notation we have $Q^\tau = q^+ + A^\tau$, i.e., Q^τ is the pre-match queue-length vector at time τ.

The following proposition provides bounds on the drift, which will allow us to apply the optimality test (Lemma 4.1) and complete the proof of the upper bound. The proof is given in Appendix B.

Proposition 4.1. *Take $\tau = \lceil \kappa\epsilon^{-1} \rceil$ for some constant $\kappa > 0$ (not dependent on ϵ). Then, the process $h^*(Q^{t_k})$, with $h^*(\cdot)$ as in (7), decreases in expectation:*

$$\mathbb{E}_q\left[h^*(Q^\tau) - h^*(q)\right] \leq -\gamma + \frac{\Gamma}{\epsilon}\mathbb{1}_{\{h^*(q)\leq B\}}, \tag{8}$$

where $B, \gamma, \Gamma > 0$ do not depend on ϵ. Consequently, there exist constants $c_1, c_2 > 0$, not dependent on ϵ, such that the process $\mathcal{L}(Q^{t_k}) := e^{h^*(Q^{t_k})}$ also decreases in expectation:

$$\mathbb{E}_q[\mathcal{L}(Q^\tau) - \mathcal{L}(q)] \leq -\frac{\gamma}{2}\mathcal{L}(q) + c_1 e^{c_2\tau}\mathbb{1}_{\{h^*(q)\leq B\}}. \tag{9}$$

Observe that inequality (9) follows from a standard mechanism, which derives an exponential Lyapunov function from a given linear one. Note that Lemma (4.1) immediately implies that under the Markov chain's unique stationary distribution, which we denote by π, we have

$$\mathbb{E}_\pi[\mathcal{L}(Q^0)] \leq \frac{2c_1}{\gamma}e^{c_2\tau}, \tag{10}$$

where $Q^0 \sim \pi$. Since $\tau = \lceil \kappa\epsilon^{-1} \rceil$, by Jensen's inequality we have

$$\mathbb{E}_\pi[h^*(Q^0)] = \mathcal{O}(\epsilon^{-1}). \tag{11}$$

The reason behind considering an exponential Lyapunov function is to be able to use geometric recurrence of the process $(Q^{t_k}, k \in \mathbb{N})$, which is crucial to prove that $\mathbb{E}[\sum_{i\in\mathcal{Q}_0} Q_i^t] = \mathcal{O}(\epsilon^{-1})$ for all $t > 0$, not only in the stationary distribution. The proof of the upper bound in Theorem 3.1 can be found in Appendix D.

5 (SPP)-acyclicity and the general position gap

In this section, we focus on a special family of matching networks to extend some of our main results, as well as providing more intuition about the general position gap ϵ.

Definition 5.1 ((SPP)-acyclic networks)**.** *Suppose that $\mathcal{G}$ satisfies* **GP** *and let (z^*, s^*) be the unique optimal solution of* (SPP)*. The* (SPP)*-residual graph is obtained by removing all redundant matches $m \in \mathcal{M}_0$ (with $z^*_m = 0$) from $\mathcal{G}$. We say that $\mathcal{G}$ is* (SPP)*-acyclic, if the* (SPP)*-residual graph is acyclic.*

If $\mathcal{G}$ (the bipartite graph representation of the hypergraph) is acyclic itself, then $\mathcal{G}$ is trivially (SPP)-acyclic. More interestingly, this is also the case if $\mathcal{G}$ itself is a simple bipartite graph (where only even cycles are allowed) with two-way matches only.

Lemma 5.1 (two-way two-sided networks)**.** *Suppose that $\mathcal{G}$ satisfies* **GP***. Let (z^*, s^*) be the unique non-degenerate optimal solution of* (SPP)*. If $|\mathcal{A}(m)| = 2$ for all $m \in \mathcal{M}$ (all matches are two-way), and $\mathcal{G}$ is bipartite (any cycle in $\mathcal{G}$ contains an even number of matches), then $\mathcal{G}$ is* (SPP)*-acyclic.*

It is important to notice that other than the network structure, (SPP)-acyclicity also depends on the optimal solution of (SPP). In turn, whether this notion of acyclicity holds or not depends not only on the matching matrix M, but also on the arrival probability vector λ and the match value vector r. Because of this dependence, one should not expect other sufficient conditions as simple and insightful as the one in Lemma 5.1.

5.1 The general position gap in (SPP)-acyclic networks

As discussed in Section 2, the general position gap can be intuitively thought of as a measure of slack in the network. In (SPP)-acyclic networks, as the next lemma shows, this slack can be viewed as an imbalance between arrival probabilities.

Lemma 5.2. *Assume that $\mathcal{G}$ is* (SPP)*-acyclic. If for every two subsets $\mathcal{A}_1 \neq \mathcal{A}_2 \subseteq \mathcal{A}$, we have*

$$\sum_{i\in\mathcal{A}_1} \lambda_i \neq \sum_{j\in\mathcal{A}_2} \lambda_j, \tag{12}$$

then (SPP) *has a non-degenerate optimal basic feasible solution*[3]*.*

[3]The opposite is not generally true.

If the arrival rates are drawn from a continuous distribution, then (12) holds almost surely. Intuitively, (**GP**) is then likely to hold in any practical setting.

We can be more precise compared to Lemma 5.2 regarding mapping the general position gap to an intuitive notion of slack. Recall that the optimal solution of (SPP) is given simply in terms of the inverse of the basis matrix as in (4). Therefore, the first step in that direction is to explicitly construct the inverse matrix Y of the optimal basis matrix. For some intuition of this construction when $\mathcal{G}$ is (SPP)-acyclic, consider the network in Figure 2. Under the optimal solution, all slack variables are 0, except $s_5^* > 0$. Then it must be that $z_1^* = \lambda_1$ (all type-1 agents are matched). Then match-2 uses the leftovers of type-2 agents, and $z_2^* = \lambda_2 - z_1^* = \lambda_2 - \lambda_1$; match$-3$ uses the leftovers (those that are not used towards match-2) of type-3 agents, and $z_3^* = \lambda_3 - z_2^* = \lambda_3 - \lambda_2 + \lambda_1$. Defining row vectors $y^1 = [1, 0, 0, 0, 0]$, $y^2 = [-1, 1, 0, 0, 0]$, $y^3 = [1, -1, 1, 0, 0]$, and $y^4 = [-1, 1, -1, 1, 0]$, we have the representation $z_m^* = y^m \lambda$ for all $m \in \mathcal{M}_+ = \{1, 2, 3, 4\}$. Similarly, we have $s_5^* = \lambda_5 - z_4^* = y^5 \lambda$, where $y^5 = [1, -1, 1, -1, 1]$ (note that $\mathcal{Q}_+ = \{5\}$). This demonstrates an instance for the general construction of the optimal solution of (SPP).

Theorem 5.1 (explicit optimal solution of (SPP))**.** *Assume that* **GP** *holds and* $\mathcal{G}$ *is* (SPP)*-acyclic. Let* (z^*, s^*) *be the unique non-degenerate optimal solution of* (SPP) *with* $\mathcal{M}_+ = \{m \in \mathcal{M} : z_m^* > 0\}$ *and* $\mathcal{Q}_+ = \{j \in \mathcal{A} : s_j^* > 0\}$*. Then there exist* $|\mathcal{M}_+|$ *vectors* $y^m \in \{-1, 0, 1\}^n$ *and* $|\mathcal{Q}_+|$ *vectors* $y^j \in \{-1, 0, 1\}^n$ *such that*

$$z_m^*(\lambda) := z_m^* = y^m \lambda > 0 \textit{ for all } m \in \mathcal{M}_+, \textit{ and } s_j^*(\lambda) := s_j^* = y^j \lambda > 0 \textit{ for all } j \in \mathcal{Q}_+.$$

Any right-hand side $\lambda > 0$ *with* $y^l \lambda > 0$ *for all* $l \in \mathcal{M}_+ \cup \mathcal{Q}_+$*, induces the optimal solution* $(z^*(\lambda), s^*(\lambda))$*.*

Recall that, also in general matching networks (not necessarily (SPP)-acyclic), the optimal solution of (SPP) takes the form as in Theorem 5.1, where y^m's and y^j's are the rows of the inverse of the optimal basis matrix; see §4.2. What is new here is that when $\mathcal{G}$ is (SPP)-acyclic, the matrix Y can be constructed explicitly; all entries of Y are either -1, 0, or -1. We prove Theorem 5.1 and provide the explicit construction of Y in Appendix C.

Without the uniqueness requirement, Lemma 5.2 has a sufficient condition for **GP** that requires the sum of total arrival probabilities—*for any* two subsets $\mathcal{A}_1$ and $\mathcal{A}_2$—to be different. But it should be clear that this requirement is too stringent. For instance in Figure 2, we would still have **GP** if $\lambda_2 = \lambda_4 = 0.2$, but that would clearly violate the requirement of Lemma 5.2. In other words, it is clear that the sets $\mathcal{A}_1$ and $\mathcal{A}_2$ need not be arbitrary.

Let us revisit the network in Figure 2. The "capacity" available to agent type-1 is λ_2. If $\lambda_1 > \lambda_2$, then queue-1 must grow with time under any matching policy. Similarly, the capacity available for agent types 2 and 4 combined is at most $\lambda_1 + \lambda_3 + \lambda_5$; the capacity slack for these two types is then $\lambda_1 + \lambda_3 + \lambda_5 - (\lambda_2 + \lambda_4)$. More generally, for each subset of agent types $S \subset \mathcal{A}$, we can define $\mathcal{N}(S)$ to be the set of agent types participating in a match with some agent type $i \in S$, and so that $\mathcal{N}(S) \cap S = \emptyset$. The capacity slack for S is then $\epsilon'(S) := |\sum_{i\in\mathcal{N}(S)} \lambda_i - \sum_{j\in S} \lambda_j|$, and the network *capacity slack* is the minimum over all subsets:

$$\epsilon' := \min_{S\subset\mathcal{A}} \epsilon'(S) = \min_{S\subseteq\mathcal{A}} \left| \sum_{i\in\mathcal{N}(S)} \lambda_i - \sum_{j\in S} \lambda_j \right|$$

This would be an intuitive notion of capacity slack, but it is still too stringent. It turns out that we do not need to consider all subsets $\mathcal{S}$ as we do in defining ϵ'. The explicit construction of the inverse matrix Y identifies for us the "relevant" subsets. Indeed, take the vector y^m as in Theorem 5.1 for some $m \in \mathcal{M}_+$. Let

$$\mathcal{A}^+(y^m) := \{i \in \mathcal{A} : (y^m)_i = 1\} \quad \text{and} \quad \mathcal{A}^-(y^m) := \{i \in \mathcal{A} : (y^m)_i = -1\}.$$

Then we have

$$y^m\lambda = \sum_{i\in\mathcal{A}^+(y^m)} \lambda_i - \sum_{i\in\mathcal{A}^-(y^m)} \lambda_i,$$

and

$$\epsilon = \min_{\ell\in\mathcal{M}_+\cup\mathcal{Q}_+} \left(\sum_{i\in\mathcal{A}^+(y^\ell)} \lambda_i - \sum_{i\in\mathcal{A}^-(y^\ell)} \lambda_i \right).$$

In turn, for (SPP)-acyclic matching networks, we can see the general position gap as a measure of capacity slack, where for each $\ell \in \mathcal{M}_+ \cup \mathcal{Q}_+$, it identifies, via y^ℓ, a subset of agent types (those in $\mathcal{A}^-(y^\ell)$) as "customers", and a subset of agent types (those in $\mathcal{A}^+(y^\ell)$) as the "servers" who serve these agent types. It then compares the total capacity to the total input.

Once ϵ is understood as a capacity slack, it is intuitively clear that achievable regret should depend on this measure. Having a large capacity slack increases the decision maker's ability to control the dynamic system and perform matches that are aligned with the deterministic counterpart (SPP). Theorem 3.1 establishes that it is feasible to achieve a regret of the order of ϵ^{-1}, and a smaller regret is not attainable.

The following remark shows that the explicit construction of the inverse matrix Y when $\mathcal{G}$ is (SPP)-acyclic allows us to give a more explicit characterization of the inter-action delay τ in Theorem 3.1 by showing that τ linearly depends on the number of agent types n. The proof reveals how the negative drift (γ) in Proposition 4.1 depends on Y, and in turn this dependence determines τ.

Remark 5.1. An immediate extension of Theorem 3.1 when $\mathcal{G}$ is (SPP)-acyclic is that the exhaustive resolving policy with $\kappa = \Theta(n)$ (so that $\tau = \Theta(n\epsilon^{-1})$) is hindsight optimal. This directly follows from the proof of Propositon 4.1 by noticing that $\omega = 1$, since any entry of the surplus vector y^j is in $\{-1, 0, 1\}$ for all $l \in \mathcal{M}_+ \cup \mathcal{Q}_+$, where ω is an upper bound for the maximum entry in Y, i.e., $\max_{i,j\in[1,n]} |Y_{i,j}| \leq \omega$.

5.2 An alternative hindsight optimal policy

Note that the match value vector plays a key role in determining the basic feasible activities under (SPP), as well as the match decisions that the exhaustive resolving policy makes. It is a natural question to ask whether good policies must further take into account the match value vector when determining which matches to perform. We are now ready to propose an alternative policy to the exhaustive resolving policy, which is also hindsight optimal when $\mathcal{G}$ is (SPP)-acyclic, where this policy does not take into account the match value vector while making match decisions. Consider the following periodic matching policy D', which acts exactly the same as the exhaustive resolving policy, except at each decision epoch t_k, we perform $z^*_m(Q^{t_k})$ many matches for all $m \in \mathcal{M}_+$, where

$$
\begin{aligned}
z^*(Q^{t_k}) \in \quad & \arg\min && \textstyle\sum_{i\in\mathcal{Q}_0} Q_i^{t_k^+} \\
& \text{s.t.} && Mz \leq Q^{t_k} \\
& && z \in \mathbb{Z}^d_{\geq 0},
\end{aligned}
\tag{13}
$$

where $Q^{t_k^+} \in \mathbb{Z}^n_{\geq 0}$ is the post-match queue-length vector right after the policy is executed at time t_k. That is, we minimize the number of agents waiting in over-demanded queues at each decision epoch.

Theorem 5.2. *Let $\mathcal{G}$ be an (SPP)-acyclic network that satisfies* **GP**, *and let ϵ be the* **GP** *gap. Then D' is hindsight optimal with the inter-action delay $\tau = \lceil \kappa\epsilon^{-1} \rceil = \Theta(\epsilon^{-1})$, where $\kappa > 0$ is some* *constant that does not depend on ϵ:*

$$\mathcal{R}^{*,t} - \mathcal{R}^{D',t} \leq \Gamma\epsilon^{-1} \quad \textit{for all } t > 0,$$

where $\Gamma > 0$ is a constant that may depend on n, d, M, and r (but not λ or ϵ).

The proof depends in explicit ways on the acyclicity; see Appendix C. We do not know if this is true for cyclic networks, where the main challenge is that we do not know how to explicitly construct the inverse matrix Y of the optimal basis matrix.

6 Delay costs

The problem of minimizing delay penalties/holding costs has been studied in earlier papers, e.g., see Gurvich and Ward (2014); Bušic and Meyn (2014). This is a complex question in general, but our results have some immediate implications on optimal delay cost scaling.

Suppose at the end of each period (after observing an arrival and possibly performing matches), we incur a delay cost c_i per type$-i$ agent in the system. Then the expected *total delay cost* by time t under a matching policy D is given by

$$\mathcal{H}^{D,t} := \mathbb{E}^D \left[\sum_{u=1}^{t} c \cdot Q^{u^+} \right].$$

The minimal delay cost for *fixed* t, is then $\mathcal{H}^{*,t} := \min_{D \in \Pi} \mathcal{H}^{D,t}$. Given delay costs c_i's, define $r = cM$; r_m is an "indirect" value per match$-m$. Note that each time that we perform match$-m$ once, the total delay cost decreases by $r_m = \sum_{i \in \mathcal{A}(m)} c_i$. With this notation, let us rewrite $\mathcal{H}^{D,t}$ as follows:

$$\begin{aligned}
\mathcal{H}^{D,t} &= \mathbb{E}^D \left[\sum_{u=1}^{t} c \cdot Q^{u^+} \right] = \mathbb{E}^D \left[\sum_{u=1}^{t} c \cdot A^u - c \cdot M D^u \right] = \mathbb{E} \left[\sum_{u=1}^{t} c \cdot A^u \right] - \mathbb{E} \left[\sum_{u=1}^{t} r \cdot D^u \right] \\
&= \mathbb{E} \left[\sum_{u=1}^{t} c \cdot A^u \right] - \sum_{u=1}^{t} \mathcal{R}^{D,u}.
\end{aligned}$$

In turn,

$$\mathcal{H}^{*,t} = \mathbb{E}^D \left[\sum_{u=1}^{t} c \cdot A^u \right] - \max_{D \in \Pi} \sum_{u=1}^{t} \mathcal{R}^{D,u},$$

and

$$\begin{aligned}\mathcal{H}^{D,t} - \mathcal{H}^{*,t} &= \max_{\pi \in \Pi} \sum_{u=1}^{t} \mathcal{R}^{\pi,u} - \sum_{u=1}^{t} \mathcal{R}^{D,u} \\ &\le \sum_{u=1}^{t} \mathcal{R}^{*,u} - \sum_{u=1}^{t} \mathcal{R}^{D,u}\end{aligned}$$

Under **GP**, our resolving policy achieves $\mathcal{R}^{*,t} - \mathcal{R}^{D,t} = \mathcal{O}(\epsilon^{-1})$ for all $t > 0$, so that

$$\mathcal{H}^{D,t} = \mathcal{H}^{*,t} + \mathcal{O}\left(t\epsilon^{-1}\right) \text{ for all } t > 0,$$

or, in terms of time-average delay cost, we have

$$\frac{1}{t}\mathcal{H}^{D,t} = \frac{1}{t}\mathcal{H}^{*,t} + \mathcal{O}(\epsilon^{-1}).$$

In the proof of the lower bound in Theorem 3.1 (see Appendix E), we show that under any matching policy, for any t_0 such that $\sum_{i \in \mathcal{Q}_0} Q_i^{t_0} \le \epsilon^{-1}$, there exists some constant $B > 0$ (that does not depend on ϵ) such that $\sum_{i \in \mathcal{Q}_0} \mathbb{E}[Q_i^{t_0 + B\epsilon^{-2}}] = \Omega(\epsilon^{-1})$. Because of this "constant shift", the set of all times when the expected sum of lengths of over-demanded queues is $\Omega(\epsilon^{-1})$ has a positive density, i.e.,

$$\liminf_{T \to \infty} \frac{\sum_{t=1}^{T} \mathbb{1}\{\mathbb{E}[\sum_{i \in \mathcal{Q}_0} Q_i^t] = \Omega(\epsilon^{-1})\}}{T} > 0.$$

In turn, it must be the case that $\mathcal{H}^{*,t} = \Omega(t\epsilon^{-1})$. We conclude then that the exhaustive resolving policy achieves the optimal delay scaling.

Allowing objectives that combine both match value and delay cost is an interesting but non-trivial research direction. Given network parameters c, r and M, consider (SPP) twice, once with r and once with $r' = cM$ (the match value maximization reformulation of the delay cost minimization). If these two instances have the same optimal basis, then it follows—from Theorem 3.1 and the delay cost derivations above—that our resolving policy achieves ϵ^{-1} all time regret for the total (match value minus scaled delay cost) objective $\mathcal{R}^{D,t} - t^{-1}\mathcal{H}^{D,t}$.

If the two bases are different, however, a possible conflict arises between match value maximization and delay cost minimization. Whether hindsight optimality is attainable in this setting and, if yes, whether it is achievable by simple policies is a worthy goal for future work.

7 Concluding remarks

The problem of dynamically allocating resources to incoming requests is central to operations research. In this paper, we seek to contribute to the study of those special settings, where requests have a dual role as demand and capacity. Our results speak to the tension between short- and long-term value maximization. We characterize networks, where maximal values can be achieved in long-term without sacrificing maximal values in short-term. We prescribe an appealingly simple dynamic matching policy that achieves this desired balance. We find that the best optimality gap that can be achieved simultaneously at all times, is inversely proportional to the general position gap ϵ. The proposed periodic resolving policy achieves this optimality gap, where the delay between consecutive decision periods is of the order of ϵ^{-1}. The general position gap in acyclic networks can be interpreted as an inherent thickness or capacity slack in the network.

This work raises several research directions. One direction is allowing objectives that combine both value and holding costs. Another direction is incorporating agents' departures. The tension between value and delay is endogenized when agents depart (abandon) without being matched. Without departures, delaying actions increases the collected value. With departures, this is no longer the case. The upper bound—given by infinitely patient agents and a decision maker that waits until the end of the horizon—is not generally achievable.

This paper reveals the importance of the general position gap in the study of departures. Since over-demanded queue lengths are of the order of ϵ^{-1} (so are their corresponding waiting times), if the patience is of the order of magnitude longer than this, the results should not change. In other words, the smaller the general position gap, the more patient we need agents to be in order to achieve hindsight optimality.

References

I. Adan, I. Kleiner, R. Righter, and G. Weiss. FCFS parallel service systems and matching models. *Performance Evaluation*, 127:253–272, 2018.

P. Afeche, R. Caldentey, and V. Gupta. On the optimal design of a bipartite matching queueing system. *Operations Research*, 2021.

R.K. Ahuja, T.L. Magnanti, and J.B. Orlin. *Network Flows: Theory, Algorithms, and Applications*. Prentice–Hall, Englewood Cliffs, NJ, 1993.

M. Akbarpour, S. Li, and S.O. Gharan. Thickness and information in dynamic matching markets. *Journal of Political Economy*, 128(3):783–815, 2020.

R. Anderson, I. Ashlagi, D. Gamarnik, and Y. Kanoria. Efficient dynamic barter exchange. *Operations Research*, 65(6):1446–1459, 2017.

A. Aouad and O. Saritac. Dynamic stochastic matching under limited time. *Working Paper*, 2020. London Business School.

N. Arnosti and P. Shi. How (not) to allocate affordable housing. In *AEA Papers and Proceedings*, volume 109, pages 204–08, 2019.

I. Ashlagi, A. Nikzad, and P. Strack. Matching in dynamic imbalanced markets. *Available at SSRN 3251632*, 2018.

I. Ashlagi, M. Burq, C. Dutta, P. Jaillet, A. Saberi, and C. Sholley. Edge weighted online windowed matching. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 729–742. ACM, 2019a.

I. Ashlagi, M. Burq, P. Jaillet, and V. Manshadi. On matching and thickness in heterogeneous dynamic markets. *Operations Research*, 67(4):927–949, 2019b.

M. Baccara, S. Lee, and L. Yariv. Optimal dynamic matching. *Theoretical Economics*, 15(3): 1221–1278, 2020.

Dimitris Bertsimas and John N Tsitsiklis. *Introduction to linear optimization*, volume Fourth Printing. Athena scientific Belmont, MA, 1997.

J.H. Blanchet, M.I. Reiman, V. Shah, and L.M. Wein. Asymptotically optimal control of a centralized dynamic matching market with general utilities. *preprint arXiv:2002.03205*, 2020.

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

A. Bušic and S. Meyn. Optimization of dynamic matching models. *arXiv preprint arXiv:1411.1044*, 2014.

R. Caldentey, E.H. Kaplan, and G. Weiss. FCFS infinite bipartite matching of servers and customers. *Advances in Applied Probability*, 41(3):695–730, 2009.

N. Collina, N. Immorlica, K. Leyton-Brown, B. Lucier, and N. Newman. Dynamic weighted matching with heterogeneous arrival and departure rates. In *International Conference on Web and Internet Economics*, pages 17–30. Springer, 2020.

J.P. Dickerson, A.D. Procaccia, and T. Sandholm. Dynamic matching via weighted myopia with application to kidney exchange. In *Proceedings of the Twenty-Sixth AAAI Conference on Artificial Intelligence*, 2012.

J.P. Dickerson, D.F. Manlove, B. Plaut, T. Sandholm, and J. Trimble. Position-indexed formulations for kidney exchange. In *Proceedings of the 2016 ACM Conference on Economics and Computation*, pages 25–42, 2016.

M.D. Donsker. An invariance principle for certain probability limit theorems. In Doob J.L., editor, *Memoirs of the American Mathematical Society*. 1951.

J. Feldman, A. Mehta, V. Mirrokni, and S. Muthukrishnan. Online stochastic matching: Beating 1-1/e. In *Proceedings of the 50th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 117–126, 2009.

P. Ferrari, W. Weimar, R.J. Johnson, W.H. Lim, and K.J. Tinckam. Kidney paired donation: principles, protocols and programs. *Nephrology Dialysis Transplantation*, 30(8):1276–1285, 2014.

S.E. Gentry and D.L. Segev. The best-laid schemes of mice and men often go awry; how should we repair them? *American Journal of Transplantation*, 15(10):2539–2540, 2015.

P.W. Glynn and A. Zeevi. Bounding stationary expectations of markov processes. In *Markov processes and related topics: a Festschrift for Thomas G. Kurtz*, pages 195–214. Institute of Mathematical Statistics, 2008.

G. Goel and A. Mehta. Online budgeted matching in random input models with applications to adwords. In *Proceedings of the nineteenth annual ACM-SIAM symposium on Discrete algorithms (SODA)*, pages 982–991, 2008.

I. Gurvich and A.R. Ward. On the dynamic control of matching queues. *Stochastic Systems*, 4(2): 479–523, 2014.

S. Jasin and S. 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.

R.J. Johnson, J.E. Allen, S.V. Fuggle, J.A. Bradley, C. Rudge, and UK Transplant NHSBT Kidney Advisory Group. Early experience of paired living kidney donation in the united kingdom. *Transplantation*, 86(12):1672–1677, 2008.

R.M. Karp, U.V. Vazirani, and V.V. Vazirani. An optimal algorithm for on-line bipartite matching. In *Proceedings of the twenty-second annual ACM symposium on Theory of computing (STOC)*, pages 352–358, 1990.

J.D. Leshno. Dynamic matching in overloaded systems. 2011.

X. Li and Y. Ye. Online linear programming: Dual convergence, new algorithms, and regret bounds. *Working Paper*, 2020. Stanford University.

Z. Li, K. Lieberman, W. Macke, S. Carrillo, C.J. Ho, J. Wellen, and S. Das. Incorporating compatible pairs in kidney exchange: A dynamic weighted matching model. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 349–367, 2019.

S. Malik and E. Cole. Foundations and principles of the canadian living donor paired exchange program. *Canadian journal of kidney health and disease*, 1(1):6, 2014.

O.L. Mangasarian and T.H. Shiau. Lipschitz continuity of solutions of linear inequalities, programs and complementarity problems. *Siam J. Control and Optimization*, 25(3), 1987.

V.H. Manshadi, S.O. Gharan, and A. Saberi. Online stochastic matching: Online actions based on offline statistics. *Mathematics of Operations Research*, 37(4):559–573, 2012.

S.P. Meyn and R.L. Tweedie. Stability of markovian processes I: Criteria for discrete-time chains. *Advances in Applied Probability*, pages 542–574, 1992.

M. Nazari and A.L. Stolyar. Reward maximization in general dynamic matching systems. *Queueing Systems*, 91(1):143–170, 2019.

P. Robert. *Stochastic networks and queues.* Springer-Verlag Berlin Heidelberg, 2003.

U.M. Ünver. Dynamic kidney exchange. *Review of Economic Studies*, 77(1):372–414, 2010.

A.W. Van der Vaart. *Asymptotic statistics.* Cambridge university press, Cambridge, 1998.

A. Vera and S. Banerjee. The bayesian prophet: A low-regret framework for online decision making. *Management Science*, 67(3):1368–1391, 2021.

A. Vera, S. Banerjee, and I. Gurvich. Online allocation and pricing: Constant regret via Bellman inequalities. *Operations Research*, 2020. Forthcoming.

W. Whitt. *Stochastic-process limits: an introduction to stochastic-process limits and their application to queues*. Springer-Verlang New York, 2002.

A Proofs from Section 2

Proof for Figure 3. Some pre-processing is useful here. It is a simple observation that under the optimal total value for a fixed t—realizable by taking no action until time t, and performing matches according to an optimal solution at that point—the optimal solution is given by setting

$$z_1^{*,t} := A_1^t \wedge A_2^t \text{ and } z_2^{*,t} := A_3^t \wedge (A_2^t - A_1^t)^+, \tag{14}$$

so that

$$\mathcal{R}^{*,t} = r_1\mathbb{E}[A_1^t \wedge A_2^t] + r_2\mathbb{E}[A_3^t \wedge (A_2^t - A_1^t)^+]. \tag{15}$$

Fix $\bar{t} = \alpha t$ for some $\alpha \in (0,1)$. Then the optimal value at time $\bar{t}$ is the same as (15), where t is replaced by $\bar{t}$. We also use the following simple fact: the multivariate central limit theorem (Van der Vaart, 1998, Example 2.1.8) applied to the multinomial random vector (A_1^t, A_2^t, A_3^t) and the continuity of the map $(x_1, x_2, x_3) \to (x_1 - x_2)$ implies that

$$\mathbb{P}\{A_1^t - A_2^t \leq \delta\sqrt{t}\} \to \Phi(\delta/\sqrt{\lambda}) \text{ as } t \to \infty, \tag{16}$$

where Φ is the cumulative density function of the standard normal distribution and $\lambda = \lambda_1 = \lambda_2 = \lambda_3$.

The proof now proceeds in two parts. We first show that *any* non-anticipating policy D that has the optimality guarantee $\mathcal{R}^{*,t} - \mathcal{R}^{D,t} = o(\sqrt{t})$, must not perform match$-2$ until late in the horizon. A consequence of this, as we will show, is that any such policy must have $\mathcal{R}^{*,t} - \mathcal{R}^{D,t} = \Omega(\sqrt{t})$.

Part 1. Fix $\alpha = 1/2$ ($\bar{t} = t/2$). The proof works for any $\alpha \in (0,1)$, but fixing $\alpha = 1/2$ is notationally convenient. For some $\kappa > 0$, let

$$\tau := \inf\left\{t \geq 0 : D_2^t \geq \kappa\sqrt{t}\right\}$$

be the first time that match-2 is used more than $\kappa\sqrt{t}$ times and fix $\delta > 2\kappa$. The following two events are independent under any non-anticipating policy D:

$$\mathcal{E}_1 := \{\tau \leq \bar{t}\} \cap \left\{A_1^{\bar{t}} - A_2^{\bar{t}} \geq -\frac{\delta}{2}\sqrt{t}\right\} \text{ and } \mathcal{E}_2 := \left\{A_1^{(\bar{t},t]} - A_2^{(\bar{t},t]} \geq \delta\sqrt{t}\right\},$$

where we introduced the increments $A_i^{(s,u]} := A_i^u - A_i^s$. On the intersection $\mathcal{E}_1 \cap \mathcal{E}_2$, we have

$A_1^t - A_2^t \geq \delta\sqrt{t}/2$, which implies $A_1^t \geq A_2^t$. Per (14), we have $z_2^{*,t} = 0$ so that, on this event, the policy loses $(r_1 - r_2)\kappa\sqrt{t}$ relative to the optimal. Using the independence of the events $\mathcal{E}_1$ and $\mathcal{E}_2$, we have

$$\mathcal{R}^{*,t} - \mathcal{R}^{D,t} \geq (r_1 - r_2)\kappa\sqrt{t}\mathbb{P}\{\mathcal{E}_1\}\mathbb{P}\{\mathcal{E}_2\}.$$

Per (16), $\mathbb{P}\{\mathcal{E}_2\} \to \eta > 0$ as $t \to \infty$. For the policy to have $\mathcal{R}^{*,t} - \mathcal{R}^{D,t} = o(\sqrt{t})$, it must be that

$$\mathbb{P}\{\mathcal{E}_1\} \to 0 \text{ as } t \to \infty.$$

Then for large enough t, we have

$$\mathbb{P}\{\tau \leq \bar{t}\} \leq \mathbb{P}\{\mathcal{E}_1\} + \mathbb{P}\left\{A_1^\tau - A_2^\tau \leq -\frac{\delta}{2}\sqrt{t}\right\} \leq 2\eta.$$

Recalling the definition of τ, this shows that a policy D that has $\mathcal{R}^{D,t} - \mathcal{R}^{*,t} = o(\sqrt{t})$ will, with high probability, avoid perfoming match-2 until time $\bar{t} = t/2$.

Part 2. We claim that any policy that has the optimality guarantee $o(\sqrt{\bar{t}}\ )$ at time $\bar{t}$, must have for all $\kappa > 0$ that

$$\mathbb{P}\{Q_2^{\bar{t}} > \kappa\sqrt{\bar{t}}\ \} \to 0 \text{ as } t \to \infty. \tag{17}$$

Before proving this claim, we will use the arguments in part 1 to show that if a policy is value optimal at t, we contradict (17), and thus the near optimality at $\bar{t}$.

Since $D_1^u \leq A_1^u$ for all $u > 0$, we have that for all $s \leq \tau$,

$$Q_2^s = A_2^s - D_1^s - D_2^s \geq (A_2^s - A_1^s - \kappa\sqrt{t})^+.$$

Thus,

$$\begin{aligned}
\mathbb{P}\{Q_2^{\bar{t}} > \kappa\sqrt{\bar{t}}\ \} &\geq \mathbb{P}\{(A_2^t - A_1^t - \kappa\sqrt{t})^+ \geq \kappa\sqrt{\bar{t}}, \tau > \bar{t}\ \} \\
&\geq \mathbb{P}\{(A_2^t - A_1^t - \kappa\sqrt{t})^+ \geq \kappa\sqrt{\bar{t}}\ \} - \mathbb{P}\{\tau \leq \bar{t}\ \} \\
&\geq \mathbb{P}\{(A_2^t - A_1^t - \kappa\sqrt{t})^+ \geq \kappa\sqrt{\bar{t}}\ \} - 2\eta.
\end{aligned}$$

Per (16), there exists $\gamma = \gamma(\kappa)$ such that $\mathbb{P}\{A_2^{\bar{t}} - A_1^{\bar{t}} \geq 2\kappa\sqrt{t})\} \geq \gamma$. Choosing δ large (and consequently, η small) so that $2\eta < \gamma$, we have that a policy that has $\mathcal{R}^{*,t} - \mathcal{R}^{D,t} = o(\sqrt{t})$, must also have $\mathbb{P}\{Q_2^{\bar{t}} > \kappa\sqrt{\bar{t}}\ \} \geq (\gamma - 2\eta) > 0$ for all $t > 0$, which contradicts (17) as required.

To conclude the proof, it remains to show that any policy with the suboptimality gap $o(\sqrt{t}\,)$, must have $\mathbb{P}\{Q_2^{\bar{t}} \geq \kappa\sqrt{\bar{t}}\,\} \to 0$ as t (and then $\bar{t} = t/2$) $\to \infty$.

Because $D_1^u + D_2^u \leq A_2^u$ for all $u > 0$, we have $Q_1^u + Q_3^u = A_1^u + A_3^u - D_1^u - D_2^u \geq A_1^u + A_3^u - A_2^u$ for all $u > 0$. Since $\lambda_1 + \lambda_3 > \lambda_2$, we have by the strong law of large numbers that

$$\mathbb{P}\{Q_1^{\bar{t}} + Q_3^{\bar{t}} \geq \kappa\sqrt{\bar{t}}\,\} \geq \mathbb{P}\{A_1^{\bar{t}} + A_3^{\bar{t}} - A_2^{\bar{t}} \geq \kappa\sqrt{\bar{t}}\,\} \to 1 \text{ as } t \to \infty.$$

If in contrast to our claim, there exists $\theta > 0$ such that $\mathbb{P}\{Q_2^{\bar{t}} \geq \kappa\sqrt{\bar{t}}\,\} \geq \theta$, then for all sufficiently large t, we have

$$\mathbb{P}\{(Q_1^{\bar{t}} + Q_3^{\bar{t}}) \wedge Q_2^{\bar{t}} \geq \kappa\sqrt{\bar{t}}\,\} \geq \theta/2.$$

On the event $\{(Q_1^{\bar{t}} + Q_3^{\bar{t}}) \wedge Q_2^{\bar{t}} \geq \kappa\sqrt{\bar{t}}\,\}$, there are $\kappa\sqrt{\bar{t}}$ unused feasible matches, which implies

$$\mathcal{R}^{*,\bar{t}} - \mathcal{R}^{D,\bar{t}} \geq \mathbb{E}[(r_1 \wedge r_2)((Q_1^{\bar{t}} + Q_3^{\bar{t}}) \wedge Q_2^{\bar{t}})] \geq (r_1 \wedge r_2)\kappa\sqrt{\bar{t}}\,\mathbb{P}\{(Q_1^{\bar{t}} + Q_3^{\bar{t}}) \wedge Q_2^{\bar{t}} \geq \kappa\sqrt{\bar{t}}\,\} \geq (r_1 \wedge r_2)\kappa\sqrt{\bar{t}}\,\theta/2,$$

contradicting the optimality guarantee $o(\sqrt{\bar{t}}\,)$ of the policy at time $\bar{t}$. □

B Proofs from Section 4

Proof of Lemma 4.1. Let $\mathcal{B}$ be the corresponding optimal basis to (z^*, s^*). Recall that $\mathcal{Q}_0 = \{i \in \mathcal{A} : s_i^* = 0\}$ and $\mathcal{M}_0 = \{m \in \mathcal{M} : z_m^* = 0\}$ are the corresponding sets of over-demanded queues and redundant matches, respectively.

Let (z, s) be *any* feasible solution of (SPP) that has $s_i = 0$ for all $i \in \mathcal{Q}_0$ and $z_m = 0$ for all $m \in \mathcal{M}_0$. Then it must be that $(z, s) = \mathcal{B}^{-1}\lambda$, and in particular, $z_m = y^m\lambda$ for all $m \in \mathcal{M}_+$. This immediately follows, since the linear system $\{Mz + s = \lambda, z \geq 0, s \geq 0\}$ with the condition we set on $s_i, i \in \mathcal{Q}_0$, and $z_m, m \in \mathcal{M}_0$, has a unique solution.

Recall also that (z^*, s^*) has a non-degenerate basis. In particular, the same conclusion holds if λ is replaced by $\tilde{\lambda} = \lambda + \zeta$ for a suitably small $\zeta \in \mathbb{R}^n$. That is, any feasible solution to the linear system $\{Mz + s = \tilde{\lambda}, z \geq 0, s \geq 0\}$ with $s_i = 0$ for all $i \in \mathcal{Q}_0$ and $z_m = 0$ for all $m \in \mathcal{M}_0$, must satisfy $z_m = y^m\tilde{\lambda}$ for all $m \in \mathcal{M}_+$.

Fix $t = \Omega(\epsilon^{-2})$. Consider a policy D that does not execute any matches in $\mathcal{M}_0$. Let $q_i := \mathbb{E}^D[Q_i^{t^+}] \leq \mathbb{E}^D[Q_i^t] = \mathcal{O}(\epsilon^{-1})$ be the post-match queue length vector and $z_m := D_m^t$. Let $\bar{z} := z/t$

and $\bar{q} := q/t$. Using the fact that $Mz + q = \lambda t$, we have

$$M\bar{z} + \bar{q} = \lambda,$$

where $\bar{q}_i = \mathcal{O}(\epsilon)$ for all $i \in \mathcal{Q}_0$ and $\bar{z}_m = 0$ for all $m \in \mathcal{M}_0$. For all $i \in \mathcal{A}$, define

$$\tilde{\lambda}_i := \lambda_i - \bar{q}_i \mathbb{1}_{\{i \in \mathcal{Q}_0\}}.$$

Let $\tilde{z}_m := \bar{z}_m$ for all $m \in \mathcal{M}_+$ and 0 otherwise. Then $(\tilde{z}, \tilde{q})$ satisfies $M\tilde{z} + \tilde{q} = \tilde{\lambda}$, where $\tilde{q}_i = 0$ for all $i \in \mathcal{Q}_0$ and $\tilde{z}_m = 0$ for all $m \in \mathcal{M}_0$. Per the above arguments, then it must be that $\tilde{z}_m = y^m\tilde{\lambda}$ for all $m \in \mathcal{M}_+$. Since $\mathcal{R}^{D,t} = t(r \cdot \bar{z}) \geq t(r \cdot \tilde{z}) = t\sum_{m \in \mathcal{M}_+} r_m y^m \tilde{\lambda}$ and $\mathcal{R}^{*,t} \leq t(r \cdot z^*)$, we have

$$\mathcal{R}^{*,t} - \mathcal{R}^{D,t} \leq t\left(r \cdot z^* - r \cdot \tilde{z}\right) = t\left(\sum_{m \in \mathcal{M}_+} r_m y^m \lambda - \sum_{m \in \mathcal{M}_+} r_m y^m \tilde{\lambda}\right) \leq t r_{max} \omega \|\lambda - \tilde{\lambda}\|_1,$$

where $r_{max} := \max_{m \in \mathcal{M}_+} r_m$, and we used the fact that the vectors y^m have all entries in $[-\omega, \omega]$. Recalling that $|\lambda_i - \tilde{\lambda}_i| = \bar{q}_i \mathbb{1}_{\{i \in \mathcal{Q}_0\}}$, we conclude that

$$\mathcal{R}^{*,t} - \mathcal{R}^{D,t} \leq t r_{max} \omega \|\lambda - \tilde{\lambda}\| = t\mathcal{O}(\epsilon) = \mathcal{O}(\epsilon^{-1}),$$

as required. □

B.1 Proof of Proposition 4.1

We first prove (8). Recall the problem

$$\begin{aligned}
h^*(u) = \quad \min \quad & \textstyle\sum_{i \in \mathcal{Q}_0} \sum_{m \in \mathcal{M}_+} (r_m y^m)_i s_i \\
\text{s.t.} \quad & z_m + y^m s = y^m u \text{ for all } m \in \mathcal{M}_+ \\
& y^j s = y^j u \text{ for all } j \in \mathcal{Q}_+ \\
& z \in \mathbb{Z}^d_{\geq 0}, s \in \mathbb{Z}^n_{\geq 0}.
\end{aligned} \tag{18}$$

Since q^+ is the post-match queue-length vector, no more matches can be performed from q^+ itself. Thus, we have $h^*(q) = h^*(q^+) = \sum_{i \in \mathcal{Q}_0} \theta_i q_i^+$. It is also immediate that for all $x \in [\mathbf{0}^n, A^\tau] \cap \mathbb{Z}^n_{\geq 0}$, we have

$$h^*(q^+ + A^\tau) \le h^*(q^+ + x) + h^*(A^\tau - x). \tag{19}$$

For $h^*(A^\tau - x)$, note that if $y^l(A^\tau - x) > 0$ for all $l \in \mathcal{M}_+ \cup \mathcal{Q}_+$, then setting $z_m = y^m(A^\tau - x)$ for all $m \in \mathcal{M}_+$, $s_i = 0$ for all $i \in \mathcal{Q}_0$, and $s_j = y^j(A^\tau - x)$ for all $j \in \mathcal{Q}_+$, is feasible for (18) with the objective function value of 0. Then it is also optimal, since the objective function is non-negative. Let

$$\mathcal{X} := \mathcal{X}(A^\tau) := \{x \in \mathbb{Z}^n_{\ge 0} : y^l(A^\tau - x) > 0 \text{ for all } l \in \mathcal{M}_+ \cup \mathcal{Q}_+\}.$$

Then we have $h^*(A^\tau - x) = 0$ for all $x \in \mathcal{X}$, and (19) implies

$$h^*(q^+ + A^\tau) \le \inf_{x \in \mathcal{X}} h^*(q^+ + x). \tag{20}$$

Our goal is to show that when $h^*(q) > B$, for a suitable choice of $\tau = \lceil \kappa \epsilon^{-1} \rceil$, we have $0 \in \mathcal{X}$ with high probability, and the inequality (20) is strict for $x = 0$. To that end, consider the following event

$$\mathcal{C} := \mathcal{C}(\tau) := \left\{ \left| y^l A^\tau - y^l \lambda \tau \right| \le \frac{1}{2} y^l \lambda \tau \text{ for all } l \in \mathcal{M}_+ \cup \mathcal{Q}_+ \right\}.$$

Since $y^l \lambda \ge \epsilon$ for all $l \in \mathcal{M}_+ \cup \mathcal{Q}_+$, we have on $\mathcal{C}$ that $y^l A^\tau \ge \epsilon\tau/2$ for all $l \in \mathcal{M}_+ \cup \mathcal{Q}_+$. Thus, for any $x \in [\mathbf{0}^n, A^\tau] \cap \mathbb{Z}^n_{\ge 0}$ such that $\|x\|_1 \le \frac{\epsilon\tau}{4\omega}$ (in particular, $|y^l x| \le \epsilon\tau/4$ for all $l \in \mathcal{M}_+ \cup \mathcal{Q}_+$), on $\mathcal{C}$ we have

$$y^l(A^\tau - x) \ge \frac{\epsilon\tau}{4} \text{ for all } l \in \mathcal{M}_+ \cup \mathcal{Q}_+.$$

In particular, we have $0 \in \mathcal{X}$ on $\mathcal{C}$, and (20) implies $h^*(q^+ + A^\tau) \le h^*(q^+)$.

Let $i \in \mathcal{Q}_0$ such that $q_i^+ > B/\theta_i$ (such i must exists if $h^*(q) = h^*(q^+) > B$). Consider $m \in \mathcal{M}_+$ such that $i \in \mathcal{A}(m)$, and set $x_j = \lfloor \frac{\kappa}{4n\omega} \rfloor > 0$ for all $j \ne i$ such that $j \in \mathcal{A}(m)$ and 0 otherwise. Note that $x_j = \lfloor \frac{\kappa}{4n\omega} \rfloor \le \frac{\epsilon\tau}{4n\omega}$ and $\|x\|_1 \le \frac{\epsilon\tau}{4\omega}$. Then B can be chosen sufficiently large so that it is feasible to perform an additional $\lfloor \frac{\kappa}{4n\omega} \rfloor$ many match$-m$'s without changing any of the other queues. Since $x \in \mathcal{X}$, we have on $\mathcal{C}$ that

$$h^*(q^+ + A^\tau) \le h^*(q^+ + x) \le h^*(q^+) - \theta_i \left\lfloor \frac{\kappa}{4n\omega} \right\rfloor \le h^*(q^+) - \underline{\theta} \left\lfloor \frac{\kappa}{4n\omega} \right\rfloor,$$

where $\underline{\theta} := \min_{i \in \mathcal{Q}_0} \theta_i > 0$.

A simple extension of Chernoff bounds for the sums $y^l A^\tau,\ l \in \mathcal{M}_+ \cup \mathcal{Q}_+$, yields

$$\mathbb{P}\left\{\left|y^l A^\tau - y^l \lambda\tau\right| \geq \frac{1}{2} y^l \lambda\tau\right\} \leq c_3 e^{-c_4 y^l \lambda\tau} \leq c_3 e^{-c_4 \epsilon\tau} \text{ for all } l \in \mathcal{M}_+ \cup \mathcal{Q}_+,$$

for some constants $c_3, c_4 > 0$, where recall that $\epsilon = \min_{m\in\mathcal{M}_+} y^m\lambda \wedge \min_{j\in\mathcal{Q}_+} y^j\lambda$. By the union bound, we have

$$\mathbb{P}\{\mathcal{C}^c\} \leq nc_3 e^{-c_4\epsilon\tau}.$$

Now we utilize the following lemma, which provides an upper bound for the expectation in (8) when we are outside of the event $\mathcal{C}$.

Lemma B.1. *For some constant $K > 0$, which does not depend on ϵ, we have*

$$\mathbb{E}_q[((h^*(Q^\tau) - h^*(q))^+)^2] \leq K^2\epsilon^2\tau^2.$$

Hölder's inequality then implies that

$$\mathbb{E}[(h^*(q + A^\tau) - h^*(q))^+ \mathbb{1}_{\{\mathcal{C}^c\}}] \leq K\epsilon\tau nc_3 e^{-c_4\epsilon\tau}.$$

Given $\delta \in (0, 1)$, set $\tau = \lceil \kappa\epsilon^{-1} \rceil$ with large enough $\kappa \geq 8n\omega$ such that

$$nc_3 e^{-c_4\epsilon\tau} \leq \delta \text{ and } K\epsilon\tau nc_3 e^{-c_4\epsilon\tau} \leq (1-\delta)\underline{\theta}\left\lfloor \frac{\kappa}{8n\omega} \right\rfloor.$$

Recalling that $h^*(q) = h^*(q^+)$, we can then conclude that if $h^*(q) > B$, then

$$\begin{aligned}\mathbb{E}_q[h^*(Q^\tau) - h^*(q)] &\leq -\mathbb{P}\{\mathcal{C}\}\underline{\theta}\left\lfloor \frac{\kappa}{4n\omega} \right\rfloor + \mathbb{E}[(h^*(q + A^\tau) - h^*(q))^+ \mathbb{1}_{\{\mathcal{C}^c\}}] \\ &\leq -(1-\delta)\underline{\theta}\left\lfloor \frac{\kappa}{4n\omega} \right\rfloor + (1-\delta)\underline{\theta}\left\lfloor \frac{\kappa}{8n\omega} \right\rfloor \\ &\leq -\gamma,\end{aligned}$$

where $\gamma := \underline{\theta}\frac{(1-\delta)}{16n\omega} > 0$. If $h^*(q) \leq B$, then clearly

$$\mathbb{E}_q[h^*(Q^\tau) - h^*(q)] \leq B + \sum_{i\in\mathcal{Q}_0} \theta_i\lambda_i\tau \leq \frac{1}{\epsilon}(B + \sum_{i\in\mathcal{Q}_0} \theta_i\lambda_i(\kappa+1)),$$

where for the last inequality, we used the fact that $\epsilon < 1$ and $\kappa + 1 \geq \epsilon\tau$. This establishes the drift property (8), and we turn to prove (9). This follows from a standard mechanism, which derives an exponential Lyapunov function from a given linear one. Note that under the exhaustive resolving policy, any match that is performed at any decision period t_k must contain at least one agent type that arrived between t_{k-1} and t_k. Thus, we have $\sum_{m\in\mathcal{M}_+}(D_m^{t_{k+1}} - D_m^{t_k}) \leq \sum_{i\in\mathcal{A}}(A^{t_{k+1}} - A^{t_k})$. Merging this fact with (1) immediately implies the following auxiliary lemma.

Lemma B.2. *Under the exhaustive resolving policy, we have*

$$\sum_{i\in\mathcal{A}} |Q_i^{t_{k+1}^+} - Q_i^{t_k^+}| \leq \sum_{i\in\mathcal{A}}(A^{t_{k+1}} - A^{t_k}) \leq n\tau \text{ for all } k \in \mathbb{N}.$$

Let $\bar{\theta} := \max_i \theta_i > 0$. Then by Lemma B.2, we have

$$C = \sup_{q\in\mathcal{S}} \mathbb{E}_q \left[e^{|h^*(Q^\tau) - h^*(q)|} \right] \leq e^{\bar{\theta} n\tau} < \infty.$$

In particular, the second condition of (Robert, 2003, Proposition 8.8) is satisfied with $\lambda = 1$ there. It also follows from the proof of (Robert, 2003, Proposition 8.8) that

$$\mathbb{E}_q[e^{h^*(Q^\tau)}] \leq e^{h^*(q)}(1 - \gamma/2), \text{ if } q \in F^c.$$

Since the linear program (7) that defines $h^*(\cdot)$ is Lipschitz continuous in the right hand side, we have $h^*(Q^\tau) \leq \max_{q\in F} h^*(q) + c_5\tau$ for some constant $c_5 > 0$. Letting $c_6 := e^{\max_{q\in F} h^*(q)}$, we have

$$\mathbb{E}_q[e^{h^*(Q^\tau)}] \leq c_6 e^{c_5\tau}, \text{ if } q \in F.$$

Overall, we obtain (9):

$$\begin{aligned} \mathbb{E}_q \left[e^{h^*(Q^\tau)} - e^{h^*(q)} \right] &\leq -\frac{\gamma}{2} e^{h^*(q)} \mathbb{1}_{\{q\in F^c\}} + c_6 e^{c_5\tau} \mathbb{1}_{\{q\in F\}} \\ &\leq -\frac{\gamma}{2} e^{h^*(q)} + c_7 e^{c_5\tau} \mathbb{1}_{\{q\in F\}} \end{aligned}$$

for some constant $c_7 > c_6$. $\square$

Proof of Lemma B.1. We will first show that given $x \in \mathbb{Z}^n_{\geq 0}$, we have

$$(h^*(q + x) - h^*(q))^+ \leq K \max_{l\in\mathcal{Q}_+\cup\mathcal{M}_+} (y^l x)^- \leq K \sum_{l\in\mathcal{Q}_+\cup\mathcal{M}_+} (y^l x)^-. \tag{21}$$

The proof than follows immediately by setting $x = A^\tau$ and using the following auxiliary result with a redefined constant K, which we prove in the end of this section.

Proposition B.1.
$$\mathbb{E}\left[\left(\min_{l\in\mathcal{Q}_+\cup\mathcal{M}_+}(y^l A^\tau)^-\right)^2\right] \leq K^2\epsilon^2\tau^2,$$
for some constant $K > 0$, which does not depend on ϵ.

We turn then to prove (21). Recall that $h^*(q + x) \leq h^*(q) + h^*(x)$, where $h^*(x) \geq 0$, and we have $h^*(x) = 0$ if $y^l x \geq 0$ for all $l \in \mathcal{M}_+ \cup \mathcal{Q}_+$. Then $(h^*(q + x) - h^*(q))^+ \leq h^*(x)$, and it suffices to show that for any $x \in \mathbb{Z}^n_{\geq 0}$, not necessarily satisfying $y^l x \geq 0$ for all $l \in \mathcal{M}_+ \cup \mathcal{Q}_+$, we have

$$h^*(x) \leq K \max_{l\in\mathcal{Q}_+\cup\mathcal{M}_+}(y^l x)^-. \tag{22}$$

Given $x \in \mathbb{Z}^n_{\geq 0}$, let $\zeta \in \mathbb{Z}^n_{\geq 0}$ be such that $y^l x = y^l\zeta$ for all $l \in \mathcal{P}^+ = \{l \in \mathcal{M}_+ \cup \mathcal{Q}_+ : y^l x \geq 0\}$, and $y^l\zeta = 0$ for all $l \in (\mathcal{M}_+ \cup \mathcal{Q}_+)\backslash\mathcal{P}^+$. Note that the linear program (7) has for all $l \in \mathcal{P}^+$ the same right-hand side for either x or ζ. The right-hand side differs only for $l \in (\mathcal{M}_+ \cup \mathcal{Q}_+)\backslash\mathcal{P}^+$, and since $y^l\zeta = 0$ for such indices, the difference in the right hand side is $|y^l\zeta - y^l x| = |y^l x|$. By the Lipschitz continuity of (SPP) (e.g., see (Mangasarian and Shiau, 1987)), we have

$$|h^*(x) - h^*(\zeta)| \leq K \max_{l\in(\mathcal{M}_+\cup\mathcal{Q}_+)\backslash\mathcal{P}^+}|y^l x| = K \max_{l\in\mathcal{M}_+\cup\mathcal{Q}_+}(y^l x)^-,$$

where we used the fact that $(y^l x)^- = 0$ for all $l \in \mathcal{P}^+$. Finally, since $y^l\zeta \geq 0$ for all $l \in \mathcal{M}_+ \cup \mathcal{Q}_+$, we have $h^*(\zeta) = 0$ so that we arrive at (22).

That existence of ζ is straightforward. Construct a "matching increment" $\mu \in \mathbb{Z}^d_{\geq 0}$ as follows:

$$\mu = \begin{cases} 0, & \text{if } m \in \mathcal{P}^+, \\ (y^m x)^-, & \text{if } m \in \mathcal{M}_+\backslash\mathcal{P}^+. \end{cases}$$

Let $\phi = M\mu$, and observe that $y^m\phi = y^m M\mu = \mu_m$ for all $m \in \mathcal{M}_+$. Letting

$$\zeta = x + \phi \geq 0,$$

we then have that $y^m\zeta = y^m x$ for all $m \in \mathcal{P}^+$ and $y^m\zeta = y^m x + (y^m x)^- = 0$ for all $m \in \mathcal{M}_+\backslash\mathcal{P}^+$ as required. □

Proof of Proposition B.1. Since

$$\mathbb{E}\left[\left(\min_{l\in\mathcal{M}_+\cup\mathcal{Q}_+}(y^l A^\tau)^-\right)^2\right] \le K\left(\sum_{l\in\mathcal{M}_+\cup\mathcal{Q}_+}\mathbb{E}\left[\left((y^l A^\tau)^-\right)^2\right]\right),$$

it suffices to establish that the bound holds for each $l \in \mathcal{M}_+ \cup \mathcal{Q}_+$ separately. Note that

$$y^l(A^\tau - \lambda\tau) = \sum_{t=1}^{\tau} y^l(\Delta A^l - \lambda),$$

where $\Delta A^t = A^t - A^{t-1}$. Observe that the variables in the sum are i.i.d, and each variable is bounded by n. Then by Hoeffding's inequality, for any $k > 0$, we have

$$\mathbb{P}\left\{|y^l(A^\tau - \lambda\tau)| \ge y^l\lambda\tau + k\epsilon\tau\right\} \le 2\exp\left(-\frac{2(y^l\lambda\tau + k\epsilon\tau)^2}{\tau n^2}\right) \le 2\exp\left(-\frac{2}{n^2}(y^l\lambda\tau + k\epsilon\tau)\right).$$

Notice that n here is the number of agent types, which does not change with ϵ or τ. In turn,

$$\begin{aligned}\mathbb{P}\{(y^l A^\tau)^- \ge k\epsilon\tau\} &= \mathbb{P}\{y^l A^\tau \le -k\epsilon\tau\} \le \mathbb{P}\{|y^l(A^\tau - \lambda\tau)| \ge y^l\lambda\tau + k\epsilon\tau\} \\ &\le \exp\left(-\frac{2}{n^2}(y^l\lambda\tau + k\epsilon\tau)\right) \le \exp\left(-\frac{2}{n^2}k\epsilon\tau\right),\end{aligned}$$

where the last inequality uses the fact that $y^l\lambda \ge 0$. This exponential tail then implies the result of the lemma by a simple integration. $\square$

C Proofs from Section 5

Proof of Lemma 5.1. The task here is to prove that under the assumption of the lemma, (SPP) can be equivalently represented by a suitable minimum-cost network flow problem. Since such a flow problem always has an acyclic optimal solution (Ahuja et al., 1993, Theorem 11.1), the lemma then follows from the assumed uniqueness under **GP**.

First, let us create the partition. Having only two-way matches allows us to represent the matching network graph as a simple graph. That is, we will have a vertex corresponding to each agent type (but not for matches), and there exists an edge between $i, j \in \mathcal{A}$ if and only if there exists $m \in \mathcal{M}$ such that $M_{im} = M_{jm} = 1$. Thus, each edge (i, j) in this simple graph representation is uniquely identified by a match, and we will write $r_{i,j}$ for the value of that match.

Our assumption—that any cycle contains an even number of matches—translates in this simple graph representation to assuming that any cycle is of even length. Since a simple graph is bipartite if and only if it does not contain any odd cycles, we have a partition of $\mathcal{A}$ into two disjoint subsets $\mathcal{A}_1$ and $\mathcal{A}_2$ such that all edges in the graph are between some $i \in \mathcal{A}_1$ and $j \in \mathcal{A}_2$.

As it is customary, we augment this graph with an origin (or supply) node s, and a destination (or target) node t. There will be directed outgoing edges from s to each $i \in \mathcal{A}_1$, as well as outgoing edges from each $j \in \mathcal{A}_2$ to t, and each edge (i, j) in this graph is directed from $i \in \mathcal{A}_1$ to $j \in \mathcal{A}_2$.

Note that the resulting directed graph, by construction, has no directed cycles. For each edge (i, j) in this graph, we place a negative cost $-r_{i,j}$ ($i \in \mathcal{A}_1, j \in \mathcal{A}_2$). We also put upper bounds $x_{s,i} \leq \lambda_i$ for all $i \in \mathcal{A}_i$ and $x_{j,t} \leq \lambda_j$ for all $j \in \mathcal{A}_2$. Consider the following minimum-cost network flow problem

$$\begin{array}{ll}
\min & -\sum_{i \in \mathcal{A}_1, j \in \mathcal{A}_2} r_{i,j} x_{i,j} \\
\text{s.t.} & \sum_{j \in \mathcal{A}_2} x_{i,j} - x_{s,i} = 0 \text{ for all } i \in \mathcal{A}_1 \\
& \sum_{i \in \mathcal{A}_1} x_{i,j} - x_{j,t} = 0 \text{ for all } j \in \mathcal{A}_2 \\
& x_{s,i} \leq \lambda_i \text{ for all } i \in \mathcal{A}_1 \\
& x_{j,t} \leq \lambda_j \text{ for all } j \in \mathcal{A}_2 \\
& x \geq 0.
\end{array}$$

This problem has a cycle free solution; e.g., see (Ahuja et al., 1993, Chapter 11.1). In particular, since the variables $x_{i,j}$ ($i \in \mathcal{A}_1, j \in \mathcal{A}_2$) have no upper or lower bounds, there is no (undirected) cycle consists of edges such that $x_{i,j} > 0$ for all edges (i, j) in the cycle.

Recall that these edges correspond to matches in the original matching network. Let $z_m = x_{i,j}$ for all $m = (i, j) \in \mathcal{M}$, $s_i = \lambda_i - x_{s,i}$ for all $i \in \mathcal{A}_1$, and $s_j = \lambda_j - x_{j,t}$ for all $j \in \mathcal{A}_2$. Then it is immediate that the minimum-cost network flow problem is equivalent to (SPP). In turn, the optimal solution to the latter problem is acyclic, where the uniqueness is assumed under **GP**. $\square$

Proof of Lemma 5.2. Let (z^*, s^*) be an optimal basic feasible solution of (SPP) such that the corresponding LP-residual graph is acyclic. By Theorem 5.1, we know that for any $m \in \mathcal{M}$ that is a basic variable, we have $z^*_m = y^m \lambda \geq 0$, and for any $i \in \mathcal{A}$ that is a basic variable, we have $s^*_i = y^i \lambda \geq 0$. Note that if $z^*_m = y^m \lambda = 0$, then condition (12) is violated, since y^m is a vector with all entries in $\{-1, 0, 1\}$. Similarly, we must have $s^*_i > 0$, which implies that the optimal basis is non-degenerate. $\square$

Construction of the surplus vectors. Removing all redundant matches $m \in \mathcal{M}_0$ from $\mathcal{G}$, decomposes the network into (possibly) multiple connected components. Throughout the construction in this section, we assume, without loss of generality, that there is a single component, i.e., $\mathcal{M}_0 = \emptyset$. Otherwise, the following procedure is applied separately to each component.

Let $\mathcal{U}_0 := \{i \in \mathcal{Q}_0 : \sum_{m\in\mathcal{M}} M_{im} = 1\}$. This is the set of queues in $\mathcal{Q}_0$ participating in exactly one match; $\mathcal{U}_0$ is a subset of the leaves in $\mathcal{G}$. The following lemma shows that $\mathcal{U}_0$ is non-empty.

Lemma C.1. *The number of leaves in $\mathcal{G}$ is at least $n-d+1$. Since $|\mathcal{Q}_+| = n-d$, at least one of the leaves must be in $\mathcal{Q}_0$, and in turn, $|\mathcal{U}_0| \geq 1$.*

For each pair of vertices $j \in \mathcal{Q}_+$ and $i \in \mathcal{U}_0$, we traverse the unique path between j and i in the (SPP)-residual graph $\mathcal{G}$. Starting from $j \in \mathcal{Q}_+$, any edge from some $i' \in \mathcal{A}$ to some $m' \in \mathcal{M}$ on this path is marked with the direction it is traversed, $i' \to m'$ or $m' \to i'$. An edge can be marked with both directions if it is traversed $i' \to m'$ on one path, but $m' \to i'$ on another. Denote the resulting directed graph by $\overrightarrow{\mathcal{G}}$.

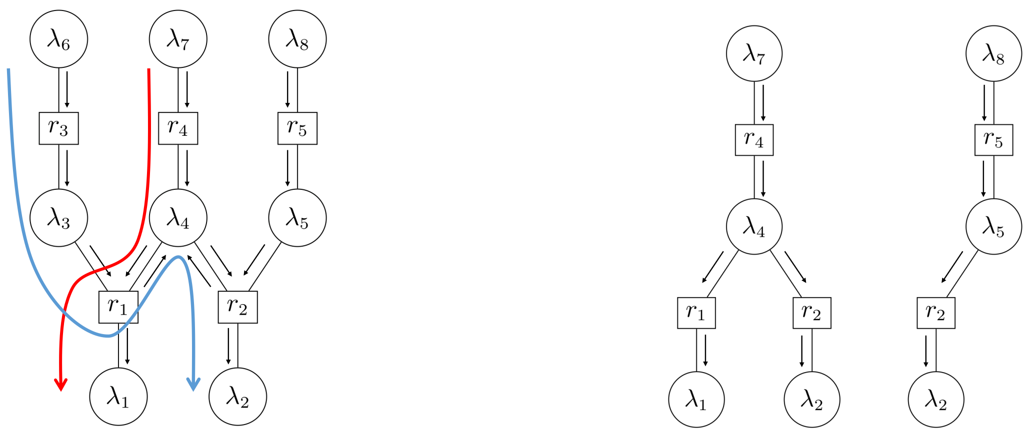

Figure 8: An example of a directed graph $\overrightarrow{\mathcal{G}}$. In this network, $\mathcal{Q}_+ = \{6, 7, 8\}$ and $\mathcal{U}_0 = \{1, 2\}$. (LEFT) The edge between match-1 and queue-4 is marked with both directions, since it is traversed on both paths $7 \to 1$ (marked in red) and $6 \to 2$ (marked in blue). (RIGHT) The subtrees rooted at queue-7 ($\mathcal{T}_7$) and queue-8 ($\mathcal{T}_8$), respectively.

Lemma C.2. *For each match $m \in \mathcal{M}$, there is a unique queue $i(m) \in \mathcal{A}(m)$, such that the edge between m and $i(m)$ has a single direction in the $\overrightarrow{\mathcal{G}}$, which is directed from m to $i(m)$.*

Given $\overrightarrow{\mathcal{G}}$, we say that a path from $j \in \mathcal{A}$ to $i \in \mathcal{U}_0$ is *uniquely directed* if for any match $m \in \mathcal{M}$ on this path, the only outgoing edge from m is to $i(m)$. For example in Figure 8, the path from queue-7 to queue-1 is uniquely directed, whereas the path from queue-6 to queue-2 is not.

Based on these uniquely directed paths, we build *subtrees* as follows. For each $i \in \mathcal{A}$, we let $\mathcal{T}_i$ be the subtree rooted at i, where $\mathcal{T}_i$ is the union of all uniquely directed paths starting from i. $\mathcal{T}_i$, by construction, is a two-way tree: for each match$-m$ in the subtree, we have $\mathcal{A}(m) = 2$; see Figure 8 for an example of a subtree. Let $\mathcal{A}(\mathcal{T}_i)$ be the set of queues in $\mathcal{T}_i$.

Let $d(i,j)$ be the length of the directed path from $i \in \mathcal{A}$ to $j \in \mathcal{A}$ in $\overrightarrow{\mathcal{G}}$. For each $i \in \mathcal{A}$, we then define the *surplus vector* $y^i \in \{-1,0,1\}^n$ as follows:

$$(y^i)_j := \begin{cases} 0, & \text{if } j \in \mathcal{A} \backslash \mathcal{A}(\mathcal{T}_i), \\ 1, & \text{if } d(i,j) \equiv 0 \pmod 4, \\ -1, & \text{if } d(i,j) \equiv 2 \pmod 4. \end{cases}$$

Note that since $d(i,i) = 0$, in particular, we have $(y^i)_i = 1$. Finally, we identify the surplus vector for each $m \in \mathcal{M}$ with the vector $y^{i(m)}$:

$$y^m := y^{i(m)} \text{ for all } m \in \mathcal{M}.$$

Proof of Theorem 5.1. Following the arguments on the structure of the optimal solution of (SPP) in §4.2, assume that $\mathcal{M}_+ = \{1, 2, ..., d - \varrho\}$ and $\mathcal{Q}_+ = \{d - \varrho + 1, d - \varrho + 2, ..., n\}$, where we let $\varrho := |\mathcal{M}_0|$. Then the optimal basis matrix takes the form

$$\mathcal{B} = \begin{bmatrix} M^0 & \mathbf{0} \\ M^+ & I \end{bmatrix},$$

where M^0 has the rows of M corresponding to the queues in $\mathcal{Q}_0$, M^+ has the remaining $n-d+\varrho$ rows, and $\mathcal{B}$ has the columns corresponding to $\mathcal{M}_+$ and $\mathcal{Q}_+$ in order; I here is an $(n-d+\varrho) \times (n-d+\varrho)$ identity matrix, and $\mathbf{0}$ is a $(d-\varrho) \times (n-d+\varrho)$ zero matrix.

Being the basis matrix, $\mathcal{B}$ is invertible and we claim that $Y = \mathcal{B}^{-1}$ has the following form

$$\mathcal{B}^{-1} = Y := \begin{bmatrix} Y^0 & \mathbf{0} \\ Y^+ & I \end{bmatrix},$$

where $[Y^0,\ \mathbf{0}]$ is a $(d-\varrho) \times n$ matrix and $[Y^+,\ I]$ is an $(n-d+\varrho) \times n$ matrix, where

1. m^{th} row of $[Y^0,\ 0]$ is y^m for each $m \in \mathcal{M}_+$, and
2. j^{th} row of $[Y^+,\ I]$ is $y^{d-\varrho+j}$ for each $d - \varrho + j \in \mathcal{Q}_+$.

In turn, the optimal solution of (SPP) can be written as

$$\begin{bmatrix} z^*_{\mathcal{M}_+} \\ s^*_{\mathcal{Q}_+} \end{bmatrix} = \mathcal{B}^{-1}\lambda = Y\lambda,$$

which implies $z_m^* = y^m\lambda > 0$ for all $m \in \mathcal{M}_+$, and $s_j^* = y^j\lambda > 0$ for all $j \in \mathcal{Q}_+$, where strict inequalities follow from the non-degeneracy of (z^*, s^*).

To prove the claim above, note that

$$Y\mathcal{B} = \begin{bmatrix} Y^0 & \mathbf{0} \\ Y^+ & I \end{bmatrix} \begin{bmatrix} M^0 & \mathbf{0} \\ M^+ & I \end{bmatrix} = \begin{bmatrix} M^0 & \mathbf{0} \\ M^+ & I \end{bmatrix} \begin{bmatrix} Y^0 & \mathbf{0} \\ Y^+ & I \end{bmatrix} = I, \tag{23}$$

is equivalent (and hence implied) by the following two properties:

1. $[Y^0,\ \mathbf{0}]' \begin{bmatrix} M^0 \\ M^+ \end{bmatrix} = I$, or $y^m M = e_m$ for all $m \in \mathcal{M}_+$, where y^m is the m^{th} row of Y, and e_m is the m^{th} row of I,

2. $Y^+M^0 + M^+ = \mathbf{0}$,

which we will prove next. Take any two matches $m, m' \in \mathcal{M}_+$, and consider the subtree $\mathcal{T}_{i(m)}$. If m' is included in $\mathcal{T}_{i(m)}$, then the queues $j \in \mathcal{A}(\mathcal{T}_{i(m)}) \cap \mathcal{A}(m')$ appear in the vector $y^{i(m)} = y^m$ with opposite signs. If m' is not included in $\mathcal{T}_{i(m)}$, then the queues that are participating in m' have 0 values in the vector y^m. Finally, since $(y^m)_{i(m)}$ has a positive sign, we have $y^m M = e_m$, and the first property holds.

For the second property, note that for each $j \in \mathcal{Q}_+$, the vector $y^j M^0$ has -1 for each match m that j participates in, and 0 otherwise. Thus, $Y^+M^0 + M^+ = \mathbf{0}$, and property 2 holds as well. $\square$

Proof of Lemma C.1. We use induction on the number of queue vertices n.

Basis. Assume that $n = 2$. Then $\mathcal{G}$ is unique with $d = 1$, and both queues correspond to a leaf in $\mathcal{G}$. Thus, $\mathcal{G}$ contains $n - d + 1 = 2$ leaves.

Inductive step. Assume that the induction hypothesis holds for all $\mathcal{G}$ with n queue vertices, $n \geq 2$. Consider $\mathcal{G}$ with $n+1$ queue vertices. Since $\mathcal{G}$ is connected and acyclic, there exists a queue vertex v that participates in exactly one matching, i.e., v is a leaf in $\mathcal{G}$. Otherwise, since all queue and match vertices have degree at least 2, there would exist a cycle.

Denote the unique match vertex that v participates in $\mathcal{G}$ by m. First, assume that the number of queues participating in m is exactly 2. Denote the other queue vertex participating in m by v'. Remove v and m from $\mathcal{G}$ and let $\mathcal{G}' = \mathcal{G} - \{v, m\}$ be the residual graph, which is clearly a matching network. By the induction hypothesis, $\mathcal{G}'$ contains at least $(n-1) - (d-1) + 1 = n - d + 1$ leaves. If v' is not a leaf in $\mathcal{G}'$, then adding back v and m increases the number of leaves by 1. Thus, $\mathcal{G}$

contains at least $n-d+2$ leaves. If v' is a leaf in $\mathcal{G}'$, then adding back v and m does not change the number of leaves. Thus, $\mathcal{G}$ contains at least $n-d+1$ leaves.

Similarly, if the number of queues participating in m is at least 3, then removing v from $\mathcal{G}$ results in a matching network with $n-1$ queue vertices. By the induction hypothesis, the residual graph $\mathcal{G}'$ contains at least $(n-1)-d+1=n-d$ leaves. Thus, adding back v increases the number of leaves by 1, and $\mathcal{G}$ contains at least $n-d+1$ leaves. Thus, the induction hypothesis holds for all $\mathcal{G}$ with $n+1$ queue vertices.

Finally, since $|\mathcal{Q}_+| = n-d$, we have $|\mathcal{U}_0| \geq 1$. $\square$

Proof of Lemma C.2. We first start with proving the following claim: $\mathcal{G}$ satisfies (i) all matches in $\mathcal{G}$ are two-way, i.e., $|\mathcal{A}(m)| = 2$ for all $m \in \mathcal{M}$, or (ii) $|\mathcal{U}_0| = 1$ if and only if all the edges in $\overrightarrow{\mathcal{G}}$ have a single direction. The necessity part is immediate. Note that if $\mathcal{G}$ only contains two-way matches, then we have $|\mathcal{Q}_+| = n-(n-1) = 1$. Thus, by the construction of $\overrightarrow{\mathcal{G}}$, all the edges are assigned with a single direction. Similarly if $|\mathcal{U}_0| = 1$, all the edges in $\overrightarrow{\mathcal{G}}$ have a single direction by construction (otherwise, $\overrightarrow{\mathcal{G}}$ would contain an undirected cycle). For the sufficiency part, assume to the contrary that there exists $m \in \mathcal{M}$ such that $|\mathcal{A}(m)| \geq 3$ and $|\mathcal{U}_0| \geq 2$. Since $n-d \geq 2$, we also have $|\mathcal{Q}_+| \geq 2$. Let $v_1, v_2 \in \mathcal{Q}_+$ and $u_1, u_2 \in \mathcal{U}_0$. By the construction of $\overrightarrow{\mathcal{G}}$, there is a directed path from v_1 to u_1, v_1 to u_2, v_2 to u_1 and v_2 to u_2 in $\overrightarrow{\mathcal{G}}$. If all the edges have a single direction, then there exists a cycle in $\mathcal{G}$ that contains v_1, v_2, u_1 and u_2, which is a contradiction.

Now let $\mathcal{E}$ be the set of all edges in $\overrightarrow{\mathcal{G}}$, which are assigned with both directions. Then removing $\mathcal{E}$ from $\overrightarrow{\mathcal{G}}$, decomposes $\overrightarrow{\mathcal{G}}$ into (possibly) multiple connected components that satisfy either (i) or (ii) in the above claim. In both cases, for each match m, there is a unique queue$-i$ in its component, such that the edge between m and i has a single direction, which is directed from m to i. $\square$

Proof of Theorem 5.2. Let us argue that we can construct a match value vector r' such that the optimal basis of (SPP) is unchanged, and all the coefficients of the objective function in (7) are equal to 1. Then Theorem 5.2 immediately follows from the proof of Theorem 3.1, since under the new match value vector r', the policy D' simply resolves (7) at each decision period t_k. It is straightforward to check that the desired match value vector is the following:

$$r'_m := \begin{cases} 1, & \text{if } \mathcal{A}(m) \cap \mathcal{Q}_+ \neq \emptyset, \\ 2, & \text{if } \mathcal{A}(m) \cap \mathcal{Q}_+ = \emptyset \ \text{ and } |\mathcal{A}(m)| = 2, \\ |\mathcal{A}(m)|, & \text{otherwise.} \end{cases}$$

$\square$

D Proof of the upper bound in Theorem 3.1

Recalling that under the exhaustive resolving policy, agents of type $i \in \mathcal{Q}_+$ are removed post-match if not used, it is straightforward to verify that the discrete time Markov chain $(Q^{t_k},\ k \in \mathbb{N})$ is irreducible and aperiodic on its state space

$$\mathcal{S} := \{Q \in \mathbb{Z}^n_{\geq 0} : Q_j \leq \tau \text{ for all } j \in \mathcal{Q}_+\}.$$

Let $F := \{Q \in \mathcal{S} : h^*(Q) \leq B\}$. Since $\theta = \sum_{m \in \mathcal{M}_+} r_m y^m \geq 0$, and in particular, $\theta_i > 0$ for all $i \in \mathcal{Q}_0$, F is clearly finite. Then the drift property (8) implies that the Markov chain is positive recurrent; e.g., see (Robert, 2003, Theorem 8.6). It also follows from Lemma 4.2 that under the Markov chain's unique stationary distribution, which we denote by π, we have

$$\mathbb{E}_\pi[\mathcal{L}(Q^0)] \leq \frac{2c_1}{\gamma} e^{c_2 \tau}, \tag{24}$$

where $Q^0 \sim \pi$. Since $\tau = \lceil \kappa \epsilon^{-1} \rceil$, by Jensen's inequality we have

$$\mathbb{E}_\pi[h^*(Q^0)] = \mathcal{O}(\epsilon^{-1}). \tag{25}$$

We next show that with the initial state $q = \mathbf{0}$, (25) holds for all $t > 0$. Let $f^0(q) := \frac{\gamma}{4}\mathcal{L}(q)$. Then (9) can be rewritten as

$$\mathbb{E}_q[\mathcal{L}(Q^\tau)] \leq \left(1 - \frac{\gamma}{4}\right)\mathcal{L}(q) - f^0(q), \text{ if } q \in F^c.$$

It then follows from (Meyn and Tweedie, 1992, Theorem 6.2) (with $\varepsilon = 1$ and $r = (1-\gamma/4)^{-1}$ there) that

$$\mathbb{E}_q\left[\sum_{k=1}^{\tau_F} r^k f^0(Q^{t_{k-1}})\right] \leq \begin{cases} \mathcal{L}(q),\ q \in F^c, \\ (1-\gamma/4)^{-1}(f^0(q) + \mathbb{E}[\mathcal{L}(Q^\tau)]),\ q \in \mathcal{S}, \end{cases}$$

where $\tau_F := \inf\{k \geq 1 : Q^{t_k} \in F\}$.

Because of the Lipschitz continuity of $h^*(\cdot)$, we have $\mathcal{L}(Q^\tau) \leq \mathcal{L}(q)e^{c\tau}$ for some $c > 0$. Setting the initial state $q = \mathbf{0}$, we then have a sufficiently large constant $\alpha > 0$ such that $\mathbb{E}_{\mathbf{0}}\left[\sum_{k=1}^{\tau_F} r^k f^0(Q^{t_{k-1}})\right] \leq \alpha e^{\alpha\tau}$.

Applying (Meyn and Tweedie, 1992, Theorem 6.1) (with $m = 1$ there), we conclude that for all

$k \geq 1$, we have

$$|\mathbb{E}_{\mathbf{0}}[f^0(Q^{t_k})] - \mathbb{E}_\pi[f^0(Q^0)]| = \frac{\gamma}{4} |\mathbb{E}_{\mathbf{0}}[\mathcal{L}(Q^{t_k})] - \mathbb{E}_\pi[\mathcal{L}(Q^0)]| \leq \alpha e^{\alpha\tau},$$

for a redefined constant α. Combining this with (24), we conclude that for all $k \geq 1$, we have

$$|\mathbb{E}_{\mathbf{0}}[f^0(Q^{t_k})]| \leq \alpha e^{\alpha\tau},$$

for a redefined constant α. Then by Jensen's inequality, we have

$$\mathbb{E}_{\mathbf{0}}[h^*(Q^{t_k})] = \mathcal{O}(\epsilon^{-1}) \text{ for all } k \geq 1.$$

Finally, note that for $t \in (t_k, t_{k+1})$, the Lipschitz continuity of $h^*(\cdot)$ implies that, because $|Q^t - Q^s| \leq |t-s|$, we have

$$\mathbb{E}_{\mathbf{0}}[h^*(Q^t)] \leq \upsilon(\mathbb{E}_{\mathbf{0}}[h^*(Q^{t_k})] + \tau) = \mathcal{O}(\epsilon^{-1}) \text{ for all } k \geq 1,$$

for some constant $\upsilon > 0$. Using the optimality test (Lemma 4.1), this proves the upper bound in Theorem 3.1.

We note here that removing agents of type $i \in \mathcal{Q}_+$ under the exhaustive resolving policy is without loss of generality. It is immediate to see that if one imposes any finite buffer size (the buffer size is τ in the proof) for the under-demanded queues, then the proof does not change since the set F is still finite. Therefore, if one focuses on finite horizon (say T) value maximization, then one can set the buffer size to be T. □

E Proof of the lower bound in Theorem 3.1

Throughout the proof, we use superscripts on expectations and probabilities to make explicit the dependence on ϵ. Assume to the contrary that there exists a matching policy, which has

$$\mathbb{E}^\epsilon \left[\sum_{i \in \mathcal{Q}_0} Q_i^t \right] = o(\epsilon^{-1}) \text{ for all } t > 0.$$

Markov's inequality then implies that for all $t > 0$, $\mathbb{P}^\epsilon\{\sum_{i \in \mathcal{Q}_0} Q_i^t \geq \epsilon^{-1}\} = o(\epsilon) \to 0$ as $\epsilon \downarrow 0$. In particular, given $t > 0$ and $0 < \delta_1 < 1$, for all sufficiently small $\epsilon > 0$, we have

$$\mathbb{P}^\epsilon \left\{ \sum_{i \in \mathcal{Q}_0} Q_i^t \leq \epsilon^{-1} \right\} \geq 1 - \delta_1 > 0. \tag{26}$$

For ease of exposition, let us fix some $t_0 > 0$, and assume that $\sum_{i \in \mathcal{Q}_0} Q_i^{t_0} \leq \epsilon^{-1}$ throughout the analysis. We will argue that this is without loss of generality at the end of the proof. First consider the case when the general position gap is determined by some active match, i.e., $\epsilon = y^m \lambda$ for some $m \in \mathcal{M}_+$. Consider the process $I^s := y^m Q^s$ for all $s \geq t_0$. Then we have

$$I^s = I^{t_0} + y^m A^{t_0,s} - D_m^{t_0,s^-} \text{ for all } s \geq t_0,$$

where for any $t > 0$ and $s > t$, we define $A^{t,s} := A^s - A^t$ and $D_m^{t,s^-} := D_m^{s^-} - D_m^t$. Since $D_m^{t,s^-} \geq 0$, we have

$$I^s \leq I^{t_0} + y^m A^{t_0,s} \text{ for all } s \geq t_0. \tag{27}$$

Define a stopping time

$$\nu := \inf\{t_0 + u : I^{t_0+u} \leq -\epsilon^{-1}, u \geq 0\}.$$

We claim, and will later prove, that given $0 < \delta_2 < \frac{1}{2}$, there exists $B > 0$ (that does not depend on ϵ) such that

$$\mathbb{P}^\epsilon\{\nu \leq t_0 + B/\epsilon^2\} \geq 1 - 2\delta_2 > 0, \tag{28}$$

for all sufficiently small $\epsilon > 0$.

Next, we use the fact that if the network is non-trivial, then y^m contains at least one negative entry. To see this, let $\mathcal{N}(m)$ be the set of all active matches that share a queue with m, i.e., $\mathcal{N}(m) := \{m' \in \mathcal{M}_+ : \mathcal{A}(m) \cap \mathcal{A}(m') \neq \emptyset\}$. Since the network is non-trivial, any $i \in \mathcal{A}(m)$ participates in at least two active matches in $\mathcal{N}(m)$. Let $c_{m'}$ be the column of M corresponding to $m' \in \mathcal{M}_+$. Assume to the contrary that $(y^m)_i \geq 0$ for all $i \in \mathcal{A}$. Since $y^m \cdot c_{m'} = 0$ for all $m' \in \mathcal{N}(m)$, we must have $(y^m)_i = 0$ for all $i \in \mathcal{A}(m) \cap \mathcal{A}(m')$, which implies that $(y^m)_i = 0$ for all $i \in \mathcal{A}(m)$. But this contradicts to the fact that $y^m \cdot c_m = 1$. Thus, y^m contains at least one negative entry.

Let $\mathcal{S}^+$ be the set of all indices of y^m that has a positive entry, and let $\mathcal{S}^-$ be the set of all indices of y^m that has a negative entry. Since $I^s = y^m Q^s = \sum_{i \in \mathcal{S}^+} (y^m)_i Q_i^s + \sum_{i \in \mathcal{S}^-} (y^m)_i Q_i^s$

$\leq -\epsilon^{-1}$ implies that $-\sum_{i\in\mathcal{S}^-}(y^m)_i Q_i^s \geq \epsilon^{-1}$, we have $-\sum_{i\in\mathcal{S}^-}(y^m)_i Q_i^\nu \geq \epsilon^{-1}$ on the event

$$\mathcal{E} := \{\nu \leq t_0 + B/\epsilon^2\}.$$

Since $-\sum_{i\in\mathcal{S}^-}(y^m)_i Q_i^s \geq -\sum_{i\in\mathcal{S}^-}(y^m)_i Q_i^{t_0} - y^m A^{t_0,s}$ for all $s \geq t_0$ by (27), we have

$$-\sum_{i\in\mathcal{S}^-}(y^m)_i Q_i^{t_0+B/\epsilon^2} \geq \inf_{\nu\leq u\leq t_0+B/\epsilon^2} \frac{1}{\epsilon} - y^m A^{\nu,u},$$

on the event $\mathcal{E}$. In particular,

$$\begin{aligned}\mathbb{P}^\epsilon\left\{-\sum_{i\in\mathcal{S}^-}(y^m)_i Q_i^{t_0+B/\epsilon^2} \geq \frac{1}{\epsilon}, \mathcal{E}\right\} &\geq \mathbb{P}^\epsilon\left\{\inf_{\nu\leq u\leq t_0+B/\epsilon^2}\left(\frac{1}{\epsilon} - y^m A^{\nu,u}\right) \geq \frac{1}{\epsilon}, \mathcal{E}\right\} \\ &\geq \mathbb{P}^\epsilon\left\{\inf_{0\leq u\leq B/\epsilon^2}\left(\frac{1}{\epsilon} - y^m A^{u}\right) \geq \frac{1}{2\epsilon}, \mathcal{E}\right\}. \end{aligned} \tag{29}$$

The process $S^u := (-y^m A^u :\ u \in \mathbb{Z}_{\geq 0})$ is a lazy random walk on $\mathbb{Z}$, with transition probabilities $\mathbb{P}\{S^{u+1} = S^u + 1\} = -\sum_{i\in\mathcal{S}^-}(y^m)_i\lambda_i$ and $\mathbb{P}\{S^{u+1} = S^u - 1\} = \sum_{i\in\mathcal{S}^+}(y^m)_i\lambda_i$, which yields $\mathbb{E}[S^{u+1} - S^u | S^u] = -y^m\lambda = -\epsilon$. Donsker's theorem (Donsker, 1951) (e.g., see also (Whitt, 2002, Page 102)) guarantees that

$$\widehat{I}^\epsilon(u) := \epsilon(-y^m A^{\lceil u/\epsilon^2\rceil}) \Rightarrow \mathcal{W}$$

where $\mathcal{W}$ is a Brownian motion with drift -1 and squared diffusion coefficient $\sigma^2 = \sum_{i\in\mathcal{S}^+} y^m\lambda_i - \sum_{i\in\mathcal{S}^-} y^m\lambda_i$. Moreover, the convergence is uniform over compact intervals. Using the continuity of the infimum map (Whitt, 2002, Section 13.4), we have

$$\mathbb{P}^\epsilon\left\{\inf_{0\leq u\leq B/\epsilon^2}\left(\frac{1}{\epsilon} - y^m A^u\right) \geq \frac{1}{2\epsilon}\right\} \to \mathbb{P}\left\{\inf_{0\leq u\leq B}\left(1 + \widehat{I}(u)\right) \geq \frac{1}{2}\right\} \geq \delta_3, \tag{30}$$

for some $\delta_3 > 0$. Finally, using (29) and (30), choosing δ_2 sufficiently small (and then B large) yields

$$\mathbb{P}^\epsilon\left\{-\sum_{i\in\mathcal{S}^-}(y^m)_i Q_i^{t_0+B/\epsilon^2} \geq \frac{1}{\epsilon}, \mathcal{E}\right\} \geq \mathbb{P}^\epsilon\left\{\inf_{0\leq u\leq B/\epsilon^2}\left(\frac{1}{\epsilon} - y^m A^u\right) \geq \frac{1}{2\epsilon}\right\} + \mathbb{P}^\epsilon\{\mathcal{E}\} - 1 \geq \frac{\delta_3}{2} - 2\delta_2 > \delta_4$$

for some $\delta_4 > 0$. We conclude that

$$\mathbb{E}^\epsilon\left[-\sum_{i\in\mathcal{S}^-}(y^m)_i Q_i^{t_0+B/\epsilon^2}\right] \geq \frac{1}{\epsilon}\,\mathbb{P}^\epsilon\left\{-\sum_{i\in\mathcal{S}^-}(y^m)_i Q_i^{t_0+B/\epsilon^2} \geq \frac{1}{\epsilon}, \mathcal{E}\right\} \geq \frac{\delta_4}{\epsilon},$$

which is a contradiction to the assumption that $\mathbb{E}^\epsilon\left[\sum_{i\in\mathcal{Q}_0} Q_i^t\right] = o(\epsilon^{-1})$ for all $t > 0$, and for all $\epsilon > 0$ sufficiently small.

So far, the analysis assumes that $\sum_{i\in\mathcal{Q}_0} Q_i^{t_0} \leq \epsilon^{-1}$ for some fixed t_0. However, this assumption is without loss of generality, since the choice of δ_1 in (26) is arbitrary. It remains to establish (28). Since

$$\nu \leq \nu_0 := \inf\{s \geq t_0 : I^{t_0} + y^m A^{t_0,s} \leq -\epsilon^{-1}\},$$

we will study ν_0 instead. Under any non-anticipating matching policy, the law of $y^m A^{t_0,s}$ is independent of I^{t_0}, and the process is a random walk with upward probability $-\sum_{i\in\mathcal{S}^-}(y^m)_i\lambda_i$ and downward probability $\sum_{i\in\mathcal{S}^+}(y^m)_i\lambda_i = -\sum_{i\in\mathcal{S}^-}(y^m)_i\lambda_i + \epsilon$. We use again the convergence of

$$\widehat{I}^\epsilon(u) := \epsilon(I^{t_0} - y^m A^{\lceil u/\epsilon^2\rceil}).$$

Note that our initialization t_0 is such that $\epsilon I^{t_0} \Rightarrow 0$. Hence, $\widehat{I}^\epsilon(u)$ converges, as before, to a Brownian motion starting at 0. From continuity of the first passage time map (Whitt, 2002, Section 13.6.3), we have

$$\epsilon^2(\nu_0 - t_0) \Rightarrow \widehat{\nu} := \inf\{s \geq 0 : \mathcal{W}(s) \leq -1\}.$$

It is known that $\mathbb{P}\{\widehat{\nu} < \infty\} > 0$ so that given $0 < \delta_2 < \frac{1}{2}$, there exists $B > 0$ (that does not depend on ϵ) such that $\mathbb{P}\{\widehat{\nu} \leq B\} \geq 1 - \delta_2$. In turn, by the weak convergence of $\epsilon^2(\nu_0 - t_0)$, $\mathbb{P}^\epsilon\{\nu_0 - t_0 \leq B/\epsilon^2\} \geq 1 - 2\delta_2 > 0$ for all $\epsilon > 0$ sufficiently small, as stated.

So far, we considered the effect of $\epsilon = z_m^*$ for some $m \in \mathcal{M}_+$, which determined the general position gap. To cover the case when the general position gap is determined by a slack variable, now we show that the case when $\epsilon = s_j^*$ for some $j \in \mathcal{Q}_+$ has a similar implication.

Similar to the previous case, y^j must contain at least one negative entry, since $y^j M = 0$ and $y^j\lambda = s_j^* > 0$. Note that $y^j M = 0$ also implies that $y^j Q^t = y^j A^t$ for all $t > 0$. Let $\mathcal{S}^+$ be the set of all indices of y^j that has a positive entry, and let $\mathcal{S}^-$ be the set of all indices of y^j that has a negative entry. Since $y^j Q^t \leq -\epsilon^{-1}$ implies that $-\sum_{i\in\mathcal{S}^+}(y^j)_i Q_i^t \geq \epsilon^{-1}$, we have

$$\mathbb{P}^\epsilon\{-\sum_{i\in\mathcal{S}^+}(y^j)_i Q_i^t \geq \epsilon^{-1}\} \geq \mathbb{P}^\epsilon\{y^j A^t \leq \epsilon^{-1}\} \text{ for all } t > 0.$$

Notice that $\mathbb{E}^\epsilon[y^j A^t] = t\epsilon = ts_j^*$. Redefining the process $I^t := -y^j A^t$, we have as before that $\widehat{I} \Rightarrow \mathcal{W}$ where $\mathcal{W}$ is a Brownian motion with drift -1. In particular, there exists $\delta, s > 0$ such that $\mathbb{P}\{\mathcal{W}(s) \leq -1\} \geq \delta$. Similarly, for any initialization t_0, there exists $t \geq t_0$ such that, for all $\epsilon > 0$ sufficiently small, we have

$$\mathbb{P}^\epsilon\left\{-\sum_{i\in\mathcal{S}^+}(y^j)_i Q_i^t \geq \frac{1}{\epsilon}\right\} \geq \frac{\delta}{2},$$

which implies $\mathbb{E}^\epsilon[-\sum_{i\in\mathcal{S}^+}(y^j)_i Q_i^t] \geq \frac{\delta}{2}\epsilon^{-1}$.

Implication to lower bound. So far, the arguments imply that over-demanded queues (queues in $\mathcal{Q}_0$) cannot be made permanently small. It remains to prove that $\sup_{t>0}(\mathcal{R}^{*,t} - \mathcal{R}^{D,t}) \geq \gamma\epsilon^{-1}$.

We will utilize the following lemma, which argues that $\mathcal{R}^{*,t}$, the optimal value at time t, is constant away from the optimal value of (SPP) when the right-hand side is scaled by t. This follows readily from the assumed non-degeneracy of (SPP) and Lipschitz continuity of (SPP) in the right-hand side.

Lemma E.1. *Suppose that* **GP** *holds. Let (z^*, s^*) be the unique optimal solution of* (SPP). *Then $(r \cdot z^*)t - \mathcal{R}^{*,t} \leq \Lambda$ for all $t > 0$, where $\Lambda > 0$ is a constant that may depend on n, d, M and r (but not on λ or ϵ).*

Note that a policy that has the state of queues Q^t at time t (such that $\mathbb{E}[\sum_{i\in\mathcal{Q}_0} Q_i^t] \geq \gamma\epsilon^{-1}$), can collect at most the value given by the following LP upper bound

$$\beta^*(Q^t, A^t) := \begin{array}{ll} \max & r \cdot z \\ \text{s.t.} & Mz \leq A^t - Q^t \\ & z \in \mathbb{Z}^d_{\geq 0}. \end{array}$$

This linear program is concave in its right-hand side so that by Jensen's inequality, we have $\mathcal{R}^{D,t} \leq \mathbb{E}^D[\beta^*(Q^t, A^t)] \leq \beta^*(\mathbb{E}^D[Q^t], \lambda t)$. Per the derivation in Section 4, we can rewrite the above linear program as

$$\beta^*(\mathbb{E}^D[Q^t], \lambda t) = \begin{array}{ll} \max & \sum_{m \in \mathcal{M}_+} r_m y^m (\lambda t - \mathbb{E}^D[Q^t]) - \sum_{i \in \mathcal{Q}_0} \sum_{m \in \mathcal{M}_+} (r_m y^m)_i s_i \\ \text{s.t.} & z_m + y^m s = y^m(\lambda t - \mathbb{E}^D[Q^t]) \text{ for all } m \in \mathcal{M}_+ \\ & y^i s = y^j (\lambda t - \mathbb{E}^D[Q^t]) \text{ for all } j \in \mathcal{Q}_+ \\ & z \in \mathbb{Z}^d_{\geq 0}, s \in \mathbb{Z}^n_{\geq 0}. \end{array}$$

Recall that $\theta_i = (\sum_m r_m y^m)_i > 0$ for all $i \in \mathcal{Q}_0$. Since $\mathbb{E}[\sum_{i \in \mathcal{Q}_0} Q^t_i] \geq \gamma \epsilon^{-1}$, we have

$$\mathcal{R}^{D,t} \leq \beta^*(\mathbb{E}^D[Q^t], \lambda t) \leq \sum_{m \in \mathcal{M}_+} r_m y^m \lambda t - \Omega(\epsilon^{-1}) \leq \mathcal{R}^{*,t} - \Omega(\epsilon^{-1}),$$

where the last inequality follows from Lemma E.1. It only remains to prove Lemma E.1. Using standard arguments, for all t sufficiently large, we have

$$\mathbb{P}\{\|A^t - \lambda t\|_1 \geq t^{3/4}\} \leq c_1 e^{-c_2 t^{1/4}},$$

for some constants $c_1, c_2 > 0$. Note that on the event $\|A^t - \lambda t\|_1 < t^{3/4}$, we have for all t sufficiently large that $y^m A^t > 0$ for all $m \in \mathcal{M}_+$. Then the optimal solution of (SPP) with the right-hand side A^t has $z^*_m(A^t) = y^m A^t$ for all $m \in \mathcal{M}_+$ and $z^*_m(A^t) = 0$ for all $m \in \mathcal{M}_0$. Outside of this event, the optimality gap is at most $\bar{r}t$, where $\bar{r} = \max_{m \in \mathcal{M}} r_m$. Thus, we have

$$(r \cdot z^*)t - \mathcal{R}^{*,t} \leq \mathcal{O}(1) + \bar{r} t c_1 e^{-c_2 t^{1/4}} = \mathcal{O}(1).$$

$\square$