The Segmentation-Thickness Tradeoff in Online Marketplaces

REZA ALIJANI, Duke University, USA
SIDDHARTHA BANERJEE, Cornell University, USA
SREENIVAS GOLLAPUDI, Google Research, USA
KOSTAS KOLLIAS, Google Research, USA
KAMESH MUNAGALA, Duke University, USA

A core tension in the operations of online marketplaces is between segmentation (wherein platforms can increase revenue by segmenting the market into ever smaller sub-markets) and thickness (wherein the size of the sub-market affects the utility experienced by an agent). An important example of this is in dynamic online marketplaces, where buyers and sellers, in addition to preferences for different matches, also have finite patience (or deadlines) for being matched. We formalize this trade-off via a novel optimization problem that we term as 'Two-sided Facility Location': we consider a market wherein agents arrive at nodes embedded in an underlying metric space, where the distance between a buyer and seller captures the quality of the corresponding match. The platform posts prices and wages at the nodes, and opens a set of virtual clearinghouses where agents are routed for matching. To ensure high match-quality, the platform imposes a distance constraint between an agent and its clearinghouse; to ensure thickness, the platform requires the flow to any clearinghouse be at least a pre-specified lower bound. Subject to these constraints, the goal of the platform is to maximize the social surplus subject to weak budget balance, i.e., profit being non-negative. Our work characterizes the complexity of this problem by providing both hardness results as well as algorithms for this setting; in particular, we present an algorithm that for any constant $\epsilon > 0$ yields a $(1 + \epsilon)$ approximation for the gains from trade, while relaxing the match quality (i.e., maximum distance of any match) by a constant factor.

1 INTRODUCTION

Online marketplaces have transformed the economic landscape of the modern world. Many of today's most important companies are platforms facilitating trade between agents: both for goods (Amazon, eBay), and increasingly, services (Lyft, Uber, Airbnb, etc). These platforms enable fine-grained monitoring, and greater control via pricing [9], terms of trade, recommendation and directed search [8], etc. The challenge of harnessing this increase in data and control has led to a growing literature in online marketplace design.

The basic algorithmic challenge facing a marketplace platform can be summarized as follows: it must decide *which buyer should match to which seller, at what time, and for what price and wage,*

Authors' addresses: Reza Alijani Duke University, USA, alijani@cs.duke.edu; Siddhartha Banerjee Cornell University, USA, sbanerjee@cornell.edu; Sreenivas Gollapudi Google Research, USA, sgollapu@google.com; Kostas Kollias Google Research, USA, kostaskollias@google.com; Kamesh Munagala Duke University, USA, kamesh@cs.duke.edu.

in order to maximize some desired objective. Though this description has much in common with other mechanism-design problems in settings such as resource allocation and monopolist pricing problems, there are a few features unique to two-sided markets which makes them particularly challenging. One notable such feature is the tension arising from the need for ensuring *budget balance* (i.e., non-negative profits) – in particular, the problem of maximizing social welfare while respecting budget-balance is at the heart of the famous Myerson-Satterthwaite impossibility result [30], and has received much attention in recent work in approximate mechanism-design [10, 12]. At its heart, however, lies the tension between balancing earnings and costs which is necessitated when one moves from a one-sided to a two-sided setting.

Our work studies another unique feature of two-sided markets – the tension between market segmentation and market thickness. At a high level, segmenting agents on both sides of a market into smaller sub-markets leads to better welfare and higher revenues in many settings; on the other hand, the quality of experience of an agent in a market improves when more agents are present on the other side. Ensuring both turns out to present severe computational challenges for the market maker in terms of pricing and scheduling policies. In particular, *our core algorithmic insight is that resolving the segmentation-thickness tradeoff in two-sided markets is intimately related to a classical 'hard' problem in combinatorial optimization: facility location.*

Before presenting this, however, we first present a natural setting that naturally exhibits the segmentation-thickness trade-off, and which forms the primary motivation for our work.

Dynamic Marketplaces with QoS Guarantees: Our motivating application is the design of a dynamic marketplace platform, and in particular, policies for pricing both buyers and sellers, and scheduling feasible matches, while providing quality of service (QoS) guarantees to the agents. To understand how this results in a segmentation-thickness tradeoff, it is useful to build up the setting piece-by-piece.

Pricing, matching and market segmentation: Consider first a simple two-sided market setting, where buyers and sellers are characterized by a *public type* and there is a single service they are interested in trading. Buyers have values, and sellers have costs, and these are drawn from known type-dependent distributions. For instance, on a platform like AirBnB, the type of a listing can represent some combination of geographic location and quality of room; on a freelancing platform, the type of a consultant can capture their skill set and the type of a task can capture its requirements; and so on.

Suppose we now incorporate dynamics as follows. Buyers and sellers belong to discrete types, and agents of any type arrive at a steady `Poisson` rate. Buyers accept prices that are lower than their value, and sellers accept wages that are higher than their costs, and the prices and wages set for different types determine the rates of arrival. We further assume buyer and seller types are located in a metric space. Typically, the buyer would be happy as long as the seller a reasonable match with her requirements, so that there is an upper bound on the metric distance of any feasible match. We will use "type" and "node" interchangeably.

A policy for the platform is to (i) set a per-type price (for buyers) and wage (for sellers); and (ii) schedule feasible matches between the arriving buyers and sellers who accept the price/wage. The goal of this process can be to maximize social surplus, or revenue, or even more complex combinations such as maximizing surplus subject to a minimum revenue guarantee. Nevertheless, for all such objectives, finding the optimal policy is not difficult – in particular, it is given by a convex flow program with prices/wages set in a fashion to balance the rates of flow from buyers to sellers and maximize social surplus. Moreover, note that such a flow program formally encodes the idea of *market segmentation*.

Impatient agents and the role of thickness: Now consider what happens when buyers and sellers have a *patience level*. They observe the price/wage, and accept/reject it. If they accept, the platform may not immediately be able to match them because a feasible match may not exist at that very moment. This will cause the buyers and sellers to wait. However, if they are not matched by the time their patience expires, they will abandon the system. The pricing and scheduling policy of the platform needs to make sure the abandonment probability is very small, since this causes dissatisfaction with the platform. Patience is a real constraint in two-sided platforms; for instance, Uber has recently introduced pricing for buyers with different patience levels [1].

Finding the optimal policy for scheduling matches under an abandonment constraint is a difficult stochastic control problem (see Section 4 for details); however, a commonly-used class of policies which admits tractable analysis is one where the platform "pools" arriving demand and supply at close-by nodes, which correspond to clearinghouses for matching – we alternately refer to these as "facilities", for reasons which will become clear soon. Demand and supply is probabilistically routed to facilities, and agents routed to a single facility are matched optimally. For such policies to have abandonment probability below a threshold, we show in Section 4 that two natural conditions are sufficient: (i) the rate of arrival of supply and demand to a facility are equal; and (ii) this rate is at least a certain threshold, so that agents find feasible matches with high probability, *i.e.*, the facility is sufficiently *thick*. In essence, the aim of the platform is to ensure that the time for finding a match is on average much smaller than the patience of any buyer or seller.

The above setting thus naturally encodes the segmentation-thickness tradeoff: on one hand, the convex flow problem tries to segment agent types into smaller flows in order to maximize the platforms' objectives; on the other hand, the need to ensure sufficient thickness at the clearinghouses places limits on the optimization. Crucially, the addition of the thickness constraints destroys the underlying convexity of the problem, and makes it much more computationally challenging – this is brought out by relating this problem to the facility-location problem, as we do next.

Before proceeding, we make a few remarks on the above model. We assume the set of different types and their embedding into a metric space is constructed by the platform based on features of the agents in a way that metric distance captures the quality of a match; however, these are not necessarily known to the agents. This provides justification for the assumption that the value (or cost) of each agent is decoupled from the metric distance. For the same reason, agents cannot choose to arrive as a different type. Moreover, note that the clearinghouses are not physical locations, but rather canonical types. An agent can be assigned this canonical type if (s)he has a type that is within a distance threshold and our policy will be constrained to match agents such that the corresponding canonical types are the same. For instance, these could be canonical room types on a renting portal, or canonical skill sets on a freelancing platform. Hence we assume there are no facility costs.

1.1 From Two-Sided Markets to Facility Location

One of our contributions (Section 4) is the reduction of a natural class of stochastic control policies for the *dynamic* pricing and scheduling problem described above to a *static* problem of locating facilities in the metric space so that these facilities are sufficiently thick. In the resulting facility location problem, buyers and sellers flow into nodes located in a metric space. The price (resp. wage) set at any buyer (resp. seller) node determines the volume of flow at that node. The platform needs to (i) open facilities in the metric space; (ii) assign prices and wages to each node; and (iii) probabilistically route the resulting flow to the open facilities so that the following service guarantees are satisfied:

(1) **Quality of service guarantees**: The flow assigned to a facility is from supply and demand nodes within distance R. (This ensures any matched demand/supply is within distance $2R$.)
(2) **Service availability guarantees**: Each facility needs to have *flow balance*, that is, equal amount of supply and demand is routed there. Further, there is a *lower bound* $\mathbf{L}$ on the flow routed to each facility, capturing the thickness constraint. As mentioned before, in Section 4, we show that these constraints arise from natural stochastic matching policies for dynamically arriving *impatient* agents.

The platform's objective is to maximize the *gains from trade* or *social surplus*, which is total value of buyers *minus* total cost of sellers, subject to weak *budget balance*, meaning the platform has non-negative total profit. We term this optimization problem as *Two-sided Facility Location*.

Such a facility location model enables us to abstract out the stochastic dynamics of scheduling with deadlines, and focus on studying the interplay between pricing and the service guarantee of finding good quality matches with high probability. Note that if we ignore the flow lower bound constraint, there is indeed a convex program for maximizing surplus. However, this may set prices and wages in such a fashion that there may not be enough nearby supply (resp. demand) to ensure thickness and prevent abandonment. On the other hand, if we do take the lower bounds into account, this may cause the prices (resp wages) to be set so low (resp. high) that the platform loses money. We illustrate this trade-off in an example in Appendix A.1, which shows that with a flow lower bound constraint, the platform can obtain much larger surplus by losing money on one facility, and making up for this loss at other facilities.

In summary, there is a three-way tension between the goals of maximizing surplus, platform profitability, and the service guarantees described above. Our problem formulation of *Two-sided Facility Location* captures this trade-off, and understanding the computational complexity of this problem is the main focus of this paper. Our model is sufficiently flexible to accommodate more complex extensions, and we present an extension to pricing patience in Appendix C.

1.2 Our Results

Our main contribution in Sections 2–3 is to show an approximation algorithm for *Two Sided Facility Location*. We present a new LP rounding framework that for any constant $\epsilon > 0$, achieves a $(1 + \epsilon)$ approximation to the social surplus objective. It relaxes the distance bound constraint by a factor of 4, while preserving the budget balance constraint, as well as the flow balance and lower bound constraints at each facility. If we allow a tiny additive error Δ in the surplus objective, our algorithm requires solving $O\left(\frac{n^{1/\epsilon}}{\epsilon} \log \frac{n W_{\max}}{\epsilon \Delta}\right)$ LPs, where n is the number of nodes, and $W_{\max}$ is the maximum possible surplus.

We show in Section 3.1 that the surplus objective is NP-HARD to approximate to a factor $o(\mathbf{L}^{\xi})$ for some constant $\xi > 0$, unless the distance bound is relaxed by at least a factor of 2.

Techniques. Our facility location variants mirror the *profit earning facility location* problem in [29]. Just like that setting, we have lower bounds on demand served at each facility and an upper bound on how far the facility can be from an assigned demand. However, there are key differences that preclude the application of existing techniques from lower-balanced facility location [20, 21, 25, 29, 32]: First, the demand or supply at each node is a variable that can be adjusted using pricing. This means the demand/supply can be zero at some "outlier" nodes, so that they do not need to be served by any facility. Secondly, each facility needs to satisfy flow balance between supply and demand, and finally, both surplus and profit involve *differences*, so the platform can potentially lose money at some facilities, but recover it at others.

The above differences make formulating an LP relaxation tricky. Note that even in [29], the version with outliers and profits that can become negative has unbounded integrality gap, because the optimal profit can be zero while the LP achieves positive profit. Unlike [29], since we can control demand/supply by pricing, we have greater flexibility in modifying the LP variables. Despite this, the integrality gap of the straightforward LP formulation for our problem is large, because there could be a facility that generates a bulk of the surplus, but has large negative profit that is compensated by other facilities. (See Appendix A.2 for an example).

This brings up our main technical contributions: We first observe that if we focus on the LP variables corresponding to a facility, we can scale these up or down by changing the fraction to which this node is an outlier. This enables us to use techniques reminiscent of improved greedy algorithms for budgeted coverage problems [18, 26]: In particular, we *strengthen the LP formulation* via guessing a few of the facilities that are opened in the optimal solution. Next, we use the guesses to develop a structural characterization for this stronger LP based on modifying variables for *pairs* of facilities. In effect this shows that there is some integrality in the neighborhood of any partially open facility, which helps us consolidate these facilities while preserving all constraints.

1.3 Related Work

Two-sided Markets. Our objective maximizes social surplus subject to budget balance (and individual rationality). This is a classic objective in two-sided market mechanisms, and originates in the celebrated work of Myerson and Satterthwaite [30], where it is termed *gains of trade*. They considered the case of a single buyer and seller. This has inspired a recent line of work on truthful mechanisms for approximate surplus maximization in markets of multiple buyers and sellers [10, 12, 28], ultimately resulting in a 2-approximation to gains of trade. This line of work assumes buyers and sellers are matched in one shot. The novelty in our work is in modeling a *dynamic setting* and incorporating *service availability* guarantees while preserving the same objectives. We therefore consider the more natural class of mechanisms that post prices and wages. Posted price mechanisms have been extensively studied in two-sided marketplaces [6, 28, 31, 34], and the main idea we borrow from this literature is the notion of *insulating tariffs* [34], which posits that market design is easier if the prices seen by buyers is disconnected from the wages seen by service providers.

Another recent line of work shows approximately optimal mechanisms for maximizing welfare in two sided markets with *goods* [16, 17]; however, theirs is a sum objective defined in terms of the final sets of items allocated to each buyer and seller, which is different from the gains of trade.

Dynamic Marketplaces. Our work on dynamic marketplaces is related to several recent works on online scheduling under stochastic arrivals of tasks on machines with limited resources [14, 15, 23]. Tasks have (private) types comprising their value, arrival time, and deadline; the platform's goal is to maximize welfare while truthfully eliciting the type. While similar to our work on pricing resources or tasks, they allow agents to choose assignments based on posted prices (*envy-freeness*). Another difference is the markets considered in their studies are one-sided.

Dynamic two-sided markets also serve as motivation for recent work on "online matching with delays" [7, 19]. Here, buyers and sellers arrive online in a metric space, and can be matched at any time subsequently. The goal is to minimize the total distance cost plus waiting cost, and the authors present a log-competitive algorithm. These models do not incorporate pricing. Further, our dynamic marketplace models are more closely related to dynamic matchings with *stochastic* arrivals, and we review this literature in Section 4.

2 PROBLEM STATEMENT

There is a metric space $G(V, E)$ with an associated distance function c. The two sided facility location problem is parametrized as TWO-SIDED FAC-LOC($\mathbf{L}, R$). Here, R is the distance threshold for assignment to a facility and $\mathbf{L}$ is the lower bound on flow routed to each open facility.

Each node $j \in V$ is associated with a demand function F_j and a supply function H_j. When offered price p, we assume the demand (*i.e.*, buyers) at node j is $d_j F_j(p)$, where $F_j(p)$ is a non-increasing function of p corresponding to the survival function of a continuous density function f_j on valuations; formally $F_j(p) = \int_{v=p}^{\infty} f_j(v) dv$. In other words, the volume of buyers is d_j, and when quoted a price p, only buyers with valuations at least p choose to participate. We assume there is a finite price $p_{\max}$ so that $F_j(p_{\max}) = 0$ for all $j \in V$.

Similarly, when offered wage w, the supply of sellers at node j is $s_j H_j(w)$, where $H_j(w)$ is a non-decreasing function of w, corresponding to the CDF of a continuous density function h_j on costs; formally $H_j(w) = \int_{c=0}^{w} h_j(c) dc$. When offered wage w, all sellers with cost at most w participate, resulting in supply $s_j H_j(w)$. We assume that $H_j(0) = 0$, *i.e.*, sellers accrue 0 utility by not participating in the platform.

Let $\mathcal{F} \subseteq V$ the set of all candidate facilities; we set $\mathcal{F} = V$. For each node j, $B_R(j) \subseteq \mathcal{F}$ denotes the set of all *compatible* facilities, i.e., $i \in \mathcal{F}$ such that $c(i, j) \leq R$. Similarly, for each facility i, we define $B_R(i)$ as the set of all *compatible* nodes, i.e., $j \in V$ such that $c(i, j) \leq R$. A solution to TWO-SIDED FAC-LOC($\mathbf{L}, R$) is specified by the following:

- An assignment of price p_j and wage w_j to each node $j \in V$. If the price (resp. wage) at node j is $p_{\max}$ (resp. 0), we assume this node generates no demand (resp. supply).
- A set of locations $S \subseteq \mathcal{F}$ for opening the facilities; and
- A routing scheme $\vec{x}_j^d$ (resp. $\vec{x}_j^s$) for each demand (resp. supply) node $j \in V$ that generates non-zero demand (resp. supply). For facility $i \in S$, if $x_{ij}^d > 0$ then $i \in B_R(j)$ (i.e., i is a compatible node for facility j). Further, $\sum_{i \in B_R(j)} x_{ij}^d = 1$ for all nodes $j \in V$ that generate non-zero demand; similarly, $\sum_{i \in B_R(j)} x_{ij}^s = 1$ for each $j \in V$ with non-zero supply.

Note that the flow of demand (resp. supply) from node $j \in V$ to facility $i \in S$ is $d_j F_j(p_j) x_{ij}^d$ (resp. $s_j H_j(w_j) x_{ij}^s$). As motivated in Section 1.1, the flows need to satisfy the *flow balance* and *flow lower bound* conditions at each facility $i \in S$:

Flow Balance. The total amount of supply and demand are equal,

Flow Lower Bound. The total amount of supply (resp. demand) routed there is at least $\mathbf{L}$.

Both these desiderata can be formalized via the following constraint:

$$\sum_j d_j F_j(p_j) x_{ij}^d = \sum_j s_j H_j(w_j) x_{ij}^s \geq \mathbf{L}, \text{ for all } i \in S.$$

Surplus (Gains from Trade) Objective. We first define the following quantities:

$$\mathcal{V}_j(p) = d_j \int_{v=p}^{\infty} v f_j(v) dv \qquad \text{and} \qquad C_j(w) = s_j \int_{c=0}^{w} c h_j(c) dc$$

respectively denote the total value of buyers generated by node j when the price there is p and the cost of sellers at node j when the wage there is w. The surplus objective can then be written as:

$$\text{Social Surplus} = \sum_{j \in V} \left(\mathcal{V}_j(p_j) - C_j(w_j) \right)$$

The goal of the platform is to maximize social surplus subject to the platform profit being non-negative. This is termed weak budget balance, and is written as:

$$\text{Profit} = \sum_{j \in V} \left(d_j p_j F_j(p_j) - s_j w_j H_j(w_j) \right) \geq 0$$

Note that simply maximizing surplus may not guarantee non-negative profit because of the flow lower bound constraint. We illustrate this in an example in Appendix A.1.

In this section, we make the standard regularity assumptions (*à la* Myerson-Satterthwaite [30]) on the density functions f_j and h_j. In particular, we assume $xF_j^{-1}(x)$ is concave in x and $yH_j^{-1}(y)$ is convex in y. This is true for instance, for all log-concave densities f_j and h_j, which includes Normal, Exponential, and Uniform distributions.

3 APPROXIMATION ALGORITHM

We characterize the approximation ratio of any algorithm for Two-Sided Fac-Loc(L, R) as (α, γ), if the resulting solution relaxes the distance bound of an assignment to a facility to αR, ensures lower bound L, and has surplus OPT/γ, where OPT is the optimal surplus. First, we show that it is NP-Hard to obtain $\gamma = o(L^\xi)$ for some constant $\xi > 0$, unless $\alpha \geq 2$. Subsequently, we present a $(4, 1 + \epsilon)$ approximation. For the algorithm to have polynomial running time, they also need lose a small additive amount in the objective; as we show later, this quantity can be exponentially small.

3.1 Hardness of Approximation

THEOREM 3.1. *It is* NP-Hard *to find a* (α, γ) *approximation for* Two-Sided Fac-Loc(L, R) *unless* $\alpha \geq 2$ *or* $\gamma \geq L^\xi$ *for some constant* $\xi > 0$.

PROOF. We reduce from Maximum Independent Set in k-regular graphs (k-MIS). Given a k-MIS instance with n vertices and $m = kn/2$ edges, construct a metric space where each edge in the k-regular graph $G(V, E)$ has length $2R$. Place a demand node at the mid-point of each edge, and a supply node at each vertex. We set L = k. Each supply node has $s_j = k$, and supply function $H^{-1}(r) = 1 - \delta$ for $r \in [0, 1]$. Similarly, each demand node has $d_j = 1$, and demand function $F^{-1}(q) = 1$ for $q \in [0, 1]$. Since the distance threshold is R, the facilities are opened at vertices of the graph. Each such facility must see k units of supply and demand, which means all neighboring demand is routed there, leading to welfare (resp. profit) $k\delta$ at that facility. Since two open facilities cannot share a demand, this means the open facilities form an independent set. Therefore, the surplus of Two-Sided Fac-Loc(k, R) is δ times the size of the maximum independent set in G. This is NP-hard to approximate to within a factor of k^ξ for some constant $\xi > 0$; see [5, 24]. Therefore, we need to relax the distance bound by at least a factor of 2. □

3.2 Linear Programming Relaxation

We now formulate Two-Sided Fac-Loc(L, R) as an integer linear program. For ease of exposition, we compare against an optimal solution that is restricted to using prices from a fixed set $\mathcal{P}$ and wages from a fixed set $\mathcal{W}$. Our solution is not restricted to using prices and wages from this set. In Appendix B, we show that our LP admits to a polynomial time solution of arbitrary additive accuracy when this assumption is relaxed, and demand/supply distributions are continuous.

Note that we assume $p_{\max} \in \mathcal{P}$ and $0 \in \mathcal{W}$, and at this price (resp. wage) the demand (resp. supply) is identically zero. This is the price (resp. wage) where this node becomes an outlier and the solution is not required to open a facility nearby.

Instead of writing our LP using prices and wages, we use the associated demand/supply values. Let $Q_j = \{q \mid q = F_j(p), p \in \mathcal{P}\}$ and $\mathcal{R}_j = \{r \mid r = H_j(w), w \in \mathcal{W}\}$. The case where the node is an

outlier now corresponds to setting $q = 0$ (resp. $r = 0$). Moreover, the valuations/costs can also be redefined using supply/demand values as follows:

$$\mathcal{V}_j(q) = d_j \int_{v=F_j^{-1}(q)}^{\infty} v f_j(v) dv \qquad \text{and} \qquad C_j(r) = s_j \int_{c=0}^{H_j^{-1}(r)} c h_j(c) dc \tag{1}$$

respectively denote the total value of buyers generated by node j when the price there is $F_j^{-1}(q)$, and the cost of sellers at node j when the wage there is $H_j^{-1}(r)$.

Variables. For each candidate facility $i \in \mathcal{F}$, let $y_i \in \{0, 1\}$ be the indicator variable that a facility is opened at that location in the metric space. Let $\alpha_{jq} = 1$ if the price at node $j \in V$ corresponds to $q \in Q_j$. Similarly define β_{jr} for $r \in \mathcal{R}_j$. The variable z_{ijq} is non-zero only if $\alpha_{jq} = 1$ and $i \in B_R(j)$. In this case, it is the fraction of j's demand that is routed to i. We define z_{ijr} similarly for supply. Note that the actual flow from j to i is $d_j q z_{ijq}$; similarly for sellers.

Objective and Weak Budget Balance. The objective of social surplus and the profit being non-negative can be captured by:

• Surplus Objective:

$$\max \sum_{j \in V} \left(\sum_{q \in Q_j} \alpha_{jq} \mathcal{V}_j(q) - \sum_{r \in \mathcal{R}_j} \beta_{jr} C_j(r) \right) \tag{2}$$

• Weak Budget Balance:

$$\sum_{j \in V} \left(\sum_{q \in Q_j} \alpha_{jq} d_j q F_j^{-1}(q) - \sum_{r \in \mathcal{R}_j} \beta_{jr} s_j r H_j^{-1}(r) \right) \geq 0 \tag{3}$$

Feasibility. The following constraints connect the variables together. We present these constraints only for buyers (that is, $q \in Q_j$); the constraints for sellers is obtained by replacing q with $r \in \mathcal{R}_j$. First, for each $q \in Q_j$, we need to choose one price for buyers (resp. sellers).

$$\sum_{q \in Q_j} \alpha_{jq} = 1 \quad \forall j \in V \tag{4}$$

$$\sum_{i \in B_R(j)} z_{ijq} = \alpha_{jq} \quad \forall j \in V, q \in Q_j \tag{5}$$

Next, if demand is fractionally routed from j to i, then i should be open and within distance R. Note that we need to ignore the case where $q = 0$ (resp. $r = 0$) since in this case, the demand (resp. supply) routed is zero, so that there is no need for a nearby facility.

$$\sum_{q \in Q_j, q>0} z_{ijq} \leq y_i \quad \forall j \in V, i \in B_R(j) \tag{6}$$

Service Availability. We finally encode *flow balance* and *flow lower bound* at each facility:

$$\sum_{j \in B_R(i)} d_j \sum_{q \in Q_j} q z_{ijq} = \sum_{j \in B_R(i)} s_j \sum_{r \in \mathcal{R}_j} r z_{ijr} \quad \forall i \in \mathcal{F} \tag{7}$$

$$\sum_{j \in B_R(i)} d_j \sum_{q \in Q_j} q z_{ijq} \geq \mathbf{L} y_i \qquad \forall i \in \mathcal{F} \tag{8}$$

If we replace the integrality constraints on $\{y_i\}$ and the $\{\alpha_{jq}, \beta_{jr}\}$ with $y_i, \alpha_{jq}, \beta_{jr} \in [0, 1]$, the above is a linear programming relaxation of the problem.

3.3 Integrality Gap and Stronger LP Formulation

The main technical hurdle arises because of the flow lower bound constraint: The LP optimum (and even an integer optimum) can now open facilities i which have positive surplus but negative profit, and compensate for the loss in profit by other facilities with positive profit. (See Appendix A.1 for an example.) Note that Constraints (6) and (4) together imply:

$$\sum_{i \in B_R(j)} y_i \geq 1 - \alpha_{j0} \quad \forall \text{ Demand nodes } j$$

and similarly for supply nodes. We call the quantities α_{j0} (resp. β_{j0}) the *outlier fraction* of node j, and correspond to the case where the node is priced in such a way that it does not generate flow. In this case, there is no need to open a facility to satisfy j. Therefore, if $\alpha_{j0}\beta_{j0} > 0$, then the above constraints could imply $\sum_{i \in B_R(j)} y_i < 1$. This means there could only be a small fractional facility open in the vicinity of j, which can account for a lot of the surplus. This makes the LP have super-polynomial integrality gap and we present an example in Appendix A.2.

Our first technical contribution involves adding constraints to the above LP formulation to bound its integrality gap. Before showing how to strengthen the LP, we present the following easy claim, which implies that once we round $\{y_i\}$, the remaining solution can easily be made integral.

LEMMA 3.2. *Given any feasible LP solution, there is an equivalent solution that assigns only one price (resp. wage) per demand (resp. supply) node, that preserves all constraints and does not decrease the objective.*

PROOF. The rounding of α_{jq}, β_{jr} is simple. Let $\hat{q}_j = \sum_{q \in Q_j} q\alpha_{jq}$ and $\hat{r}_j = \sum_{r \in \mathcal{R}_j} r\beta_{jr}$. Set the price of location j to be $F_j^{-1}(\hat{q}_j)$ and the wage at j to be $H_j^{-1}(\hat{r}_j)$. In other words, set $\hat{\alpha}_{j\hat{q}_j} \leftarrow 1$ and $\hat{\beta}_{j\hat{r}_j} \leftarrow 1$. Further set $\hat{z}_{ij\hat{q}_j} \leftarrow \sum_{q \in Q'_j} z_{ijq} \frac{q}{\hat{q}_j}$ and $\hat{z}_{ij\hat{r}_j} \leftarrow \sum_{r \in \mathcal{R}'_j} z_{ijr} \frac{r}{\hat{r}_j}$.

Note that this process preserves the demand and supply from node j to facility i, which preserves all the constraints in the LP formulation. Note further that the function $qF_j^{-1}(q)$ is concave in q by the regularity of the demand function. Therefore,

$$\sum_q \alpha_{jq} qF_j^{-1}(q) \leq \left(\sum_q \alpha_{jq}q\right) F_j^{-1}\left(\sum_q \alpha_{jq}q\right) = \hat{q}_j F_j^{-1}(\hat{q}_j)$$

Similarly, since we assumed $rH_j^{-1}(r)$ is convex in r (by regularity of supply), we have $\sum_r \beta_{jr} rH_j^{-1}(r) \geq \hat{r}_j H_j^{-1}(\hat{r}_j)$. Therefore, this transformation preserves weak budget balance. Next, we note that $\mathcal{V}_j(q)$ is always a concave function of q and $C_j(r)$ is always a convex function. Therefore, the above argument also implies the welfare (social surplus) does not decrease in the above transformation. □

Define a variable for the surplus W_i and profit R_i of facility i respectively as:

$$W_i = \sum_{j \in B_R(i)} \left(\sum_{q \in Q'_j} \mathcal{V}_j(q)z_{ijq} - \sum_{r \in \mathcal{R}'_j} C_j(r)z_{ijr}\right) \tag{9}$$

$$R_i = \sum_{j \in B_R(i)} \left(\sum_{q \in Q'_j} d_j qF_j^{-1}(q)z_{ijq} - \sum_{r \in \mathcal{R}'_j} s_j rH_j^{-1}(r)z_{ijr}\right) \tag{10}$$

Then the objective can be rewritten as: Maximize $\sum_i W_i$, and weak budget balance is $\sum_i R_i \geq 0$. Further note that $W_i \geq R_i$ since for any q, r, we have $\mathcal{V}_j(q) \geq d_j qF_j^{-1}(q)$, and $C_j(r) \leq s_j rH_j^{-1}(r)$ if we integrate the expressions in Equation (1) by parts.

Stronger LP Formulation. Let $\epsilon > 0$ be any constant, and let $\theta = \frac{1}{\epsilon}$. We guess the θ facilities in the optimum solution that have the most surplus. There are two cases. First, if the optimum solution opens fewer than θ facilities, we can perform a brute force search over all integer solutions that open at most θ facilities. This can be done in $O(n^\theta)$ time, where $n = |V|$. For each selection of facilities, Lemma 3.2 implies that solving the LP formulation with the corresponding y_i set to 1 and the rest to zero yields the optimal surplus (or results in declaring infeasibility). We can therefore find the surplus maximizing solution among these in polynomial time; call this surplus W_1.

In the other case, the optimum solution opens more than θ facilities. In this case, for every choice of parameter $\mathcal{W} \geq 0$ scaled to powers of $(1 + \epsilon)$, and every subset $S \subseteq \mathcal{F}$ with $|S| = \theta$, define $LP(\mathcal{W}, S)$ as having all of Constraints (3) – (8), *plus* the following new ones:

$$W_i \leq \mathcal{W} y_i \qquad \forall i \in \mathcal{F} \setminus S \tag{11}$$

$$y_i = 1 \qquad \forall i \in S \tag{12}$$

$$\sum_{i \in S} W_i \geq \mathcal{W}\theta(1 - \epsilon) \tag{13}$$

Let OPT denote the optimum surplus, and also define $W_2 = \max\{LP(\mathcal{W}, S) \mid \mathcal{W} \geq 0, S \text{ s.t. } |S| = \theta\}$. Then, it is easy to see that $OPT \leq \max(W_1, W_2)$: If OPT opens fewer than θ facilities, then clearly $W_1 \geq OPT$, since W_1 opens all possible choices of at most θ facilities. Otherwise, let W_i^* denote the surplus generated by open facility i in OPT. Let W^* denote the θth largest value of W_i^*. Choose $\mathcal{W} \in [W^*, W^*(1 + \epsilon)]$, and S as the set of θ facilities in OPT with $W_i^* \geq W^*$. This induces a feasible solution to the above constraints, so that the LP optimum is at least OPT.

3.4 Structural Characterization of LP Optimum

Our second technical contribution is a new structural characterization about the LP optimum. This is crucial for the rounding that we present subsequently, since it allows sufficient mass of facility to be located in roughly the same neighborhood.

Recall α_{j0}, β_{j0} are the fractions to which node j is an outlier, *i.e.* has zero flow. These variables are the reason the simpler LP had large integrality gap, since they allow facilities in $B_R(j)$ to be open to small fractions. Our main observation is the following:

LEMMA 3.3 (STRUCTURAL CHARACTERIZATION). *There is a* $(1 + \epsilon)$ *approximation to the objective of* $LP(\mathcal{W}, S)$ *that satisfies:*

$$\forall i \in \mathcal{F}, \qquad y_i \in (0, 1) \qquad \Rightarrow \qquad \exists j \in B_R(i) \text{ s.t. } \alpha_{j0}\beta_{j0} = 0$$

High level Idea. Before presenting the proof, we present the high level idea. Consider a facility that violates the statement. If it has $R_i > 0$, then consider all LP variables $\{y_i, z_{ijq}, z_{ijr}\}$ corresponding to some such facility i and uniformly increase them. This increases both profit and surplus. We can decrease the fractions $\{\alpha_{j0}, \beta_{j0}\}$ to which any node j connected to i is assigned as outlier to compensate the fraction to which it is assigned i. Note that Constraints (7) and (8) are local to a single facility. Since we scale up *all* variables corresponding to a facility, we preserve these constraints. If we keep up this process, then either the facility is completely open ($y_i = 1$); or some demand/supply node assigned to it has $\alpha_{j0} = 0$ or $\beta_{j0} = 0$. (This must hold in the LP optimum.)

On the other hand, if $W_i > 0$ but $R_i < 0$, then increasing its LP variables would hurt profit, which may violate the budget balance constraint; while reducing the variables would increase profit but hurt the surplus. The idea now is the following: Take any pair of such facilities; increase the variables for one facility while decrease them for the other. There is always a way of doing this so that *both* the total profit and surplus do not decrease – this is essentially a fractional knapsack

argument. Again, since we uniformly scale all variables corresponding to a facility, we preserve all constraints. Note that the process can also stop when a facility closes ($y_i = 0$). Eventually, we run out of pairs, so that for all but one facility, the above characterization holds.

At this point, the strengthened LP kicks in. The singleton facility violating the above lemma was fractionally open and had $R_i < 0$. It has surplus at most $\mathcal{W}$ by Constraint (11). But we have integrally open facilities that generate surplus at least $\mathcal{W}\frac{(1-\epsilon)}{\epsilon}$ by Constraint (13), which means closing the singleton facility reduces surplus by at most $(1 - \epsilon)$, and preserves budget balance.

Proof of Lemma 3.3. We first simplify the LP. Let $Q'_j = Q_j \setminus \{0\}$ and $\mathcal{R}'_j = \mathcal{R}_j \setminus \{0\}$. Let $\eta_j = \sum_{q \in Q'_j} \alpha_{jq}$ and $\phi_j = \sum_{r \in \mathcal{R}'_j} \beta_{jr}$ respectively denote the fractions to which j is assigned prices (wages) that correspond to non-zero demand (supply). We can rewrite the constraints (4) and (5) as:

$$\eta_j = \sum_{q \in Q'_j} \sum_{i \in B_R(j)} z_{ijq} \leq 1 \qquad \text{and} \qquad \phi_j = \sum_{r \in \mathcal{R}'_j} \sum_{i \in B_R(j)} z_{ijr} \leq 1 \qquad \forall j \in V \tag{14}$$

and set $\alpha_{j0} = 1 - \eta_j$, and $\beta_{j0} = 1 - \phi_j$. Recall from Equations (9) and (10) that W_i and R_i are respectively the surplus and profit of facility i in the LP optimum.

We call node j *fully demand-utilized* if $\eta_j = 1$, and *fully supply-utilized* if $\phi_j = 1$. We say that node j is *partially demand-connected* to facility $i \in \mathcal{F}$ if $\sum_{q \in Q'_j} z_{ijq} > 0$, and *partially supply-connected* if $\sum_{r \in \mathcal{R}'_j} z_{ijr} > 0$. Let $J_D(i)$ denote the set of nodes that are partially demand-connected to $i \in \mathcal{F}$, and $J_S(i)$ be the set that is partially supply-connected.

LEMMA 3.4. *In the LP optimum, for any $i \in \mathcal{F}$ with $y_i > 0$, we have $W_i > 0$. Furthermore, all except one facility with $y_i > 0$ satisfy the following condition: either $y_i = 1$; or there exists $j \in J_D(i)$, such that j is fully demand-utilized; or there exists $j \in J_S(i)$ such that j is fully supply-utilized.*

PROOF. First, note that $W_i \geq R_i$. Suppose an open facility has $W_i \leq 0$. This implies $R_i \leq 0$. Consider a different solution that sets $y_i = 0$ and $z_{ijq} = z_{ijr} = 0$ for all $j \in V, q \in Q'_j, r \in \mathcal{R}'_j$. We adjust η_j and ϕ_j for each $j \in V$ to preserve constraint (14). This new solution has $W_i = R_i = 0$ and has at least as large surplus and profit. Since we set all LP variables corresponding to i to zero, this satisfies constraints (6), (7), and (8), and is therefore feasible for the LP.

We therefore only focus on facilities whose $W_i > 0$. Consider the set of these facilities and split them into two groups. Let

$$S_1 = \{i \in \mathcal{F} \mid y_i \in (0, 1) \text{ and } R_i < 0\} \qquad \text{and} \qquad S_2 = \{i \in \mathcal{F} \mid y_i \in (0, 1) \text{ and } R_i \geq 0\}$$

Assume that for all of these facilities, there is no $j \in J_D(i)$, such that j is *fully demand-utilized* and no $j \in J_S(i)$ such that j is *fully supply-utilized*

First consider the facilities in set S_2, we can increase the LP variables till the condition of the lemma is satisfied; this process only increases both profit and surplus, preserving all constraints. We do this as follows: Suppose no $j \in J_D(i)$ is fully demand-utilized and no $j \in J_S(i)$ is fully supply-utilized. In this case, let

$$\theta = \min\left(\frac{1}{y_i}, \min_{j \in J_D(i)}\left(\frac{1 - \sum_{q \in Q'_j} \sum_{i' \neq i} z_{i'jq}}{\sum_{q \in Q'_j} z_{ijq}}\right), \min_{j \in J_S(i)}\left(\frac{1 - \sum_{r \in \mathcal{R}'_j} \sum_{i' \neq i} z_{i'jr}}{\sum_{r \in \mathcal{R}'_j} z_{ijr}}\right)\right)$$

Since $\eta_j < 1$ for all $j \in J_D(i)$ and $\phi_j < 1$ for all $j \in J_S(i)$, we have $\theta > 1$. Suppose we increase y_i, z_{ijq} for all $j \in J_D(i), q \in Q'_j$, and z_{ijr} for all $j \in J_S(i), r \in \mathcal{R}'_j$ by a factor of θ. We will still have $\eta_j \leq 1$ for all $j \in J_D(i)$ and $\phi_j \leq 1$ for all $j \in J_S(i)$. However, either y_i or one of these values will become exactly 1. Note that since we scaled all LP variables corresponding to i by the same factor,

this preserves constraints (6), (7), and (8). The surplus and profit of this facility increase by a factor $\theta > 1$, which contradicts the optimality of the LP solution. Therefore, the facilities in S_2 all have a neighboring j that is either fully demand-utilized or fully supply-utilized.

Next consider the facilities in set S_1. Suppose the condition in the lemma is not satisfied, so that there are two facilities i and i' with $y_i, y_{i'} \in (0, 1)$, and with no neighboring j that is either fully demand-utilized or fully supply-utilized. Suppose $W_i/|R_i| = a$ and $W_{i'}/|R_{i'}| = b$ with $a \geq b$. We multiply each LP variable corresponding to i by a factor of $(1 + \delta)$, and multiply each LP variable corresponding to i' by a factor of $\left(1 - \frac{W_i}{W_{i'}}\delta\right)$. Using the same argument as above, this process preserves the constraints that are specific to a facility, since all variables are changed by the same factor. The increase in surplus of facility i is δW_i, and the decrease in surplus of facility i' is δW_i, so the overall surplus is preserved. The decrease in profit of facility i is $|R_i|\delta$, and the increase in profit of facility i' is $|R_{i'}|\frac{W_i}{W_{i'}}\delta \geq |R_i|\delta$ by our assumption that $a \geq b$. Therefore, this process cannot decrease profit, hence all constraints are preserved. We choose δ as the smallest value that either makes facility i have $y_i = 1$ or one neighboring j either fully supply or demand utilized, or that sets the variables of facility i' to zero. In all cases, the size of set S_1 reduces by one. We repeat this process till there is only one facility in S_1, completing the proof.   □

The following corollary now restates Lemma 3.3, completing its proof.

COROLLARY 3.5. *There is a $(1 + \epsilon)$-approximation to the LP optimum where any facility with $y_i > 0$ satisfies the following condition: either $y_i = 1$; or there exists $j \in J_D(i)$, such that j is* fully demand-utilized*; or there exists $j \in J_S(i)$ such that j is* fully supply-utilized.

PROOF. By Constraint (12), there is a set of facilities S that are fully open (*i.e.*, $y_i = 1$) and $\sum_{i \in S} W_i \geq \mathcal{W}\frac{1-\epsilon}{\epsilon}$ by Constraint (13). The rounding in Lemma 3.4 does not touch these facilities, since we only increase/decrease variables corresponding to partially open facilities (*i.e.*, those with $y_i \in (0, 1)$). Lemma 3.4 implies there is only facility i that violates the condition of the corollary. This facility must have surplus $W_i \leq \mathcal{W}y_i \leq \mathcal{W}$ by Constraint (11). This means closing this facility (setting all its associated variables to zero) reduces the LP optimum by at most a factor of $(1 - \epsilon)$. Since the previous lemma implies this facility had $R_i < 0$, this means closing it only increases profit, preserving weak budget balance.   □

3.5   Rounding the LP Relaxation

The rounding now follows approaches similar to those in [20, 35]. We first present the high-level idea. Note that if a node j has $\alpha_{j0} = 0$ or $\beta_{j0} = 0$, then Constraint (6) implies $\sum_{i \in B_R(j)} y_i \geq 1$. Consider an *independent set* of such nodes, such that no two are fractionally assigned to the same facility. For any j in this set, move all partially open facilities in $B_R(j)$ to j itself, so that there is a facility integrally opened at j. Since we move an entire facility, we preserve all flows, so that flow balance and lower bound are preserved, and so is profit. Now a demand/supply can be assigned a distance $2R$ away, and the opened facilities are integral.

At this point, consider any fractionally open facility i. It must have a node j adjacent to it that satisfies the condition in Lemma 3.3. If j has a facility completely open at its location, then move i to location j. Otherwise, j was not part of the independent set in the previous step, which means j and j' shared a fractionally open facility, and the previous step opened a facility completely at j'. In this case, we move i to j', again preserving all flows. This means any demand/supply moves distance at most $4R$, preserving all the LP constraints.

3.5.1   Rounding Facilities. We now present the rounding algorithm in detail. Initially, all facilities $i \in \mathcal{F}$ with $y_i > 0$ are *partially open*. Node $j \in V$ is *untouched* if for all i such that $j \in J_D(i) \cup J_S(i)$,

the facility i is partially open. Let U be the set of *untouched* nodes, and let Z be the set that is either fully demand-utilized or fully supply-utilized. Let $U_f = U \cap Z$.

Phase 1. Consider any $j \in U_f$. W.l.o.g., assume $\eta_j = 1$; the case where $\phi_j = 1$ is symmetric. Let $N(j) = \{i | j \in J_D(i)\}$. For every $i \in N(j)$, we "move" i to location j; call the new facility at location j as i^*. This means we set

- $\bar{y}_{i^*} \leftarrow \sum_{i \in N(j)} y_i$ and $\bar{y}_i \leftarrow 0 \qquad \forall i \in N(j)$;
- $\bar{z}_{i^* j' q} \leftarrow \sum_{i \in N(j)} z_{ij'q}$ and $\bar{z}_{ij'q} \leftarrow 0 \qquad \forall j' \in V, q \in Q_{j'}, i \in N(j)$
- $\bar{z}_{i^* j' r} \leftarrow \sum_{i \in N(j)} z_{ij'r}$ and $\bar{z}_{ij'r} \leftarrow 0 \qquad \forall j' \in V, r \in \mathcal{R}_{j'}, i \in N(j)$

From constraint (6), and the fact that $\eta_j = 1$, we have:

$$\bar{y}_{i^*} = \sum_{i \in N(j)} y_i \geq \sum_{i \in B_R(j)} \sum_{q \in Q'_j} z_{ijq} = \eta_j = 1$$

Subsequently, we mark every $i \in N(j)$ as *closed*, and mark i^* as *completely open*. Furthermore, we mark every j' that was reassigned in the above steps as *touched*.

Note that in the last three steps, any agents at a node j' that was initially assigned to $i \in N(j)$ is now assigned to i^*. Since each of the distances $j' \rightarrow i$ and $i \rightarrow j$ is at most R, the distance from j' to i^* is at most $2R$. Therefore, this step relaxes the distance of a feasible assignment to a facility from R to $2R$.

This process trivially preserves the objective and weak budget balance, as well as constraints (14). Moreover, constraints (6) and (7) are satisfied since we add both sides of the constraints corresponding to $i \in N(j)$ to obtain the constraint for i^*. Finally, to see that (8) is satisfied for i^*, note that

$$\sum_{j' \in J_D(i^*)} \sum_{q \in Q'_{j'}} d_{j'} q \bar{z}_{i^* j' q} = \sum_{i \in N(j)} \sum_{j' \in J_D(i)} \sum_{q \in Q'_{j'}} d_{j'} q z_{ij'q} \geq \mathsf{L} \sum_{i \in N(j)} y_i \geq \mathsf{L}$$

We continue this process, finding a node $j \in U_f$, and merging all facilities in $N(j)$ to one location. At the end of this process, the set U_f is empty.

Phase 2. At the end of Phase 1, each node j which is *touched* (including all *fully utilized* nodes) route some fraction of their demand (or supply) to at least one facility that is completely open. However, there could still be partially open facilities with $y_i \in (0, 1)$ to which demand and supply are assigned. Consider these partially open facilities in arbitrary order. Suppose we are considering facility i and there exists *touched* and *fully utilized* node j such that $\sum_{q \in Q'_j} z_{ijq} > 0$ (resp. $\sum_{r \in \mathcal{R}'_j} z_{ijr} > 0$). Consider the completely open facility i^* such that $\sum_{q \in Q'_j} \bar{z}_{i^* jq} > 0$ or $\sum_{r \in \mathcal{R}'_j} \bar{z}_{i^* jr} > 0$. We move the facility i to location i^*, updating the variables just as in Phase 1; *i.e.*, we set

- $\bar{y}_{i^*} \leftarrow \bar{y}_{i^*} + y_i$ and $\bar{y}_i \leftarrow 0$;
- $\bar{z}_{i^* j' q} \leftarrow \bar{z}_{i^* j' q} + z_{ij'q}$ and $\bar{z}_{ij'q} \leftarrow 0 \qquad \forall j' \in V, q \in Q_{j'}$
- $\bar{z}_{i^* j' r} \leftarrow \bar{z}_{i^* j' r} + z_{ij'r}$ and $\bar{z}_{ij'r} \leftarrow 0 \qquad \forall j' \in V, r \in \mathcal{R}_{j'}$

The argument that all constraints are preserved follows just as before. For any j' that was partially assigned to i, the new assignment is to i^*. This distance is at most

$$c(j', i^*) \leq c(j', i) + c(i, j) + c(j, i^*) \leq R + R + 2R = 4R$$

where we note that the distance $j \rightarrow i^*$ was at most $2R$ because j was potentially reassigned to i^* in Phase 1. We mark all nodes j' that are reassigned in this process as *touched*.

At the end of this process, suppose there are still partially open facilities with $y_i \in (0, 1)$. By Corollary 3.5, each of these facilities i must have some $j \in Z$ partially assigned to it. At the end of Phase 1, we have the invariant that $j \notin U_f$, since U_f is empty. This means j was touched on Phase 1. But in that case, i must have been reassigned in Phase 2, which is a contradiction. Therefore, at

this point, all facilities are either closed ($\bar{y}_i = 0$) or completely open ($\bar{y}_i \geq 1$). Furthermore, for any variable $\bar{z}_{ijq} > 0$ (resp. $\bar{z}_{ijr} > 0$), the facility i is completely open; the distance from j to i is at most $4R$; for each completely open facility, the rate of supply equals the rate of demand (Constraint (7)), and finally, the total flow is at least $\mathbf{L}$ (Constraint (8)).

3.5.2   Final Steps and Running Time. At this point, the facilities are opened integrally. Lemma 3.2 now implies that we can choose one price/wage per node preserving all constraints and the objective. This completes the proof of the following theorem.

THEOREM 3.6. *For any constant $\epsilon > 0$, there is a $(4, 1 + \epsilon)$ approximation algorithm for* TWO-SIDED FAC-LOC($\mathbf{L}, R$).

Note that the surplus can become arbitrarily close to zero. Therefore, for parameter $\Delta > 0$, we will allow additive error Δ in the surplus objective. Note that the maximum possible surplus is $W_{\max} = (\sum_j d_j)p_{\max}$, which is an upper bound on $\mathcal{W}$. If we assume the surplus is at least Δ, then $\max_i W_i^* \geq \Delta/n$. Since the top $1/\epsilon$ facilities on *OPT* have surplus $\mathcal{W}(1 - \epsilon)/\epsilon$, this means we can set $\mathcal{W} \geq \frac{\epsilon\Delta}{2n}$ Therefore, the number of choices of $\mathcal{W}$ is $O\left(\frac{1}{\epsilon}\log\frac{nW_{\max}}{\epsilon\Delta}\right)$. For each choice of $\mathcal{W}$, we need to solve $O(n^{1/\epsilon})$ LPs, so that the overall number of LPs is $O\left(\frac{n^{1/\epsilon}}{\epsilon}\log\frac{nW_{\max}}{\epsilon\Delta}\right)$. Note that Δ can be exponentially small, and our algorithm for solving the LP in Appendix B will lose such an additive factor in the objective anyway. Omitting details, we note that if the objective is profit, we can achieve optimal objective by directly rounding the *single LP* in Section 3.2 (details similar to Appendix C).

4   QUEUEING-THEORETIC JUSTIFICATION: DYNAMIC MARKETPLACES

In the facility location model discussed above, we imposed a lower bound $\mathbf{L}$ on the flow routed to any facility. We now present a *dynamic marketplace* model that provides queueing-theoretic justification for these constraints. This model has the following features:

- Buyer and seller types are located in a metric space just as before.
- Buyers at node j arrive as a Poisson process with rate $d_jF_j(p)$ when quoted price p, and when quoted a wage w; similarly, sellers follow a Poisson process with rate $s_jH_j(w)$.
- We assume each buyer and seller has a private patience level or deadline; if not matched within their deadline, they abandon the system. The platform knows the patience distribution.

The stochastic control problem that we term *dynamic marketplace problem* can be summarized by two control decisions:

(1) *Pricing decision.* Choose static prices p_j and wages w_j at each node $j \in V$; and
(2) *Scheduling decision.* This matches feasible buyer-seller pairs and removes them from the system. This decision is dynamic, depending on the entire state of the system as captured by the number of unmatched buyers and sellers at different nodes at any point of time.

The goal is to design a stochastic control policy that maximizes the long-term average surplus subject to long-term budget balance. We insist all scheduled matches must involve a current buyer and seller with metric distance at most R. The key difference is in the *service availability guarantee*: Given the stochastic nature of our arrivals, there is always some probability that an incoming buyer or seller exhausts her patience before being matched. A more realistic goal is to design policies that guarantee a minimum level of service availability. We quantify this via the long-term average probability of abandonment of agents. Formally, given a parameter $\epsilon > 0$ as input, the goal of the platform is to make the abandonment probability at most ϵ.

Scheduling Policies. Constructing the optimal policy for the dynamic marketplace problem is closely related to several lines of work in *dynamic matchings* over a compatibility graph – in kidney

exchanges [4] where patients abandon the exchange if their health fails; in control of matching queues for housing allocation [13]; and more generally in service system design [2, 3, 22], wherein customers and servers arrive stochastically and are matched according to a compatibility graph. In all these models, the choice of whom to match an arriving agent to depends on the entire set of agents waiting at different nodes, leading to the "curse of dimensionality".

Given this curse of dimensionality, we consider the restricted sub-class of matching policies where the platform creates facilities in the metric space, and uses each facility to cater to a different set of mutually-compatible agent types. Arriving agents are randomly routed to a compatible facility, where they are queued up to be matched to agents on the other side. The probabilistic routing is fixed over time, and does not depend on the state of the facilities.

Each facility maintains a queue of active buyers and sellers that have been assigned there, ignores what location they came from, and matches them up using an optimal scheduling policy for minimizing abandonment rate using only the current state of that particular queue. We will enforce the constraint that for any facility, the long-term abandonment probability is at most ϵ, which in turn will ensure the overall abandonment probability is at most ϵ.

Now assume buyer deadlines are distributed as $\texttt{Exponential}(\kappa)$, and seller deadlines are distributed as $\texttt{Exponential}(\gamma)$. Though these distributions are known, the scheduling decisions at any facility are made without knowing the patience level of any individual agent. Then, any work-conserving policy (including FIFO) is optimal. If an agent's deadline expires and there is no agent to match it with in the queue, this agent is considered abandoned. We build on results from queueing theory [33] to bound this abandonment rate tightly as follows:

THEOREM 4.1. *Suppose λ and μ be the (Poisson) arrival rates of buyers and sellers into a facility. Assume buyer deadlines are distributed as* $\texttt{Exponential}(\kappa)$*, and seller deadlines are distributed as* $\texttt{Exponential}(\gamma)$*. Then the FIFO policy has abandonment rate at most ϵ when:*
  (1) *There is* flow balance*, that is, $\lambda = \mu$; and*
  (2) *There is a* flow lower bound*, that is, $\lambda \geq \frac{3}{2}\left(\frac{\min(\gamma,\kappa)}{\epsilon^2}\right)$.*

PROOF. The behavior of a facility is captured via the following birth-death Markov chain: consider the state-space $\{\ldots, s(2), s(1), 0, b(1), b(2), \ldots\}$, where $0 = s(0) = b(0)$ denotes the state that the facility is empty, while for any $n \geq 1$, the state $b(n)$ denotes that there are n buyers queued up, and state $s(n)$ denote that there are n sellers queued up. For any $n \geq 1$, the transition rate from $b(n)$ to $b(n + 1)$ is λ, and for $s(n)$ to $s(n + 1)$ is μ; on the other hand, the rate of transition from $s(n)$ to $s(n-1)$ is $n\gamma + \lambda$, while from $b(n)$ to $b(n-1)$ is $n\kappa + \mu$. Here, the term $n\gamma$ corresponds to the rate of abandonment of sellers as their deadlines expires, and similarly $n\kappa$ is the rate of abandonment of buyers.

Assume that $\lambda = \mu$. Let q_0 denote the steady state probability of the queue being empty. Let

$$\Pr[\text{State } = s(n)] = \alpha_n \qquad \Pr[\text{State } = b(n)] = \beta_n$$

where $\alpha_0 = \beta_0 = q_0$. We have the following balance equations:

$$\alpha_n n\gamma = \lambda(\alpha_{n-1} - \alpha_n) \qquad \text{and} \qquad \beta_n n\kappa = \lambda(\beta_{n-1} - \beta_n)$$

Adding these equations, we have: $\sum_{n=1}^{\infty} (\alpha_n n\gamma + \beta_n n\kappa) = 2\lambda q_0$. Note that the LHS here is the total abandonment rate, and since the total arrival rate of agents (buyers and sellers) is 2λ, this means the abandonment probability is exactly q_0.

Since $\alpha_n = \frac{\lambda}{\lambda + n\gamma}\alpha_{n-1}$ and $\beta_n = \frac{\lambda}{\lambda + n\kappa}\beta_{n-1}$, we have by telescoping:

$$\alpha_n = q_0 \frac{\lambda^n}{\prod_{j=1}^n (\lambda + j\gamma)} = q_0 \prod_{j=1}^n \frac{1}{\left(1 + j\frac{\gamma}{\lambda}\right)}$$

$$\beta_n = q_0 \frac{\lambda^n}{\prod_{j=1}^n (\lambda + j\kappa)} = q_0 \prod_{j=1}^n \frac{1}{\left(1 + j\frac{\kappa}{\lambda}\right)}$$

Since these probability values sum to one, this implies

$$\frac{1}{q_0} = 1 + \sum_{n=1}^\infty \left(\prod_{j=1}^n \frac{1}{\left(1 + j\frac{\gamma}{\lambda}\right)} + \prod_{j=1}^n \frac{1}{\left(1 + j\frac{\kappa}{\lambda}\right)} \right) \tag{15}$$

For given κ and γ, this is an increasing function of λ. Therefore, $q_0 \le \epsilon$ translates to a bound of the form $\lambda \ge \mathbf{L}$. An upper bound $\mathbf{L}_\epsilon$ on $\mathbf{L}$ can be computed as follows. Let $c = \frac{\min(\gamma, \kappa)}{\lambda}$. Then,

$$\frac{1}{q_0} \ge 1 + \sum_{n=1}^\infty \left(\prod_{j=1}^n \frac{1}{\left(1 + j\frac{\min(\gamma, \kappa)}{\lambda}\right)} \right)$$

$$\ge \sum_{n=0}^\infty e^{-cn^2/2} - e^{-c} \ge \int_0^\infty e^{-cx^2/2} dx - e^{-c}$$

$$= \sqrt{\frac{\pi}{2c}} - e^{-c} \ge \sqrt{\frac{2}{3c}}$$

where the second inequality uses $1 + x \le e^x$ for all $x \ge 0$. Therefore, if we insist $\sqrt{\frac{2\lambda}{3\min(\gamma, \kappa)}} \ge \frac{1}{\epsilon}$, this ensures the abandonment probability is at most ϵ. This translates to the following lower bound:

$$\lambda \ge \mathbf{L}_\epsilon = \frac{3}{2} \cdot \frac{\min(\gamma, \kappa)}{\epsilon^2}$$

□

We have therefore shown that a sufficient condition for bounding the abandonment probability at any facility by ϵ reduces to saying the flow to the facility is balanced, and facility is *thick* – there is a lower bound $\mathbf{L} = \frac{3}{2}\left(\frac{\min(\gamma, \kappa)}{\epsilon^2}\right)$ on how much demand or supply needs to be routed there. This reduces the *dynamic* control problem with atomic agents to a *static* problem where demand/supply are fluid – exactly TWO-SIDED FAC-LOC($\mathbf{L}, R$) for suitable $\mathbf{L}$ that depends on ϵ.

5 CONCLUSIONS AND OPEN QUESTIONS

The advantage of our LP rounding framework for *Two-sided Facility Location* is that it gracefully handles more complex variants motivated by the dynamic marketplace setting. For instance, Uber has recently introduced pricing for buyers with different patience levels [1]. When mapped to our model, consider the dynamic problem where the platform uses prices and wages to truthfully elicit patience of agents, and subsequently matches them optimally using Earliest Deadline First (EDF) scheduling in each virtual market. Motivated by this, we consider an *envy-free* variant of *Two-sided Facility Location* in Appendix C, where each node (or agent type) is composed of sub-types that envy one another. The platform sets prices/wages for each sub-type so that each agent truthfully chooses its sub-type. Each sub-type has a weight and the service availability constraints are captured by flow balance of supply and demand at each virtual market, and a lower bound on the total *weight*

routed to each open market. Our LP rounding framework easily extends to yield optimal profit while relaxing the distance constraint by factor 4.

Our work is a first step in understanding the problem of jointly pricing and scheduling in dynamic matching facilities. We now mention several open questions that arise. For the facility location problems, there are other variants that we do not yet have good algorithms for, for instance, general policies for the envy-free version where the routing can be correlated with the sub-type elicited. Further, our model imposes a uniform distance bound the match of any agent. Extending it to average match distance will require new techniques; the basic filtering step in facility location rounding fails in our case since the demand value itself is a variable. Finally, it is an interesting question to extend our techniques to when facilities can be priced, and agents choose facilities to optimize their utility, in particular, extending techniques for stochastic scheduling in one-sided facilities [14, 15].

For the dynamic marketplace problems, we reduced a special type of scheduling policy that used facilities to the facility location problems. One can ask: *What about approximating the overall optimal policy?* Such a policy need not use facilities, and is poorly understood even when pricing is not involved (for instance, see [4]). Our model here also assumed Poisson arrivals whose rate is constant over time. A different approach is to use online algorithms. In particular, it would be interesting to incorporate pricing and wages into the "online matching with delays" models considered in [7, 19].

REFERENCES

[1] Uber is experimenting with letting riders wait longer in exchange for cheaper fares. *Quartz Magazine*, June 2018.

[2] Ivo Adan, Ana Busic, Jean Mairesse, and Gideon Weiss. Reversibility and further properties of fcfs infinite bipartite matching. *arXiv preprint arXiv:1507.05939*, 2015.

[3] Ivo Adan and Gideon Weiss. Exact FCFS matching rates for two infinite multitype sequences. *Operations research*, 60(2):475–489, 2012.

[4] Mohammad Akbarpour, Shengwu Li, and Shayan Oveis Gharan. Dynamic matching market design. In *ACM Conference on Economics and Computation, EC '14, Stanford , CA, USA, June 8-12, 2014*, page 355, 2014.

[5] Noga Alon, Uriel Feige, Avi Wigderson, and David Zuckerman. Derandomized graph products. *Comput. Complex.*, 5(1):60–75, January 1995.

[6] Mark Armstrong. Competition in two-sided markets. *The RAND Journal of Economics*, 37(3):668–691, 2006.

[7] Yossi Azar, Ashish Chiplunkar, and Haim Kaplan. Polylogarithmic bounds on the competitiveness of min-cost perfect matching with delays. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2017, Barcelona, Spain, Hotel Porta Fira, January 16-19*, pages 1051–1061, 2017.

[8] Siddhartha Banerjee, Sreenivas Gollapudi, Kostas Kollias, and Kamesh Munagala. Segmenting two-sided markets. In *Proceedings of the 26th International Conference on World Wide Web, WWW 2017, Perth, Australia, April 3-7, 2017*, pages 63–72, 2017.

[9] Siddhartha Banerjee, Ramesh Johari, and Carlos Riquelme. Pricing in ride-sharing platforms: A queueing-theoretic approach. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, EC '15, pages 639–639, 2015.

[10] Liad Blumrosen and Yehonatan Mizrahi. Approximating gains-from-trade in bilateral trading. In *Proceedings of the 12th International Conference on Web and Internet Economics*, WINE 2016, pages 400–413, 2016.

[11] Nam Kyoo Boots and Henk Tijms. A multiserver queueing system with impatient customers. *Management Science*, 45(3):444–448, 1999.

[12] Johannes Brustle, Yang Cai, Fa Wu, and Mingfei Zhao. Approximating gains from trade in two-sided markets via simple mechanisms. *CoRR*, abs/1706.04637, 2017.

[13] René Caldentey, Edward H Kaplan, and Gideon Weiss. Fcfs infinite bipartite matching of servers and customers. *Advances in Applied Probability*, 41(3):695–730, 2009.

[14] Shuchi Chawla, Nikhil R. Devanur, Alexander E. Holroyd, Anna R. Karlin, James Martin, and Balasubramanian Sivan. Stability of service under time-of-use pricing. In *ACM STOC*, 2017.

[15] Shuchi Chawla, Nikhil R. Devanur, Janardhan Kulkarni, and Rad Niazadeh. Truth and regret in online scheduling. In *ACM EC*, 2017.

[16] Riccardo Colini-Baldeschi, Bart de Keijzer, Stefano Leonardi, and Stefano Turchetta. Approximately efficient double auctions with strong budget balance. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2016, Arlington, VA, USA, January 10-12, 2016*, pages 1424–1443, 2016.

[17] Riccardo Colini-Baldeschi, Paul W. Goldberg, Bart de Keijzer, Stefano Leonardi, Tim Roughgarden, and Stefano Turchetta. Approximately efficient two-sided combinatorial auctions. In *Proceedings of the 2017 ACM Conference on Economics and Computation*, EC '17, pages 591–608, 2017.

[18] Brian C. Dean, Michel X. Goemans, and Jan Vondák. Approximating the stochastic knapsack problem: The benefit of adaptivity. *Mathematics of Operations Research*, 33(4):945–964, 2008.

[19] Yuval Emek, Shay Kutten, and Roger Wattenhofer. Online matching: haste makes waste! In *Proceedings of the 48th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2016, Cambridge, MA, USA, June 18-21, 2016*, pages 333–344, 2016.

[20] Zachary Friggstad, Mohsen Rezapour, and Mohammad R. Salavatipour. Approximating connected facility location with lower and upper bounds via LP rounding. In *15th Scandinavian Symposium and Workshops on Algorithm Theory, SWAT 2016, June 22-24, 2016, Reykjavik, Iceland*, pages 1:1–1:14, 2016.

[21] Sudipto Guha, Adam Meyerson, and Kamesh Munagala. Hierarchical placement and network design problems. In *41st Annual Symposium on Foundations of Computer Science, FOCS 2000, 12-14 November 2000, Redondo Beach, California, USA*, pages 603–612, 2000.

[22] Itai Gurvich, Amy Ward, et al. On the dynamic control of matching queues. *Stochastic Systems*, 4(2):479–523, 2014.

[23] Mohammad Taghi Hajiaghayi, Robert D. Kleinberg, Mohammad Mahdian, and David C. Parkes. Online auctions with re-usable goods. In *Proceedings 6th ACM Conference on Electronic Commerce (EC-2005), Vancouver, BC, Canada, June 5-8, 2005*, pages 165–174, 2005.

[24] Elad Hazan, Shmuel Safra, and Oded Schwartz. On the complexity of approximating k-set packing. *Comput. Complex.*, 15(1):20–39, May 2006.

[25] David R. Karger and Maria Minkoff. Building steiner trees with incomplete global knowledge. In *41st Annual Symposium on Foundations of Computer Science, FOCS 2000, 12-14 November 2000, Redondo Beach, California, USA*, pages 613–623, 2000.

[26] Samir Khuller, Anna Moss, and Joseph (Seffi) Naor. The budgeted maximum coverage problem. *Information Processing Letters*, 70(1):39 – 45, 1999.

[27] Aukasz Kruk, John Lehoczky, Kavita Ramanan, and Steven Shreve. Heavy traffic analysis for EDF queues with reneging. *Ann. Appl. Probab.*, 21(2):484–545, 2011.

[28] R Preston McAfee. The gains from trade under fixed price mechanisms. *Applied Economics Research Bulletin*, 1(1):1–10, 2008.

[29] Adam Meyerson. Profit-earning facility location. In *Proceedings of the Thirty-third Annual ACM Symposium on Theory of Computing*, STOC '01, pages 30–36, 2001.

[30] Roger B Myerson and Mark A Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2):265 – 281, 1983.

[31] Jean-Charles Rochet and Jean Tirole. Platform Competition in Two-Sided Markets. *Journal of the European Economic Association*, 1(4):990–1029, 2003.

[32] Zoya Svitkina. Lower-bounded facility location. In *Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2008, San Francisco, California, USA, January 20-22, 2008*, pages 1154–1163, 2008.

[33] Amy R. Ward and Peter W. Glynn. A diffusion approximation for a markovian queue with reneging. *Queueing Systems*, 43(1):103–128, 2003.

[34] E. Glen Weyl. A price theory of multi-sided platforms. *American Economic Review*, 100(4):1642–72, September 2010.

[35] David P Williamson and David B Shmoys. *The design of approximation algorithms*. Cambridge university press, 2011.

A EXAMPLES

A.1 Tradeoff Between Surplus, Profit, and Thickness

We show that the surplus optimum solution can open a facility with negative profit. To be more specific, for any given constant $c < 1$ we present a simple example in which c fraction of the total surplus is generated by a facility with negative profit.

Let $\{v, v'\}$ be two nodes infinitely far apart. Let L be the lower bound for the total amount of demand (supply) at each open facility. For node v, assume that the volume of demand and supply are $d_v = s_v = L$, the valuation of buyers is uniformly distributed over the interval $[2, 3]$, and the cost of sellers is uniformly distributed over the interval $[0, 1]$. For node v', assume the volume of demand and supply are $d_{v'} = s_{v'} = L$, the valuation of buyers is uniformly distributed over the interval $[c' - 1, 2c' + 1]$, and the cost of sellers is uniformly distributed over the interval $[0, c']$ where $c' = \frac{2c}{1-c}$. We claim that the optimum solution for this example is to open a facility at each of the nodes and set the price and wage at node v to 2 and 1 respectively, and set the price and wage at node v' to $c' - 1$ and c' respectively.

First, we show that this solution is feasible. At each node the price is not more than the valuation of any arriving buyer. Therefore, all the buyers choose to participate. Similarly, since the wage is not less than the cost of any arriving seller, all the sellers choose to participate. This solution satisfies flow balance for each of the facilities because the volume of sellers and buyers are equal at the corresponding node, and all of them choose to participate. In addition, flow lower bound is also satisfied, since this volume is at least L. Finally, the profit of the facility at v is d_v and the loss of the facility at node v' is $d_{v'}$. Therefore, the total profit is 0 and profit of the facility at node v compensates for the loss at the other facility.

Now, we show that the welfare of the facility with negative profit is a fraction c of the total welfare. The welfare at node v is $d_v \times (2.5 - 0.5) = 2L$ and the welfare at node v' is $d_{v'} \times (3c'/2 - c'/2) = c'L$. Therefore, $c'/(c' + 2) = c$ fraction of the welfare is generated at node v'.

Finally, we need to show that this solution is optimum. The nodes are far from each other and we cannot send the buyers and sellers from different nodes to a common facility. The only option for opening a facility at each of the nodes is to set the price and wage at each node in a way that all the arriving buyers and sellers choose to participate (otherwise, the flow lower bound cannot be satisfied). Therefore, this problem has three feasible integral solutions: no facility is opened, a facility at node v is opened, and a facility at each of the nodes is opened. Note that the solution which only opens a facility at v' is not feasible because it does not satisfy budget balance – in order to generate L volume of demand (supply), the platform must lose money here. The welfare of those solutions are 0, $2L$, and $(2 + c')L$ respectively. Therefore, the third solution is optimum.

A.2 Integrality Gap Example in Section 2

Now we slightly modify the previous example to show that the LP has unbounded integrality gap. We only change the distribution of the valuation of the buyers at node v'. The valuation of the buyers is now uniformly distributed over the interval $[c' - 1 - \epsilon, 2c' + 1 + \epsilon]$ for a small positive constant ϵ. After this change, the integral solution which opens a facility at each node is not feasible anymore because it violates weak budget balance constraint. Therefore, the optimum integral solution has $2L$ welfare.

On the other hand we claim that there is a fractional solution which has $(\frac{1}{1+\epsilon} \times \frac{2c}{1-c} + 2)L$ welfare. Set the price and wage at node v to 2 and 1 and open the facility at that node ($y_v = 1$). For the node v' we can only open the facility partially. Set $y_{v'} = \frac{1}{1+\epsilon}$ and the price and wage at node v' to $c' - 1 - \epsilon$ and c' with probability $\frac{1}{1+\epsilon}$ and to p_{max} and 0 with probability $\frac{\epsilon}{1+\epsilon}$. In other words, set $\alpha_{v'1} = \beta_{v'1} = \frac{1}{1+\epsilon}$ and $\alpha_{v'0} = \beta_{v'0} = \frac{\epsilon}{1+\epsilon}$. This solution is feasible and generates $(\frac{1}{1+\epsilon} \times \frac{2c}{1-c} + 2)L$

welfare, while the optimum integer solution generates only $2L$ welfare. Note that c can be arbitrarily close to 1 and therefore the integrality gap is unbounded.

B   SOLVING THE LP FORMULATION IN SECTION 3

We now show how to use the Ellipsoid algorithm to efficiently solve the LP formulation in Section 3 to arbitrary additive accuracy even when the demand and supply distributions are continuous, so that the sets Q_j (resp. $\mathcal{R}_j$) are continuous. First we get rid of weak budget balance by take a Lagrangian of surplus and the profit. For any parameter $\lambda \geq 0$, define:

$$\mathcal{V}_j^\lambda(q) = \mathcal{V}_j(q) + \lambda d_j q F_j^{-1}(q)$$

and

$$C_j^\lambda(r) = C_j(r) + \lambda s_j r H_j^{-1}(r)$$

Since we assumed regular supply and demand distributions, it is easy to show that $\mathcal{V}_j^\lambda(q)$ is concave in q and $C_j^\lambda(r)$ is convex in r. The Lagrangian objective is then:

$$\text{Maximize} \sum_{j \in V} \left(\sum_{q \in Q_j} \sum_{i \in B_R(j)} z_{ijq} \mathcal{V}_j^\lambda(q) - \sum_{r \in \mathcal{R}_j} \sum_{i \in B_R(j)} z_{ijr} C_j^\lambda(r) \right)$$

$$
\begin{array}{rcll}
\sum_{q \in Q_j} \sum_{i \in \mathcal{F}} z_{ijq} & \leq & 1 & \forall j \in V \\
\sum_{r \in \mathcal{R}_j} \sum_{i \in \mathcal{F}} z_{ijr} & \leq & 1 & \forall j \in V \\
\sum_{q \in Q_j} z_{ijq} & \leq & y_i & \forall j \in V, i \in B_R(j) \\
\sum_{q \in \mathcal{R}_j} z_{ijr} & \leq & y_i & \forall j \in V, i \in B_R(j) \\
\sum_{j \in B_R(i)} d_j \sum_{q \in Q_j} q z_{ijq} & = & \sum_{j \in B_R(i)} s_j \sum_{r \in \mathcal{R}_j} r z_{ijr} & \forall i \in \mathcal{F} \\
\sum_{j \in B_R(i)} d_j \sum_{q \in Q_j} q z_{ijq} & \geq & \mathbf{L} y_i & \forall i \in \mathcal{F} \\
z_{ijq}, z_{ijr}, y_i & \geq & 0 & \forall i, j, q, r
\end{array}
$$

The dual is the following:

$$\text{Minimize} \sum_{j \in V} (a_j + b_j)$$

$$
\begin{array}{rcll}
a_j + \eta_{ij} + d_j q (\zeta_i - \rho_i) & \geq & \mathcal{V}_j^\lambda(q) & \forall j \in V, i \in B_R(j), q \in Q_j \\
b_j + \theta_{ij} - s_j r \zeta_i + C_j^\lambda(r) & \geq & 0 & \forall j \in V, i \in B_R(j), r \in \mathcal{R}_j \\
\mathbf{L} \rho_i & \geq & \eta_{ij} + \theta_{ij} & \forall j \in V, i \in B_R(j) \\
\eta_{ij}, \theta_{ij}, \rho_i & \geq & 0 & \forall j \in V, i \in B_R(j)
\end{array}
$$

For fixed dual variables, since $\mathcal{V}_j^\lambda(q)$ is concave in q and $C_j^\lambda(r)$ is convex in r, it is easy to check that for each i, j, the separation oracle either involves maximizing a concave function in q (for the first set of constraints) or minimizing a convex function in r (for the second set of constraints). In either case, finding the separating hyperplane involves one-dimensional convex optimization. This implies the LP admits to an efficient additive approximation even for continuous distributions over a bounded domain. We omit the standard details.

C   ENVY-FREE PRICING AND PROFIT MAXIMIZATION

The idea from Section 2 of independently scaling up/down LP variables corresponding to individual facilities is fairly general, and leads naturally to approximation algorithms for more complex variants that are motivated by different scheduling policies for the dynamic marketplace problem. In this section, we present one such formulation that generalizes the model discussed in Section 2. In

Appendix C.2, we show that this model corresponds to the dynamic marketplace setting when the platform uses prices to elicit patience of agents, and uses Earliest Deadline First (EDF) scheduling in each virtual market.

We assume each node (type) j of buyer/seller has a collection of subtypes $\mathcal{S}_j$. There is a DAG $G_j(\mathcal{S}_j, E_j)$ on $\mathcal{S}_j$ that captures *envy*. If there is an edge $(k_1, k_2) \in E_j$, then sub-type k_1 envies sub-type k_2. The platform announces a price (resp. wage) p_{jk} (resp. w_{jk}) for each sub-type $k \in \mathcal{S}_j$. In order to preserve incentive compatibility, we require that if $(k_1, k_2) \in E_j$, then $p_{jk_1} \leq p_{jk_2}$; resp. $w_{jk_1} \geq w_{jk_2}$. This prevents an agent of sub-type k_1 from reporting its type to be k_2. Note that since the graph G_j is a DAG, such a price (resp. wage) assignment is feasible. We term such an assignment of prices (resp. wages) at each j as a *price (resp. wage) ladder*.

As before, there is a non-increasing demand function $d_{jk}F_{jk}(p_{jk})$ for each buyer sub-type $k \in \mathcal{S}_j$, and a non-decreasing supply function $s_{jk}H_{jk}(w_{jk})$ for each seller sub-type $k \in \mathcal{S}_j$. Each sub-type $k \in \mathcal{S}_j$ is also associated with a weight $\mathcal{G}_{jk}$. The platform learns which sub-type any agent chooses.

Lottery Pricing and Assignment. The platform opens a set of "virtual markets". For each node j and $k \in \mathcal{S}_j$, buyers (resp. sellers) arriving at the node and choosing that type are probabilistically routed to virtual markets which are within distance R from the node. We assume the platform shows a lottery over price (resp. wage) ladders as follows: For each node $j \in V$ the platform maintains a distribution $\mathcal{Z}_j$ of virtual markets within distance bound R, and for each virtual market in this set, it maintains a distribution $\mathcal{L}_{ij}$ of price (resp. wage) ladders. Given an agent arriving at this node, the platform first chooses a market i from $\mathcal{Z}_j$, and then a ladder from $\mathcal{L}_{ij}$ and shows it to the agent. After the agent chooses the price or wage (hence revealing its sub-type), she is routed to market i. We note that the routing policy makes the market the agent is routed to be independent of the sub-type elicited. Though this assumption is somewhat restrictive, it prevents the agent from choosing a sub-type to optimize for the virtual market they get assigned to.

Service Availability Guarantee. As before, we capture service availability by ensuring that each virtual market i has balanced supply and demand, and is also sufficiently *thick*. However, we now capture thickness by a lower bound $\mathbf{L}$ on the *total weight* of the sub-types assigned there. Formally, let x_{ijk} denote the expected flow of sub-type $k \in \mathcal{S}_j$ to virtual market i.

Flow Balance. The expected amount of supply and demand assigned there are the same.
Weight Lower Bound. The expected weight assigned there is large: $\sum_{j \in V, k \in \mathcal{S}_j} \mathcal{G}_{jk} x_{ijk} \geq \mathbf{L}$.

The objective is to maximize the expected profit of the solution. We term this problem ENVY-FREE FL($\mathbf{L}, R$). We note that similar ideas can be used to maximize other objectives; we present the profit objective for simplicity.

In the dynamic marketplace setting presented in Appendix C.2, the sub-types correspond to different deadlines, and the weight of a sub-type is precisely the deadline value. We show there that the weight lower bound corresponds to the condition for the EDF scheduling policy to have low abandonment rate.

C.1 Approximation Algorithm

Our LP formulation and rounding are similar to the one for TWO-SIDED FAC-LOC($\mathbf{L}, R$), and we highlight the differences. As before, we assume there is a candidate set $\mathcal{P}$ and $\mathcal{W}$ of prices and wages for each node, respectively. The set of all candidate virtual markets in the metric space is denoted by $\mathcal{F}$; since we assume the metric space is explicitly specified as input, we set $\mathcal{F} = V$. For each node j, $B_R(j) \subseteq \mathcal{F}$ denotes the set of all the virtual markets $i \in \mathcal{F}$ such that $c(i, j) \leq R$. For each virtual market i, define $B_R(i)$ as the set of all the nodes $j \in V$ such that $c(i, j) \leq R$.

C.1.1 Linear Programming Relaxation. **Variables.** For each candidate virtual market $i \in \mathcal{F}$, let $y_i \in \{0, 1\}$ be the indicator variable that a virtual market exists at that location in the metric space. These are the only integer variables in our formulation. Variables x_{ij}^d and x_{ij}^s are non-zero only if $y_i = 1$ and $i \in B_R(j)$. In this case, those are respectively the probability that buyers and sellers at node j are routed to virtual market i. Note that there is some probability that all prices at node j are set to $p_{\max}$, which corresponds to not routing node j anywhere. Let z_{ijkp} be the probability that buyers at node j with sub-type $k \in \mathcal{S}_j$ are assigned to virtual market i and offered price p. Similarly, z_{ijkw} denotes the probability that sellers at node j with sub-type k are assigned to virtual market i and offered wage w.

Objective and Constraints. The objective is to maximize the profit.

$$\max \sum_{j \in V} \sum_{k \in \mathcal{S}_j} \sum_{i \in B_R(j)} \left(\sum_{p \in \mathcal{P}} p d_{jk} F_{jk}(p) z_{ijkp} - \sum_{w \in \mathcal{W}} w s_{jk} H_{jk}(w) z_{ijkw} \right) \tag{16}$$

The following constraints connect the variables together. We present these constraints only for buyers (that is, $p \in \mathcal{P}$); the constraints for sellers is obtained by replacing p and x^d with $w \in \mathcal{W}$ and x^s. Since we route the buyers at node j probabilistically to one of the virtual markets, or to no market by offering all deadlines a price $p_{\max}$:

$$\sum_{i \in B_R(j)} x_{ij}^d \le 1 \quad \forall j \in V \tag{17}$$

Next, a price should be offered to each buyer with sub-type k at node j assigned to market i:

$$\sum_{p \in \mathcal{P}} z_{ijkp} = x_{ij}^d \quad \forall j \in V, i \in B_R(j), k \in \mathcal{S}_j \tag{18}$$

Next, if demand is fractionally routed from j to i, then i should be open and within distance R:

$$x_{ij}^d \le y_i \quad \forall j \in V, i \in B_R(j) \tag{19}$$

We next enforce that the prices and wages form a distribution over ladders. Note that the policy first chooses the virtual market to route to, and then chooses from a distribution over ladders. This reduces to a stochastic dominance condition for the distributions corresponding to z:

$$\sum_{p' \le p, p' \in \mathcal{P}} z_{ijkp'} \quad \le \quad \sum_{p' \le p, p' \in \mathcal{P}} z_{ijk'p'} \quad \forall p \in \mathcal{P}, (k, k') \in E_j, \forall j \in V, \forall i \in B_R(j) \tag{20}$$

$$\sum_{w' \le w, w' \in \mathcal{W}} z_{ijkw'} \quad \ge \quad \sum_{w' \le w, w' \in \mathcal{W}} z_{ijk'w'} \quad \forall w \in \mathcal{W}, (k, k') \in E_j, \forall j \in V, \forall i \in B_R(j) \tag{21}$$

Finally, we encode the service availability constraints. We first capture *flow balance* at each virtual market: the rate of arrival of buyers and sellers are equal.

$$\sum_{j \in B_R(i)} \sum_{k \in \mathcal{S}_j, p \in \mathcal{P}} d_{jk} F_{jk}(p) z_{ijkp} = \sum_{j \in B_R(i)} \sum_{k \in \mathcal{S}_j, w \in \mathcal{W}} s_{jk} H_{jk}(w) z_{ijkw} \quad \forall i \in \mathcal{F} \tag{22}$$

We finally encode *weighted flow lower bound* on the total deadline of buyers and sellers at the market:

$$\sum_{j \in B_R(i)} \sum_{k \in \mathcal{S}_j} \mathcal{G}_{jk} \left(\sum_{p \in \mathcal{P}} d_{jk} F_{jk}(p) z_{ijkp} + \sum_{w \in \mathcal{W}} s_{jk} H_{jk}(w) z_{ijkw} \right) \ge \mathbf{L} y_i \quad \forall i \in \mathcal{F} \tag{23}$$

C.1.2 Rounding. If we ignore the integrality constraints on y_i, the above is a linear programming relaxation of the problem. We will now show how to round the resulting solution.

We generalize Lemma 3.4 using the following definitions of *fully utilized*. We say that $j \in V$ is *fully demand utilized* if $\sum_{i \in B_R(j)} x_{ij}^d = 1$; similarly, it is *fully supply-utilized* if $\sum_{i \in B_R(j)} x_{ij}^s = 1$. We say j is *partially demand-connected* to market $i \in \mathcal{F}$ if $x_{ij}^d > 0$, and *partially supply-connected* if $x_{ij}^s > 0$. Let $J_D(i)$ denote the set of nodes that are partially demand-connected to $i \in \mathcal{F}$, and $J_S(i)$ be the set that is partially supply-connected. As before, we define the profit of a virtual market $i \in \mathcal{F}$ as:

$$R_i = \sum_{j \in B_R(i)} \sum_{k \in \mathcal{S}_j} \left(\sum_{p \in \mathcal{P}} p d_{jk} F_{jk}(p) z_{ijkp} - \sum_{w \in \mathcal{W}} w s_{jk} H_{jk}(w) z_{ijkw} \right)$$

LEMMA C.1. *In the LP optimum, for any $i \in \mathcal{F}$, $R_i \geq 0$. Further, if $R_i > 0$, either $y_i = 1$; or there exists $j \in J_D(i)$, s.t. j is fully demand-utilized; or there exists $j \in J_S(i)$, s.t. j is fully supply-utilized.*

The proof of the above lemma follows the same argument as Lemma 3.4: If a market has negative R_i, we can set all its variables to zero without violating any constraints. If the condition in the Lemma is violated for $i \in \mathcal{F}$, then we can increase all variables corresponding to i by the same factor till the condition is satisfied. Since all constraints involve single markets, this process preserves them while increasing the objective. For this transformation to work, it is crucial Constraints (20) are defined separately for each (i, j) pair; in other words, we crucially need to assume the policy chooses a market first and then chooses a distribution over ladders for that market.

The rounding now proceeds in the same way as in Section 2: In Phase 1, we identify untouched and *fully utilized* j and merge all i to which it is partially connected to one market. Note that the total y_i of these markets is at least 1 by the LP constraints. At the end of this phase, we move the remaining partially open i as in Phase 2 of the rounding scheme. This preserves the profit, and satisfies the flow balance and lower bound constraints (B_R is replaced by B_{4R} in the constraints), yielding the following theorem:

THEOREM C.2. *There is a feasible solution $\{\bar{x}, \bar{y}, \bar{z}\}$ to the above linear program, whose objective is optimal, and all of whose constraints are satisfied. For each $i \in \mathcal{F}$, either $\bar{y}_i = 0$ or $\bar{y}_i \geq 1$.*

Final Policy. The final choice of prices and wages, and the routing policy is the following. We present it only for buyers; the policy for sellers is symmetric.

- At node j, choose a market i with probability $\bar{x}_{ij}$. If no market is chosen, the price is set to $p_{\max}$.
- If market i is chosen, choose α uniformly at random in $[0, 1]$. For each $k \in \mathcal{S}_j$, find $p_k \in \mathcal{P}$ such that $\sum_{p' < p_k, p' \in \mathcal{P}} \frac{\bar{z}_{jkp'}}{\bar{x}_{ij}} \leq \alpha$ and $\sum_{p' \leq p_k, p' \in \mathcal{P}} \frac{\bar{z}_{jkp'}}{\bar{x}_{ij}} > \alpha$. Post prices $\{p_1, p_2, \ldots, p_K\}$.
- If the buyer accepts price p_k, route her to virtual market i.

Constraints (20) imply that regardless of the choice of α, the prices $\{p_1, p_2, \ldots, p_K\}$ in the second step form a ladder, so that $p_1 \geq p_2 \geq \cdots \geq p_K$. A similar statement holds for wages. Therefore, the second step produces a lottery over ladders. Further, if Z_{ijkp} denote the event that the price for sub-type $k \in \mathcal{S}_j$ is p and market i is chosen, then it is an easy exercise to check that $\mathbf{E}[Z_{ijkp}] = \bar{z}_{jkp}$. Therefore, the randomized policy exactly implements the solution found in Theorem C.2, so that it maximizes profit. We omit the details and state the final theorem.

THEOREM C.3. *There is a polynomial time $(4, 1)$ approximation for ENVY-FREE FL($\mathbf{L}, R$). That is, we obtain the optimal expected profit by relaxing the distance constraint by a factor of 4.*

C.2 Justification via Dynamic Marketplace Model

In this section, we present a dynamic marketplace model that justifies the problem statement of
Envy-Free FL. As in Section 4, we assume buyers and sellers have an inherent patience level or
deadline. If they are not matched within their deadline, they drop out of the system. We assume
every agent m is associated with a patience level v_m; unlike Section 4, we do not assume these
are Exponentially distributed. The platform advertises a fixed set of patience levels, or deadlines,
denoted by $\mathcal{S}_j = \{v_{j1}, v_{j2}, \ldots, v_{jK}\}$, which is a guarantee on the time by which a buyer or seller
choosing that deadline is guaranteed to be matched. We assume $v_{j1} \leq v_{j2} \leq \cdots \leq v_{jK}$. For simplicity,
we use $k \in \mathcal{S}_j$ and $v_k \in \mathcal{S}_j$ interchangeably.

Incentive-compatibility. We assume the platform sets a lottery of prices and wages at each node
j, that are independent of time. Consider the issue of eliciting deadlines truthfully. Consider buyers
first. At node j, suppose the platform offers price p_{jk} for deadline v_{jk}. Every buyer can choose one
deadline in $\mathcal{S}_j$, in which case he pays price p_{jk}, and is guaranteed to be matched within time v_{jk}
from his arrival. We assume any buyer m has very large negative utility for being matched after his
patience level v_m, therefore he will choose a k such that $v_{jk} \leq v_m$. Subject to this, he will choose
k with smallest p_{jk}, since this maximizes his valuation minus price. A symmetric model can be
posited for sellers, where we replace price with wage, and the seller chooses the largest wage such
that the corresponding deadline is smaller than his own patience level.

Since the goal of the platform is to elicit patience levels truthfully, the platform chooses a *price
ladder* $p_{j1} \geq p_{j2} \geq \cdots \geq p_{jK}$ and *wage ladder* $w_{j1} \leq w_{j2} \leq \cdots \leq w_{jK}$ at each node j. This ensures
that agents with $v_m \in [v_{jk}, v_{jk+1}]$ report deadline v_{jk}.

Each deadline level $v_{jk} \in \mathcal{S}_j$ gets associated with non-increasing demand function $d_{jk}F_{jk}(p_{jk})$,
which is the Poisson rate at which buyers m with patience $v_m \in [v_{jk}, v_{jk+1}]$ arrive when the price
of deadline v_{jk} is p_{jk}. Similarly, deadline level $k \in \mathcal{S}_j$ is associated with a non-decreasing supply
function $s_{jk}H_{jk}(w_{jk})$, which is the Poisson rate at which sellers m with patience $v_m \in [v_{jk}, v_{jk+1}]$
arrive when the wage for deadline v_{jk} is w_{jk}. These deadline levels correspond to the sub-types
described before.

Scheduling Policy. As in Section 4, the platform opens a set of "virtual markets". For each node
j and deadline level k, buyers (resp. sellers) arriving at the node and choosing that deadline are
probabilistically routed to virtual markets which are within distance R from the node. Buyers
and sellers arriving at the virtual market are queued up, and optimally matched to minimize
abandonment. Since the platform knows which deadline was chosen by the agent, the optimal
matching policy is now a variant of Earliest Deadline First (EDF): When the deadline of some buyer
(resp. seller) expires, it is matched to that seller (resp. buyer) in the queue whose deadline will
expire earliest in the future. If an agent's deadline expires and there is no agent to match it with in
the queue, this agent is abandoned. It is an easy exercise to show that this policy maximizes the
number of matches made in any virtual market.

As in Section 4, the goal of the platform is to design a joint pricing and scheduling policy to
maximize profit, while ensuring bounded match distance and bounded abandonment probability.

C.2.1 Bounding Abandonment Rate. We will now show that the weight lower bound can be
interpreted as a sufficient condition for the abandonment rate of the EDF policy to be at most ϵ,
where the weight of a sub-type is simply its deadline value.

The main technical assumption we require in this part is that the desired abandonment probability,
ϵ is small, in particular that $\epsilon \ll \frac{v_{\min}}{v_{\max}}$. As noted above, the scheduling policy within a virtual market
is a variant of EDF. Unlike the Patience-oblivious model where the behavior of a virtual market

could be modeled as a variant of a $M/M/1$ queue, the optimal abandonment probability in a two-sided EDF queue clearly depends on the entire distribution of deadlines of buyers and sellers, which in turn depends on the pricing scheme and assignment policy. However, we crucially need a closed-form bound on this probability in order to plug into an LP relaxation for the overall problem. We use recent results from queueing due to Kruk *et al.* [27] to construct such a closed-form bound, whose very existence we find non-trivial and surprising!

Kruk *et al.* [27] present an approximation to the abandonment probability of a one-sided queue $M/M/1$ queue with EDF scheduling. They approximate the queueing process via a reflected Brownian motion. We adapt their result to our setting, and rephrase it below. Consider the queue associated with a virtual market. Let $\bar{S}$ denote the average deadline of a seller arriving to this queue, and $\bar{D}$ denote the average deadline of a buyer arriving to the queue. Note that the distribution of deadlines as well as arrival rate depends on the overall pricing and assignment policies.

Recall that we assumed ϵ is small, in particular that $\epsilon \ll \frac{v_{\min}}{v_{\max}}$. We first enforce that supply and demand arrive to the queue at the same rate; call this rate λ. Next suppose w.l.o.g. that $\bar{D} > \bar{S}$. Consider the policy that instantaneously matches arriving sellers to the queued buyer with earliest deadline; if the queue is empty, the seller is abandoned. This exactly mimics a one-sided $M/M/1$ queue with EDF scheduling. We quote the following result informally from [27]:

> Consider a one sided $M/M/1$ queue with arrival rate and service rate equal to λ. Suppose deadlines of jobs are independently distributed with mean $\bar{D}$, and the scheduler uses the EDF policy. Then holding λ and $\frac{v_{\max}}{v_{\min}}$ fixed, in the regime where $v_{\min}$ becomes very large, the abandonment probability approaches $\frac{1}{\lambda \bar{D}}$.

Though part of their argument is heuristic, they perform simulations to show that this approximation is indeed accurate. Since we need abandonment probability of $\frac{1}{\lambda \bar{D}}$ to be at most $\epsilon \ll 1$, and since we assumed $\frac{v_{\min}}{v_{\max}} \gg \epsilon$, this automatically enforces that all deadlines are much larger than the mean inter-arrival time, satisfying their precondition for our setting.

Since the optimal policy for a two-sided queue only has lower abandonment probability, we use $\frac{1}{\lambda \bar{D}}$ as an upper bound on this quantity. Since we assumed $\bar{D} \geq \bar{S}$, we will instead use $\frac{2}{\lambda(\bar{D}+\bar{S})}$ as the upper bound, which we will set to be at most ϵ.

We now show that this is the best possible upper bound that only depends on $\bar{D}$ and $\bar{S}$. Suppose buyers deadlines are deterministic with value $\bar{D}$, and seller deadlines are deterministic with value $\bar{S}$. Then the optimal policy matches without waiting in a FIFO fashion. This means the loss probability assuming the queue has buyers is the same as that of a $M/M/1$ queue with deadlines $\bar{D}$, which from [11] is exactly

$$P_1 = \frac{1}{1 + \lambda \bar{D}}$$

Similarly, when there are sellers in the queue, the loss probability is

$$P_2 = \frac{1}{1 + \lambda \bar{S}}$$

Conditioned on the queue being empty and a buyer arriving, the expected time after which the queue next becomes empty is $T_b = \frac{1}{P_1} = 1 + \lambda \bar{D}$, in which period the loss probability is P_1. Similarly, if a seller arrives when the queue is empty, the expected time after which the queue again becomes empty is $T_s = \frac{1}{P_2} = 1 + \lambda \bar{D}$, in which period the loss probability is P_2. Since a buyer or seller arrives with equal probability when a queue is empty, the expected loss probability is

$$P = \frac{T_b P_1 + T_s P_2}{T_b + T_s} = \frac{2}{2 + \lambda \bar{D} + \lambda \bar{S}} \approx \frac{2}{\lambda \left(\bar{D} + \bar{S}\right)}$$

assuming $\lambda(\bar{D} + \bar{S}) \gg 1$.

In summary, each virtual market needs to satisfy the following two sufficient conditions for its abandonment probability to be at most ϵ:

(1) The rate of arrival of supply and demand should be the same; call this rate λ.
(2) If $\bar{S}$ denote the average deadline of a seller, and $\bar{D}$ denote the average deadline of a buyer, then $\lambda(\bar{D} + \bar{S}) \geq \frac{2}{\epsilon}$.

Therefore, to reduce the scheduling policy to an instance of ENVY-FREE FL$(R, \mathbf{L})$, we set $\mathcal{G}_{jk} = v_{jk}$ and $\mathbf{L} = \frac{2}{\epsilon}$, so that the second condition above translates to the weight lower bound. This justifies the ENVY-FREE FL$(R, \mathbf{L})$ problem as capturing the optimal scheduling policy for the dynamic marketplace problem presented above.

Note that the resulting lower bound on λ derived by the above condition is a significant improvement over the patience-oblivious case, since the lower bound now depends on $\frac{1}{\epsilon}$ instead of $\frac{1}{\epsilon^2}$. This intuitively means that in order to achieve comparable profit and abandonment probability, we can aim for a higher quality of match by reducing the radius R. A similar observation that even partial information about deadlines significantly reduces abandonment is made in [4], albeit for a different model.