# Blind Dynamic Resource Allocation in Closed Networks via Mirror Backpressure

Yash Kanoria* Pengyu Qian†

**Abstract**

We study the problem of maximizing payoff generated over a period of time in a general class of closed queueing networks with finite, fixed number of supply units which circulate in the system. We consider general controls including entry control, pricing, and assignment. Motivating applications include shared transportation platforms and scrip systems.

Inspired by the mirror descent algorithm for optimization and the backpressure policy for network control, we introduce a novel family of *Mirror Backpressure* (MBP) control policies. The MBP policies are simple and practical, and crucially do not need any statistical knowledge of the demand (job) arrival rates.

Under mild conditions, we show that MBP policies lose at most an $O\left(\frac{K}{T} + \frac{1}{K}\right)$ fraction of the payoff achieved by the optimal policy that knows the system parameters, where $K$ is the number of supply units and $T$ is the total number of arrivals over the time horizon, and the guarantee holds for finite $T$ and $K$. The key technical challenge we overcome is that the number of supply units at any node can never be negative (i.e., reflecting boundary). Simulation results in a realistic environment support our theoretical findings.

**Keywords:** queueing network; control; backpressure; maximum weight; mirror descent; dynamic pricing; shared transportation platforms; scrip systems.

*Graduate School of Business, Columbia University. Email: `ykanoria@columbia.edu`

†Graduate School of Business, Columbia University. Email: `PQian20@gsb.columbia.edu`

## 1 Introduction

The control of complex systems with circulating resources such as shared transportation platforms and scrip systems has been the subject of intensive study in recent years. Closed queueing networks provide a powerful abstraction for these applications (Banerjee et al. 2016, Braverman et al. 2019). A widely adopted approach for this problem is to solve the deterministic optimization problem that arises in the continuum limit (often called *the static planning problem*), and show that the resulting policy is near-optimal in a certain asymptotic regime. However, this approach only works under the assumption that *(1) the system parameters (demand arrival rates) are precisely known, and (2) the system is in steady state.* As is pointed out in Banerjee et al. (2016), relaxing either of these assumptions would be of great interest.

In this paper, we relax *both* assumptions. We propose a family of simple, practical control policies that are *blind* in that they use *no* prior knowledge of the system, and prove strong transient and steady state performance guarantees for these policies. In simulations, our policies achieve excellent performance that beats the state of the art even in an unequal contest where the exact demand arrival rates are provided beforehand to the state-of-the-art policies whereas our proposed policies are given *no* prior information about demand arrival rates.

**Informal description of the model.** For ease of exposition, our baseline setting is one where entry control is the only available control lever. Later we allow other controls including dynamic pricing of relocating resources, and flexible assignment of resources, and show that our machinery extends seamlessly.

In our baseline entry control model, we consider a closed queueing network consisting of $m$ nodes with infinite-buffer queues, and a fixed number $K$ of supply units that circulate in the system. Demand units with different origin-destination node pairs arrive stochastically over slotted time. The controller dynamically decides whether to admit an incoming demand unit. Each admission decision has two effects: it generates a certain *payoff* depending on the origin and destination of the demand unit, and it causes a supply unit to *relocate* from the origin to the destination instantaneously, if the origin node is non-empty. The goal of the system is to maximize the collected payoff over a period of time.

**Preview of main result.** We now give a preview of our main performance guarantee. We propose a large class of simple and practical control policies that are blind (i.e., require no estimates of the demand arrival rates), and show that, under a mild connectivity assumption on the network, the optimality gap (with payoffs scaled to be order 1) is bounded above as

$$O\left(\frac{K}{T}+\frac{1}{K}\right) \qquad \text{as } K\to\infty\,,$$

where $K$ is the number of supply units in the system, $T$ is the number of demand units that arrive during the time horizon of interest. Our result is *non-asymptotic*, i.e., our performance guarantee holds for a finite number of supply units, and over a finite horizon, and thus covers both transient and steady state performance. In particular, taking $T\to\infty$, we obtain a steady state optimality gap of $O\left(\frac{1}{K}\right)$, matching the state-of-the-art policy of Banerjee et al. (2016), though that policy requires perfect estimates of demand arrival rates, in sharp contrast to our policy which is completely blind. Our bound further provides a guarantee on transient performance: the horizon-dependent term $K/T$ in our bound on optimality gap is small if the total number of arrivals $T$ over the horizon is large compared to the number of supply units $K$.

Next, we motivate and introduce our family of control policies. We begin by describing how our problem is one of controlling a closed queueing network. Next, we describe the celebrated backpressure methodology for blind control of queueing networks. We then outline the central challenge in using backpressure in settings like ours. Finally, we introduce our proposed policies which significantly generalize backpressure, and may be broadly useful.

**Analogy with control of a closed queueing network.** Our problem can be viewed as one of optimal control of a closed queueing network. In the terminology of classic queueing theory, the $K$ supply units are "jobs", and each location in our model has both a queue of jobs (supply units) as well as a "server" which receives a "service token" each time a demand unit arrives with that location as the origin. Our model also specifies the "routing" of jobs: service tokens are labeled with a destination queue to which the served job (supply unit) moves. Since jobs circulate in the system (they do not arrive or leave), our setup is a *closed* queueing network. (Networks where jobs arrive, go through one or more services, and then leave, are called *open* networks.)

**Backpressure.** Our control approach is inspired by the celebrated backpressure methodology of Tassiulas and Ephremides (1992). Backpressure is a popular approach to the control of queueing networks: it simply uses queue lengths as the shadow prices to the flow constraints (the flow constraint for each queue is that the inflow must be equal to the outflow in the long run), and chooses a control decision at each time which maximizes the myopic payoff inclusive of shadow costs. This simple approach has been used very effectively in a range of practical and theoretical settings arising in cloud computing, networking, etc. (see, e.g., Georgiadis et al. 2006). Backpressure is near-optimal in many settings because it has the property of executing dual stochastic gradient descent (SGD) on the controller's deterministic (continuum limit) optimization problem. As we discuss next, this property breaks down when the so-called "no-underflow constraint" binds, making it very challenging to use backpressure in our setting (indeed, this difficulty may be the reason that backpressure has not yet been proposed as a solution to the control of ride-hailing platforms).

**Main challenge: no-underflow constraint.** The control policy must satisfy the *no-underflow constraint*, namely, that each decision to admit a demand unit needs to be backed by an available supply unit at the origin node of the demand. This constraint couples together the present and future decisions, and presents a challenge in deploying the backpressure methodology in numerous settings, including ours.

In certain settings this constraint does *not* pose a problem: For example, when the goal is to maximize throughput (as in[1] Dai and Lin 2005 and the well known "crossbar switch" problem in Maguluri and Srikant 2016), there are no "payoffs" apart from the shadow prices, so backpressure only recommends to serve a queue with positive length (taking queue lengths as congestion costs, backpressure only serves a queue if it is longer than the destination queue) and so the no-underflow constraint does not bind. In several works that do include payoffs, the authors *make strong assumptions* to similarly ensure the constraint does not bind.[2] In our setting, payoffs are essential (there is value generated by serving a customer), and so the constraint *does* bind.

A machinery that introduces *virtual queues* has been developed to extend backpressure to settings where the constraint binds (see, e.g., Jiang and Walrand 2009). The main idea is to introduce a "fake" supply unit into the network each time the constraint binds, to preserve the SGD property of backpressure. In open queueing networks, these fake supply units eventually leave the system, and so have a small effect (under appropriate assumptions). In our closed network setting, these fake supply units, once created, never leave and so would build up in the system, leading to very poor performance. In Section 4.3, we provide a detailed discussion of the challenge posed by the no-underflow constraint, and how it prevents us from using backpressure as is.

**Our solution: Mirror Backpressure.** In solving this problem, we introduce a novel class of policies which we call *Mirror Backpressure*. MBP generalizes the celebrated backpressure (BP) policy and is as simple and practical as BP. Whereas BP uses the queue lengths as congestion costs, MBP employs a flexibly chosen *congestion function* to translate from queue lengths to congestion

[1]Dai and Lin (2005) assume that the network satisfies a so-called EAA condition, which ensures that the no-underflow constraint does not bind for backpressure.

[2]For example, Stolyar (2005) assumes that payoffs are generated only by the source nodes, which have infinite queue lengths. Huang and Neely (2011) consider networks where the payoffs are generated only by the output nodes, and show that a simple tweak of backpressure avoids underflow entirely under this assumption. Gurvich and Ward (2014) assume that the network satisfies a so-called DI condition.

costs. MBP features a simple and intuitive structure: for example, in the entry control setting, the platform admits a demand only if the payoff of serving it outweighs the difference between congestion costs at the destination and origin of the demand. Crucially, the congestion function is chosen so that MBP has the property that it executes dual *stochastic mirror descent* (Nemirovsky and Yudin 1983, Beck and Teboulle 2003) on the platform's optimization problem, with the chosen mirror map[3]. The mirror map can be flexibly chosen to fit the problem geometry arising from the no-underflow constraints. Roughly, we find better performance with congestion functions which are steep for small queue lengths, the intuition being that this makes MBP more aggressive in protecting the shortest queues (and hence preventing underflow).

Notably, we combine the standard technical machinery for proving convergence of mirror descent and the well-understood Lyapunov analysis in stochastic networks, and show that they can be generalized almost seamlessly to study MBP policies, providing a ready Lyapunov function and yielding a guarantee of near optimal performance. This is in sharp contrast to much of the queuing literature, where finding a suitable Lyapunov function to prove convergence for one problem at a time has been quite an "art form" (see, e.g., Maguluri and Srikant 2016).

Our work fits into the broad literature on the control of stochastic processing networks (Harrison 2000, 2003). Our MBP methodology of designing blind control policies and obtaining provable guarantees applies to open queueing networks as well. We are optimistic that our problem will be one of many in the control of queueing networks that MBP will help solve.

**Main contributions.** To summarize, we make two main contributions in this paper:

1. **Propose a class of near-optimal control policies that are completely blind.** In general settings that consider entry control, pricing, and flexible assignment, we propose a family of dynamic control policies, the Mirror Backpressure policies, that have a strong transient and steady state performance guarantee. The MBP policies are simple and practical, and do not require any prior knowledge of demand arrival rates, making MBP promising for applications.

2. **Provide a framework for systematic design and analysis of policies for control of queueing networks.** Our control framework has a tight connection with mirror descent, which makes the process of policy design and analysis systematic, and allows us to handle the challenging no-underflow constraint. The technical machinery for mirror descent can be

[3]The special case of the congestion function being the identity function corresponds to standard BP, which has the property of executing stochastic gradient descent, a special case of mirror descent (Eryilmaz and Srikant 2007).

seamlessly leveraged to design policies and obtain rigorous guarantees for a variety of settings. This is in contrast with various intricate approaches in the queueing literature that do not easily generalize.

## 1.1 Applications

In Section 6 we generalize the baseline model (which allows entry control only) and include pricing and flexible assignment as control levers. Our general model includes a number of key ingredients common to many applications. We illustrate its versatility with the following applications, which are also summarized in Table 1. In Section 7 we discuss in detail the application to scrip systems (Item 3 below).

1. *Online control of ride-hailing platforms.* Ride-hailing platforms make dynamic decisions to optimize their objectives (e.g., revenue, welfare, etc.). Here the nodes in the model correspond to geographical locations, while supply units and demand units correspond to cars and customers, respectively. For most ride-hailing platforms in North America, pricing is used to modulate demand, which we capture by the pricing model we will introduce in[4] Section 6. In certain countries such as China, however, pricing is a less acceptable lever, hence admission control of customers is used as a control lever instead. This corresponds to the entry control model described above. In both cases, the platform can also decide where (near the demand's origin) to dispatch a car from, and where (near the demand's destination) to drop off a customer. These scenarios are well captured by the joint entry-assignment (JEA) and joint pricing-assignment (JPA) models to be introduced in Section 6.

2. *Dynamic incentive program for bike-sharing systems.* A major challenge faced by bike-sharing systems such as Citi Bike in New York City is the frequent out-of-bike and out-of-dock events caused by demand imbalance. One popular solution is to incentivize certain trips by awarding points (with cash value) depending on a trip's origin and destination (Chung et al. 2018). The optimal design of a dynamic incentive program corresponds to the pricing setting of our model. As we briefly discuss in Section 6, a simple adaptation of our control methodology allows us seamlessly handle the additional constraints in a docked bikesharing system that there are only a certain number of docks at each location.

3. *Scrip systems.* A scrip system is a nonmonetary trade economy where agents use scrips

[4] Our model there is similar to that of Banerjee et al. (2016), except that the platform does not know demand arrival rates and we allow a finite horizon.

(tokens, coupons, artificial currency) to exchange services. These systems are typically implemented when monetary transfer is undesirable or impractical. For example, Agarwal et al. (2019) suggest that in kidney exchange, to align the incentives of hospitals, the exchange should deploy a scrip system that awards points to hospitals that submit donor-patient pairs to the central exchange, and deducts points from hospitals that conduct transplantations. Another well-known example is Capitol Hill Babysitting Co-op (Sweeney and Sweeney 1977, see also Johnson et al. 2014), where married couples pay for babysitting services by another couples with scrips. A key challenge in these markets is the design of the service provider selection rule: When there are several possible providers for a trade, who should be selected for service? If an agent is running low on scrip balance, should they be allowed to request services? We postpone the detailed discussion of how our model can facilitate the design of a near optimal provider selection rule to Section 7.

| *Application* | *Control lever* | *Corresponding setting in this paper* |
|---|---|---|
| Ride-hailing in USA, Europe | Pricing & Dispatch | Joint pricing-assignment |
| Ride-hailing in China | Admission & Dispatch | Joint entry-assignment |
| Bike-sharing | Reward points | Pricing (finite buffer queues) |
| Scrip system | Provider selection | Joint entry-assignment |

Table 1: Summary of the applications that can be captured by our model, the control levers therein and their corresponding settings in this paper. See Section 2-5 for the entry control setting, and Section 6 for all other settings. For each setting, we design MBP policies that achieve $O\left(\frac{K}{T}+\frac{1}{K}\right)$ sub-optimality bound.

For the ride-hailing application, our assumption that cars relocate immediately is a common simplification in the existing literature (e.g. Banerjee et al. 2016, Balseiro et al. 2019). Moreover, our realistic simulations based on NYC yellow cab data show that our theoretical findings strikingly retain their power even with relocation times, achieving excellent performance over both short and long time horizons (Section 8).

## 1.2 Literature Review

**MaxWeight/backpressure policy.** Backpressure (also known as MaxWeight, see Tassiulas and Ephremides 1992, Georgiadis et al. 2006) are well-studied dynamic control policies in constrained queueing networks for workload minimization (Stolyar 2004, Dai and Lin 2008), queue length minimization (Eryilmaz and Srikant 2012) and utility maximization (Eryilmaz and Srikant 2007), etc. One of the main attractive features of MaxWeight/backpressure policies is that they can achieve provably good performance without knowing anything beforehand about arrival/service rates. Most of this literature considers the open queueing networks setting, and there is much less work on closed networks. An exception is a recent paper on state dependent control of closed networks (Banerjee et al. 2018), which shows the large deviations optimality of scaled MaxWeight policies. In Banerjee et al. (2018) the objective is throughput maximization, and the demand arrival rates are assumed to satisfy a strong near balance assumption hence it suffices to consider a non-idling policy (i.e., with assignment control only). In this paper, however, the objective is more general *and* we allow very general demand arrival rates, which makes it necessary to consider idling policies (e.g. entry control, pricing) to manage the system. Another main contribution of this paper to the literature is that the existing works use queue lengths or their power as congestion costs (Stolyar 2004), while our policies use a general increasing function (e.g., the logarithm) of queue lengths as congestion costs.

**Applications: shared transportation, scrip systems.** Most of the ride-hailing literature studied controls that require the exact knowledge of system parameters. Ozkan and Ward (2016) studied payoff maximizing assignment control in an open queueing network model, Braverman et al. (2019) derived the optimal state independent routing policy that sends empty vehicles to under-supplied locations, Banerjee et al. (2016) adopted the Gordon-Newell closed queueing network model and considered various controls that maximize throughput, welfare or revenue. Balseiro et al. (2019) considered a dynamic programming based approach for a specific network of star structure. Banerjee et al. (2018) which assumes a near balance condition on demands and equal pickup costs may be the only one that does not use the knowledge about system parameters. Comparing with Banerjee et al. (2016) which obtains a steady state regret scaling of $O\left(\frac{1}{K}\right)$ (in the absence of travel times) assuming *perfect* knowledge of demand arrival rates, our control policy achieves the same steady state regret with *no* knowledge of demand arrival rates, and further achieves a *transient* regret scaling of $O\left(\frac{K}{T}+\frac{1}{K}\right)$ for a finite horizon $T$.

Our model can be applied to the design of dynamic incentive programs for bike-sharing systems (Chung et al. 2018) and service provider rules for scrip systems (Johnson et al. 2014, Agarwal et al. 2019). For example, the "minimum scrip selection rule" proposed in Johnson et al. (2014) is a special case of backpressure policy, and our methodology leads to control rules in much more general settings as described in Section 7. See Section 1.1 for more discussion on these applications.

**Other related works.** There is a related stream of research on online stochastic bipartite matching, see, e.g., Caldentey et al. (2009), Adan and Weiss (2012), Bušić and Meyn (2015), Mairesse and Moyal (2016). The main difference between their setting and ours is that we study a *closed* system where supply units never enter or leave the system. Jordan and Graves (1995), Désir et al. (2016), Shi et al. (2019) and others study how process flexibility can facilitate improved performance, analogous to our use of assignment control to maximize payoff (when all pickup costs are equal), but the focus there is more on network design than on control policies. Along similar lines, network revenue management is a classical dynamic resource allocation problem, see, e.g., Gallego and Van Ryzin (1994), Talluri and Van Ryzin (2006), and recent works, e.g., Bumpensanti and Wang (2018). Again, this setting is "open" in that each supply unit can be used only once, in contrast to our closed setting.

### 1.3 Organization of the Paper

The remainder of our paper is organized as follows. We describe the connection to our motivating applications in Section 1.1. From Section 2 to Section 5 we focus on the entry control setting as an illustrative example of our approach: Section 2 presents our model and the platform objective. Section 3 introduces the Mirror Backpressure policy and presents our main theoretical result, i.e., a performance guarantee for the MBP policies. Section 4 introduces the static planning problem and describes the connection between the MBP policies and mirror descent. Section 5 outlines the proof of our main result. In Section 6, we consider MBP policies for general settings, e.g. pricing and flexible assignment, demonstrating the versatility of our approach. In Section 7, we discuss the application to scrip systems. In Section 8, we describe our simulation study of MBP policies using NYC yellow cab data. We conclude in Section 9 and discuss the future directions.

**Notation.** All vectors are column vectors if not specified otherwise. The transpose of vector or matrix $\mathbf{x}$ is denoted as $\mathbf{x}^\top$. We use $\mathbf{e}_i$ to denote the $i$-th unit column vector with the $i$-th coordinate being 1 and all other coordinates being 0, and $\mathbf{1}$ ($\mathbf{0}$) to denote the all 1 (0) column vector, where the dimension of the vector will be indicated in the superscript when it is not clear from the context, e.g., $\mathbf{e}_i^n$.

## 2 Illustrative Model: Dynamic Entry Control

In this section, we formally define our model of dynamic entry control in (closed) queueing networks. We will use this model as an illustrative example of our methodology.

We consider a finite-state Markov chain model with slotted time $t = 0, 1, 2, \ldots$, where a fixed number (denoted by $K$) of identical *supply units* circulate among $m$ *nodes* (locations), indexed by $j \in V$.

**Queues (system state).** At each node $j$, there is an infinite-buffer queue of supply units. (See Section 6 for extension to finite-buffer queues.) The *system state* is the vector of queue lengths at time $t$, which we denote by $\mathbf{q}[t] = [q_1[t], \cdots, q_m[t]]^\top$. Denote the state space of queue lengths by $\Omega_K \triangleq \{\mathbf{q} : \mathbf{q} \in \mathbb{N}^m, \mathbf{1}^\top \mathbf{q} = K\}$, and the normalized state space by $\Omega \triangleq \{\mathbf{q} : \mathbf{q} \in \mathbb{R}^m, \mathbf{1}^\top \mathbf{q} = 1\}$.

**Demand Types and Arrival Process.** We assume exactly one demand unit (customer) arrives at each period $t$, and denote his type by $(o[t], d[t]) \in V \times V$, where $o[t]$ is his origin and $d[t]$ is his destination. With probability $\phi_{jk}$, we have $(o[t], d[t]) = (j, k)$, independent of demands in earlier periods.[5] Let $\boldsymbol{\phi} \triangleq (\phi_{jk})_{j\in V, k\in V}$. Importantly, the system can observe the type of the arriving demand at the beginning of each time slot, but *the probabilities (arrival rates)* $\boldsymbol{\phi}$ *are not known.* Thus we substantially relax the assumption in previous works that the system has complete knowledge of demand arrival rates (see, e.g., Ozkan and Ward 2016, Banerjee et al. 2016, Balseiro et al. 2019).

**Entry Control and Payoff.** At time $t$, after observing the demand type $(o[t], d[t]) = (j, k)$, the system makes a binary decision $x_{jk}[t] \in \{0, 1\}$ where $x_{jk}[t] = 1$ stands for serving the demand, $x_{jk}[t] = 0$ means rejecting the demand. A supply unit moves and payoff is collected (or not) accordingly as follows:

- If $x_{jk}[t] = 1$, then a supply unit relocates from $j$ to $k$, immediately. Meanwhile, the platform collects payoff $v[t] = w_{jk}$ in this period. Without loss of generality, let $\max_{j,k\in V} |w_{jk}| = 1$.
- If $x_{jk}[t] = 0$, then supply units remain where they are and $v[t] = 0$.

Because the queue lengths are non-negative by definition, we require the following *no-underflow constraint* to be met at any $t$:

$$x_{jk}[t] = 0 \quad \text{if} \quad q_j[t] = 0\,. \tag{1}$$

As a convention, let $x_{j'k'}[t] = 0$ if $(o[t], d[t]) \neq (j', k')$. A *feasible policy* specifies, for each time $t \in \{0, 1, 2, \dots\}$, a mapping from the history so far of demand types $(o[t'], d[t'])_{t'\le t}$ and states $(\mathbf{q}[t'])_{t'\le t}$ to a decision $x_{jk}[t] \in \{0, 1\}$ satisfying (1), where $(j, k) = (o[t], d[t])$ as above. We allow $x_{jk}[t]$ to be randomized, although our proposed policy will be deterministic. The set of feasible policies is denoted by $\mathcal{U}$.

**System Dynamics and Objective.** The dynamics of system state $\mathbf{q}[t] \in \Omega_K$ is as follows:

$$\mathbf{q}[t+1] = \mathbf{q}[t] + \sum_{j\in V, k\in V} x_{jk}[t](-\mathbf{e}_j + \mathbf{e}_k)\,. \tag{2}$$

[5] This is equivalent to considering a continuous time model where the arrivals of different types of demands follow independent Poisson processes with rates proportional to the $(\phi_{jk})$s. The discrete time model considered is the embedded chain of the continuous time model.

We use $v^\pi[t]$ to denote the payoff collected at time $t$ under control policy $\pi$. Let $W_T^\pi$ denote the average payoff per period collected by policy $\pi$ in the first $T$ periods, and let $W_T^*$ denote the optimal payoff per period in the first $T$ periods over all admissible policies. Mathematically, they are defined respectively as:

$$W_T^\pi \triangleq \min_{\mathbf{q}\in\Omega_K} \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}[v^\pi[t]|\mathbf{q}[0]=\mathbf{q}]\,, \qquad W_T^* \triangleq \sup_{\pi\in\mathcal{U}} W_T^\pi\,. \tag{3}$$

Note that we consider the worst-case initial system state. Define the infinite-horizon per period payoff $W^\pi$ collected by policy $\pi$ and the optimal per period payoff of all admissible policy $W^*$ respectively as:

$$W^\pi \triangleq \operatorname{liminf}_{T\to\infty} W_T^\pi\,, \qquad W^* \triangleq \sup_{\pi\in\mathcal{U}} W^\pi\,. \tag{4}$$

We measure the performance of a control policy $\pi$ by its *optimality gap* (“loss”):

$$L_T^\pi = W_T^* - W_T^\pi \qquad \text{and} \qquad L^\pi = W^* - W^\pi\,. \tag{5}$$

# 3 The MBP Policies and Main Result

In this section, we propose a family of blind online control policies, and state our main result for these policies, which provides a strong transient and steady state performance guarantee for finite systems. We only require a mild connectivity assumption on the network for our result bounding the optimality gap.

## 3.1 The Mirror Backpressure Policies

We propose a family of online control policies which we call *Mirror Backpressure* (MBP) policies. Each member of the MBP family is specified by a mapping of normalized queue lengths $\mathbf{f}(\bar{\mathbf{q}}):\Omega\to\mathbb{R}^m$, where $\mathbf{f}(\bar{\mathbf{q}})\triangleq[f(\bar{q}_1),\cdots,f(\bar{q}_m)]^\top$ and $f$ is a monotone increasing function.[6] We will refer to $f(\cdot)$ as the *congestion function*, which maps each queue length to a congestion cost at that location, based on which MBP will make its control decisions.

We will later clarify the precise role of the congestion function: we will show that MBP executes dual stochastic mirror descent (Beck and Teboulle 2003) on the fluid limit problem with mirror map equal to the inverse of the congestion function. Similar to the design of effective mirror descent

[6]The methodology we will propose will seamlessly accommodate general mappings $\mathbf{f}(\cdot)$ such that $\mathbf{f}=\nabla F$ where $F(\cdot):\Omega\to\mathbb{R}$ is a strongly convex function, a special case of which is $\mathbf{f}(\bar{\mathbf{q}})\triangleq[f_1(\bar{q}_1),\cdots,f_m(\bar{q}_m)]^\top$ some monotone increasing $(f_j)$s. Here it suffices to consider a single congestion function, whereas in Section 6 we will employ queue-specific congestion functions $f_j(q_j)$.

algorithms, the choice of congestion function should depend on the constraints of the setting. This leads to an interesting interplay between problem geometry and policy design.

For conciseness, in this section we will state our main result for the congestion function

$$f(\bar{q}_j) \triangleq -\sqrt{m} \cdot \bar{q}_j^{-\frac{1}{2}}, \tag{6}$$

and postpone the results for other choices of congestion functions to Appendix D (see also Remark 1). For technical reasons, we need to keep $\bar{\mathbf{q}}$ in the *interior* of the normalized state space $\Omega$, i.e., we need to ensure that all normalized queue lengths remain positive. This is achieved by defining the normalized queue lengths $\bar{\mathbf{q}}$ as

$$\bar{q}_i \triangleq \frac{q_i + \delta_K}{\tilde{K}} \quad \text{for} \quad \delta_K \triangleq \sqrt{K} \quad \text{and} \quad \tilde{K} \triangleq K + m\delta_K\,. \tag{7}$$

Note that this definition leads to $\mathbf{1}^\top \bar{\mathbf{q}} = 1$ therefore $\bar{\mathbf{q}} \in \Omega$.

Our proposed MBP policy for the entry control problem is given in Algorithm 1. MBP admits a demand of type $(j, k)$ if and only if the *score*

$$w_{jk} + f(\bar{q}_j) - f(\bar{q}_k) \tag{8}$$

is nonnegative *and* the origin node $j$ has at least one supply unit (see Figure 1 for illustration of the score). The score (8) is nonnegative if and only if the payoff $w_{jk}$ of serving the demand outweighs the difference of congestion costs (given by $f(\bar{q}_k)$ and $f(\bar{q}_j)$) between the demand's destination $k$ and origin $j$. Roughly speaking, MBP is more willing to take a supply unit from a long queue and add it to a short queue, than vice versa. The policy is not only completely blind, but also semi-local, i.e., it only uses the queue lengths at the origin and destination. Figure 2 plots the (example) congestion function $f(\bar{q}_j)$ given in (6). Note that the congestion cost increases with queue length (as required), and furthermore decreases sharply as queue length approaches zero. Observe that such a choice of congestion function makes the MBP very reluctant to take supply units from short queues and helps to enforce the no-underflow constraint (1). See Section 4.3 for detailed discussion on the no-underflow constraint.

**ALGORITHM 1:** Mirror Backpressure (MBP) Policy for Entry Control

At the start of period $t$, the platform observes $(o[t], d[t]) = (j, k)$.

**if** $w_{jk} + f(\bar{q}_j[t]) - f(\bar{q}_k[t]) \geq 0$ **and** $q_j[t] > 0$ **then**

  $x_{jk}[t] \leftarrow 1$, i.e., serve the incoming demand;

**else**

  $x_{jk}[t] \leftarrow 0$, i.e., drop the incoming demand;

**end**

The queue lengths update as $\bar{\mathbf{q}}[t+1] = \bar{\mathbf{q}}[t] - \frac{1}{\tilde{K}} x_{jk}[t](\mathbf{e}_j - \mathbf{e}_k)$.

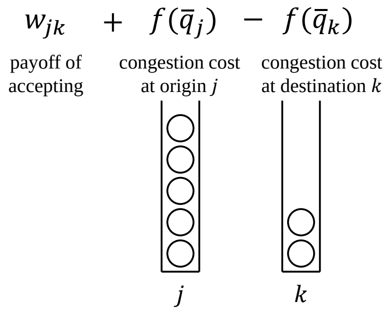


Figure 1: The score (8) based upon which MBP makes admission decisions. MBP admits a demand unit only if the score is non-nonnegative, i.e., MBP is eager to take a supply unit from a longer queue and add it to a shorter queue.

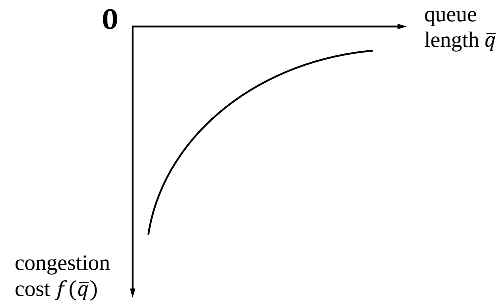


Figure 2: An example of a congestion function, i.e., a mapping from queue lengths to congestion costs, as given in (6). Congestion cost is increasing (nonlinearly, in general) in queue length.

### 3.2 Main Performance Guarantee of MBP Policies

We now formally state the main performance guarantee of our paper for the dynamic entry control model introduced in Section 2. We will outline the proof in Section 5, and extend the result to more general settings in Section 6.

Our result only requires a mild connectivity assumption on the demand arrival rates $\boldsymbol{\phi}$.

**Condition 1** (Strong Connectivity of $\boldsymbol{\phi}$)**.** *Define the* connectedness *of $\boldsymbol{\phi}$ as*

$$\alpha(\boldsymbol{\phi}) \triangleq \min_{S \subsetneq V, S \neq \emptyset} \sum_{j \in S, k \in V \setminus S} \phi_{jk} \,. \tag{9}$$

*We assume that $\boldsymbol{\phi}$ is* strongly connected*, namely, that $\alpha(\boldsymbol{\phi}) > 0$.*

Note that Condition 1 is equivalent to requiring that for every ordered pair of locations $(j, k)$, there is a sequence of demand types with positive arrival rate that would take a supply unit from $j$ eventually to $k$.

**Theorem 1.** *Consider a set of $m \in \mathbb{N}$ locations and any demand arrival rates $\boldsymbol{\phi}$ that satisfies Condition 1. Then there exists $K_1 = \mathrm{poly}\left(m, \frac{1}{\alpha(\boldsymbol{\phi})}\right)$, $M_1 = O(m)$ and $M_2 = O(m^2)$, such that the following holds.*[7] *For the congestion function $f(\cdot)$ defined in* (6)*, for any $K \geq K_1 = \mathrm{poly}\left(m, \frac{1}{\alpha(\boldsymbol{\phi})}\right)$, the following guarantees hold for Algorithm 1*

$$L_T^{\mathrm{MBP}} \leq M_1 \cdot \frac{K}{T} + M_2 \cdot \frac{1}{K} \,, \qquad \text{and} \qquad L^{\mathrm{MBP}} \leq M_2 \cdot \frac{1}{K} \,.$$

[7] Here "poly" indicates a polynomial. Also, $M_1$ and $M_2$ do not depend on $\alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})$ and $\bar{\mathbf{d}}$.

**Remark 1.** *In Section 6 we generalize Theorem 1 to broader settings that allow pricing and flexible assignment (Theorem 2, 3). In Appendix D, we further generalize Theorem 1, 2 and 3 by showing that similar performance guarantees hold for a whole class of congestion functions that satisfy certain growth conditions. For example, for both the logarithmic congestion function, i.e.* $f(\bar{q}) = c \cdot \log(\bar{q})$*, and the linear congestion function, i.e.* $f(\bar{q}) = c \cdot \bar{q}$ *with* $c > c_0$ *for some* $c_0 = \mathrm{poly}\left(m, \frac{1}{\alpha(\boldsymbol{\phi})}\right)$*, the same guarantee as in Theorem 1 holds with* $K_1 = \mathrm{poly}\left(c, m, \frac{1}{\alpha(\boldsymbol{\phi})}\right)$, $M_1 = \mathrm{poly}(c, m)$, $M_2 = \mathrm{poly}(c, m)$*. However, the specific polynomials depend on the choice of congestion functions.*

There are several attractive features of the performance guarantee provided by Theorem 1 for the simple and practically attractive Mirror Backpressure policy:

1. **The policy is completely blind.** At best the platform operator has access to an imperfect estimate of the demand arrival rates $\boldsymbol{\phi}$, so it is a very attractive feature of the policy that it does not need any estimate of $\boldsymbol{\phi}$ whatsoever. It is worth noting that the consequent bound of $O\left(\frac{1}{K}\right)$ on the steady state optimality gap remarkably matches that provided by Banerjee et al. (2016) even though MBP requires no knowledge of $\boldsymbol{\phi}$, whereas the policy of Banerjee et al. (2016) requires exact knowledge of $\boldsymbol{\phi}$. The policy of Banerjee et al. (2016) will generically suffer a long run (steady state) per period optimality gap of $\Omega(1)$ (as $K \to \infty$) if the estimate of demand arrival rates is imperfect.[8]

2. **Non-trivial regret bound for finite $T$.** In contrast with Banerjee et al. (2016) which provides only a steady state bound for finite $K$, we are able to provide a performance guarantee for finite horizon and finite (large enough) $K$. The horizon-dependent term $K/T$ in our bound on optimality gap is small if the total number of arrivals $T$ is large compared to the number of supply units $K$.

   It is worth noting that our bound *does not* deteriorate as the system size increases in the *large market regime*, where the number of supply units $K$ increases proportionally to the demand arrival rates (this regime is natural in ride-hailing settings taking the trip duration to be of order 1 time unit and where a non-trivial fraction of cars are busy at any time, see, e.g., Braverman et al. 2019). Define $T_K^{\mathrm{real}}$ as the physical time for a system with $K$

[8] Fix any demand arrival rates $\boldsymbol{\phi}$ with $\phi_{jk} > 0$ for all $j \in V, k \in V$. Let $\hat{\boldsymbol{\phi}}$ be the estimated demand arrival rates. Then the set of $\hat{\boldsymbol{\phi}}$ for which the optimality gap of the policy of Banerjee et al. (2016) (parameterized by $\hat{\boldsymbol{\phi}}$) is $\Omega(1)$ is open and dense in the set of possible estimates $\{\hat{\boldsymbol{\phi}} \in \mathbb{R}_{++}^{m \times m} : \mathbf{1}^\top \hat{\boldsymbol{\phi}} \mathbf{1} = 1\}$. Here $\mathbb{R}_{++}$ denotes the set of strictly positive reals. A similar result holds for any given sparsity pattern of $\boldsymbol{\phi}$ (i.e., when the arrival rates are zero for a subset of origin-destination pairs), where the platform knows the sparsity pattern and estimates only the positive entries. We omit the proofs; the main idea is that a fluid limit solution based on $\hat{\boldsymbol{\phi}}$ generically violates flow balance at the nodes, so the fluid solution itself has an $\Omega(1)$ optimality gap, and so a randomized policy based on the fluid solution as in Banerjee et al. (2016) inherits an $\Omega(1)$ optimality gap.

vehicles. As $K$ increases in the large market regime, the primitive $\boldsymbol{\phi}$ remains unchanged, while $T = \Theta(K \cdot T_K^{\mathrm{real}})$ since there are $\Theta(K)$ arrivals per unit of physical time. Hence, we can rewrite our performance guarantee as

$$W_T^* - W_T^{\mathrm{MBP}} \leq M \left( \frac{1}{T_K^{\mathrm{real}}} + \frac{1}{K} \right),$$

where the horizon-dependent term $\frac{M}{T_K^{\mathrm{real}}}$ is scale-invariant as $K$ increases.

3. **Flexibility in the choice of congestion function.** Because of the richness of the class of congestion functions covered in Appendix D which generalizes Theorem 1, the system controller now has the additional flexibility to choose a suitable congestion function $f$. For example, in our setting the performance guarantee for the congestion function given in (6) (Theorem 1) is more attractive than that for the linear congestion function $f(\bar{q}) = c \cdot \bar{q}$ (Remark 1) in the following way: in the latter case the coefficient $c$ needs to be larger than a threshold that depends on connectedness $\alpha(\boldsymbol{\phi})$ for a non-trivial performance guarantee to hold.[9] From a practical perspective, this flexibility can allow significant performance gains to be unlocked by making an appropriate choice of $f$, as evidenced by our numerical experiments in Section 8.

# 4 The MBP Policies and Mirror Descent

In this section, we discuss the main intuition behind the success of MBP policies, namely, that they execute (dual) mirror descent algorithms on a certain deterministic optimization problem. In Section 4.1, we define the deterministic optimization problem which arises in the continuum limit: the *static planning problem* (*SPP*), whose value we use to upper bound the optimal finite (and infinite) horizon per period $W_T^*$ (and $W^*$) defined in (3) and (4). In Section 4.2, we first review the interpretation of the celebrated Backpressure (BP) policy as a stochastic gradient descent algorithm on the dual of the SPP, and then proceed to generalize the argument to informally show that MBP executes mirror descent on the dual of SPP. In Section 4.3 we discuss the main challenge in turning the intuition into a proof, namely, the no-underflow constraint.

[9] Thus, in order to choose $c$ the platform needs to know $\alpha(\boldsymbol{\phi})$, whereas no knowledge of $\alpha(\boldsymbol{\phi})$ is needed when using the congestion function (6).

### 4.1 The Static Planning Problem

We first introduce a linear program (LP) that will be used to upper bound $W_T^*$ and $W^*$. The LP, called the static planning problem (SPP) (see, e.g., Harrison 2000, Dai and Lin 2005), is defined as:

$$\text{maximize}_{\mathbf{x}} \sum_{j,k \in V} w_{jk} \cdot \phi_{jk} \cdot x_{jk} \tag{10}$$

$$\text{s.t.} \sum_{j,k \in V} \phi_{jk} \cdot x_{jk}(\mathbf{e}_j - \mathbf{e}_k) = \mathbf{0} \qquad \text{(flow balance)} \tag{11}$$

$$x_{jk} \in [0,1] \qquad \forall j,k \in V\,. \qquad \text{(demand constraint)} \tag{12}$$

One interprets $x_{jk}$ as the fraction of type $(j,k)$ demand which is accepted, and the objective (10) as the rate payoff is generated under the fractions $\mathbf{x}$. In the SPP (10)-(12), one maximizes the rate of payoff generation subject to the requirement that the average inflow of supply units to each node in $V$ must equal the outflow (constraint (11)), and that $\mathbf{x}$ are indeed fractions (constraint (12)). Let $W^{\text{SPP}}$ be the optimal value of SPP. The following proposition formalizes that, as is typical in such settings $W^{\text{SPP}}$ is an upper bound on the optimal steady state (per period) payoff $W^*$. It further establishes that the optimal finite horizon per period payoff $W_T^*$ cannot be much larger than $W^{\text{SPP}}$.

**Proposition 1.** *For any horizon $T < \infty$ and any $K$, the finite and infinite horizon average payoff $W_T^*$ and $W^*$ are upper bounded as*

$$W_T^* \le W^{\text{SPP}} + m \cdot \frac{K}{T}\,, \qquad W^* \le W^{\text{SPP}}\,. \tag{13}$$

We obtain the finite horizon upper bound to $W_T^*$ in (13) by slightly relaxing the flow constraint (11).

### 4.2 MBP Executes Dual Stochastic Mirror Descent on SPP

The celebrated BP policy and our proposed MBP policies are closely related to the (partial) dual of the SPP:

$$\text{minimize}_{\mathbf{y}}\ g(\mathbf{y})\,, \qquad \text{where } g(\mathbf{y}) \triangleq \sum_{j,k \in V} \phi_{jk} \cdot \max_{x_{jk} \in [0,1]} x_{jk}\left(w_{jk} + y_j - y_k\right). \tag{14}$$

Here $\mathbf{y}$ are the dual variables corresponding to the flow balance constraints (11), and have the interpretation of "congestion costs" (see, e.g., Srikant 2012), i.e., $y_j$ can be thought of as the "cost" of having one extra supply unit at node $j$.

In the rest of this subsection, we *informally* discuss the interpretation of BP as stochastic gradient descent, and the interpretation of MBP as stochastic mirror descent.

**Review of the interpretation of BP as dual stochastic subgradient descent.** Rich dividends have been obtained by treating the (properly scaled) current queue lengths $\mathbf{q}$ as the dual variables $\mathbf{y}$, resulting in the celebrated backpressure (BP, also known as MaxWeight) control policy, introduced in the thesis of Tassiulas (Tassiulas 1992), see also Stolyar (2005), Eryilmaz and Srikant (2007). Formally, BP sets the current value of $\mathbf{y}$ to be proportional to the current normalized queue lengths, i.e., $\mathbf{y}[t] = c \cdot \bar{\mathbf{q}}[t]$ for some $\bar{\mathbf{q}} \in \Omega$ defined, e.g., as in (7), and greedily maximize the inner problem in (14) for every origin $j$ and destination $k$, i.e.,

$$x_{jk}^{\mathrm{BP}}[t] = \begin{cases} 1 & \text{if } w_{jk} + c \cdot \bar{q}_j[t] - c \cdot \bar{q}_k[t] \geq 0 \text{ and } q_j[t] > 0\,, \\ 0 & \text{otherwise}\,. \end{cases} \tag{15}$$

The main attractive feature of this policy is that it is extremely simple and does not need to know demand arrival rates $\boldsymbol{\phi}$. The BP policy can be viewed as a *stochastic subgradient descent (SGD)* algorithm on the dual problem (14) in the *interior* of the state space, i.e., when $q_j > 0$ for all $j \in V$ (Stolyar 2003, Huang and Neely 2009). To see this, denote the subdifferential (set of subgradients) of function $g(\cdot)$ at $\mathbf{y}$ as $\partial g(\mathbf{y})$. Observe that *the expected change of queue lengths under BP is proportional to the negative of a subgradient of* $g(\cdot)$ *at* $\mathbf{y} = c \cdot \bar{\mathbf{q}}[t]$, in particular

$$-\frac{\tilde{K}}{c} \cdot \mathbb{E}[\mathbf{y}[t+1] - \mathbf{y}[t]] = -\mathbb{E}[\mathbf{q}[t+1] - \mathbf{q}[t]] = \sum_{j,k \in V} \phi_{jk} \cdot x_{jk}^{\mathrm{BP}}[t](\mathbf{e}_j - \mathbf{e}_k) \in \partial g(\mathbf{y}[t])\,, \tag{16}$$

where the first equality follows from the definition $\mathbf{y}[t] = c \cdot \bar{\mathbf{q}}[t]$ and second equality is just the system dynamics (2). Here $\sum_{j,k \in V} \phi_{jk} \cdot x_{jk}^{\mathrm{BP}}[t](\mathbf{e}_j - \mathbf{e}_k) \in \partial g(\mathbf{y}[t])$ since $g$ is a maximum of linear functions of $\mathbf{y}$ parameterized by $\mathbf{x}$, hence $g$ is convex and the gradient of a linear function among these which is an argmax at $\mathbf{y}[t]$ (in particular, the linear function parameterized by $\mathbf{x}^{\mathrm{BP}}[t]$) is a subgradient of $g$ at $\mathbf{y}[t]$.

(16) shows that the evolution of $\mathbf{y}[t]$ when $q[t] > 0$ is exactly an iteration of SGD with step size $\frac{c}{\tilde{K}}$. This interpretation of BP as stochastic subgradient descent leads to desirable properties including stability, minimization of delay/workload, and revenue maximization in certain networks (see, e.g., Georgiadis et al. 2006, Eryilmaz and Srikant 2007, etc.). However, as we will see in Section 4.3, in our setting the SGD property of backpressure breaks on the *boundary* of state space, i.e., when there exists $j' \in V$ such that $q_{j'} = 0$, due to the *no-underflow* constraints $\mathbf{q} \geq \mathbf{0}$.

**MBP executes dual stochastic mirror descent on the SPP.** The key innovation of our approach is to design a family of policies generalizing BP (MBP given in Algorithm 1) that *executes stochastic mirror descent on the partial dual problem* (14) *(with flow constraints dualized), with* $\bar{\mathbf{q}}[t]$ *given by* (7) *being the mirror point and the inverse mirror map being the (vector) congestion function*

$\mathbf{f}(\bar{\mathbf{q}}) \triangleq [f(\bar{q}_1), \cdots, f(\bar{q}_m)]^\top$. Mathematically, if $\mathbf{q} > 0$, we have

$$-\tilde{K} \cdot \mathbb{E}[\bar{\mathbf{q}}[t+1] - \bar{\mathbf{q}}[t]] = -\mathbb{E}[\mathbf{q}[t+1] - \mathbf{q}[t]] = \sum_{j,k \in V} \phi_{jk} \cdot x_{jk}^{\mathrm{MBP}}[t](\mathbf{e}_j - \mathbf{e}_k) \in \partial g(\mathbf{y})\Big|_{\mathbf{y}=\mathbf{f}(\bar{\mathbf{q}}[t])}, \quad (17)$$

where $\mathbf{x}^{\mathrm{MBP}}[t]$ is the control defined in Algorithm 1; notice that the entry rule $\mathbf{x}^{\mathrm{MBP}}[t]$ has the same form as that for BP (15) except that it uses a general congestion function $f(q_j)$, leading to (17) for MBP via the same reasoning that led to (16) for BP. Thus, MBP performs stochastic mirror descent on the partial dual problem (14), which generalizes the previously known fact that BP performs stochastic gradient descent.

A main advantage of mirror descent over gradient descent is that it can better capture the geometry of the state space for an appropriately chosen mirror map (see, e.g., Nemirovsky and Yudin 1983, Beck and Teboulle 2003). The analysis of mirror descent is almost completely parallel to that of gradient descent: they both rely on Lyapunov functions, which represents the proximity between current variables and the set of their optimal values. The difference lies in the choice of Lyapunov functions: for gradient descent, squared Euclidean distance is often used as the proximity measure; for mirror descent, the Bregman divergence determined by the mirror map is used (e.g., the Kullback-Leibler divergence is the Lyapunov function for the exponential mirror map $q_j = e^{y_j} \Leftrightarrow y_j = f(q_j) = \log q_j$, i.e., for the logarithm congestion function, see Remark 1). In our setting, the congestion function $\mathbf{f}(\bar{\mathbf{q}})$ is the inverse mirror map and can be flexibly chosen.

Our approach blending backpressure and mirror descent with a flexibly chosen mirror map is novel. We believe it can serve as a general framework for systematic design of near optimal backpressure-like control policies for queueing networks in settings with hairy practical constraints.

### 4.3 Challenge: No-underflow Constraints

As we have discussed earlier, the no-underflow constraints pose a challenge when applying backpressure to various settings. The following simple example illustrates how BP fails when the proportionality constant $c$ is not chosen to be sufficiently large.

**Example 1** (BP is far from optimal if $c$ is not large enough). *Consider a network with three nodes $\{1, 2, 3\}$, demand arrival rates $\phi_{12} = \epsilon$, $\phi_{23} = \frac{1}{3} + \epsilon$, $\phi_{21} = \phi_{32} = \frac{1}{3} - \epsilon$ (where $\epsilon < \frac{1}{3}$), and payoffs $w_{12} = w_{23} = w_{21} = 0$, $w_{23} = w > 0$. Suppose the platform employs backpressure where the shadow prices are taken to be proportional to (normalized) queue lengths $\mathbf{y}[t] = c \cdot \bar{\mathbf{q}}[t]$. Let $\bar{\mathbf{q}}^*$ be the queue lengths correspond to the optimal dual variables in* (14) *with the additional constraint that the queue lengths sum to $K$. Simple algebra yields $\bar{\mathbf{q}}^* = (\frac{c-w}{3c}, \frac{c-w}{3c}, \frac{c+2w}{3c})$. When $c < w$, $\bar{\mathbf{q}}^*$ lies outside the normalized state space $\bar{\mathbf{q}}^* \notin \Omega$, hence the $\bar{\mathbf{q}}[t]$ will never converge to $\bar{\mathbf{q}}^*$ and BP is far* *from optimal.*

Even if the platform uses BP with sufficiently large $c$ to ensure that $\bar{\mathbf{q}}^* \in \Omega$, the existing analysis of BP still fails, as is demonstrated below.

**Example 2** (BP has positive Lyapunov drift at a certain state)**.** *Again consider Example 1 and let $c \geq w$. The existing analysis of BP is based on establishing that the "drift" defined by*

$$\mathbb{E}\Big[ \|\bar{\mathbf{q}}[t+1] - \bar{\mathbf{q}}^*\|_2^2 \,\Big|\, \bar{\mathbf{q}}[t] \Big] - \|\bar{\mathbf{q}}[t] - \bar{\mathbf{q}}^*\|_2^2$$

*is strictly negative when $\|\bar{\mathbf{q}}[t] - \bar{\mathbf{q}}^*\|_2 = \Omega(K)$. Suppose at time $t$ we have[10] $\bar{\mathbf{q}}[t] = (\frac{2c-w}{2c}, 0, \frac{w}{2c})$; in particular, queue 2 is empty. Note that at $\bar{\mathbf{q}}[t]$, BP suggests fulfilling the demand going from 2 to 3, but the platform cannot do so because of the no-underflow constraint. Straightforward calculation shows that the "drift" is positive for large enough $K$ if $\epsilon < \frac{w}{6c}$.*

In the following analysis, we show that the underflow problem is provably alleviated by MBP policies with an appropriately chosen congestion function. For example, the MBP policy with congestion function given in (6) is more aggressive in preserving supply units in near-empty queues compared to BP, making the system less likely to violate the no-underflow constraints. Besides analytical tractability, the MBP policy also achieves better performance than BP in simulations (Section 8).

# 5 Proof Sketch of Theorem 1

In this section we provide an overview of the key lemmas and ideas that lead to a proof of Theorem 1. Our proof consists of three steps:

1. In Section 5.1, we use Lyapunov analysis to upper bound the suboptimality that MBP incurs in one period by the sum of several auxiliary terms (Lemma 1). The auxiliary terms are easier to control and have clear interpretations.

2. In Section 5.2, we utilize the structure of the dual problem (14) to bound the auxiliary terms introduced in the first step (Lemma 3).

3. In Section 5.3, we average the one-step optimality gap obtained in previous steps over a finite/infinite horizon, and conclude the proof of Theorem 1.

[10] The integrality of the components of $\mathbf{q}[t]$ is non-essential, hence we assume all components of $\mathbf{q}[t]$ are integers. Also, here we take the normalized queue lengths to be defined as $\bar{\mathbf{q}}[t] \triangleq \mathbf{q}[t]/K$ to simplify the expressions.

### 5.1 Single Period Analysis of MBP via Lyapunov Function

This part of the proof relies on the key observation we made in Section 4, i.e., that MBP policy executes stochastic mirror descent on the dual objective function $g(\mathbf{y})$ (the dual problem was defined in (14)) except when underflow happens. As a result, our proof combines the standard approach for stochastic mirror descent algorithms (see, e.g., Nemirovsky and Yudin 1983, Beck and Teboulle 2003) and a novel argument that bounds the suboptimality contributed by underflow.

We use the antiderivative of $\mathbf{f}(\cdot)$, which, for the congestion function $f$ defined in (6), is

$$F(\bar{\mathbf{q}}) = -2\sqrt{m}\sum_{j=1}^{m}\sqrt{\bar{q}_j}\,, \tag{18}$$

as the Lyapunov function. Note that it is standard in the analysis of mirror descent to use the Bregman divergence generated by $F(\cdot)$ as the Lyapunov function. Here we combine that approach with the Lyapunov drift method (see, e.g., Neely 2010) in network control and use a slightly different Lyapunov function, which streamlines the proof and yields sharper results. Recall that $W^{\mathrm{SPP}}$ is the optimal value of SPP (10)-(12), $v^{\mathrm{MBP}}[t]$ denotes the payoff collected under the MBP policy in the $t$-th period, and $g(\cdot)$ is the dual problem (14). We have the following result:

**Lemma 1** (Suboptimality of MBP in one period)**.**

$$\begin{aligned} W^* - \mathbb{E}[v^{\mathrm{MBP}}[t]|\bar{\mathbf{q}}[t]] \leq\ & \underbrace{\tilde{K}\left(F(\bar{\mathbf{q}}[t]) - \mathbb{E}[F(\bar{\mathbf{q}}[t+1])|\bar{\mathbf{q}}[t]]\right)}_{\mathcal{T}_1} + \underbrace{\frac{1}{2\tilde{K}}\cdot\max_{j\in V}\left|f'(\bar{q}_j[t])\right|}_{\mathcal{T}_2} \\ & + \underbrace{\left(W^{\mathrm{SPP}} - g(\mathbf{f}(\bar{\mathbf{q}}[t]))\right)}_{\mathcal{T}_3} + \underbrace{\mathbb{1}\left\{q_j[t]=0, \exists j\in V\right\}}_{\mathcal{T}_4}\,. \end{aligned} \tag{19}$$

In Lemma 1, the LHS of (19) is the suboptimality incurred by MBP in a single period. On the RHS of (19), $\mathcal{T}_1$ and $\mathcal{T}_2$ come from the standard analysis of mirror descent; $\mathcal{T}_3$ is the negative of the dual suboptimality at $\mathbf{y} = (\bar{\mathbf{q}}[\mathbf{t}])$, hence it is always non-positive; $\mathcal{T}_4$ is the payoff loss because of underflow.

In the next subsection, we outline our proof that the sum of the last three terms $\mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4$ is small. As a result, $\mathcal{T}_1$ is the main term on the right-hand side. Observe that it is proportional to the *Lyapunov drift*: the negative of the expected change in the Lyapunov function in one time step. The main intuition leading to the finite horizon performance guarantee in Theorem 1 is then that if the suboptimality of MBP in some period is large, then (19) implies that there is also a large negative Lyapunov drift.

## 5.2 Bounding Single Period Payoff Loss

In this section we proceed to upper bound $\mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4$ on the RHS of (19). Observe that the terms $\mathcal{T}_2$ and $\mathcal{T}_4$ are non-negative, while $\mathcal{T}_3$ is non-positive, thus the goal is to show that $\mathcal{T}_3$ compensates for $\mathcal{T}_2 + \mathcal{T}_4$. First notice that $\mathcal{T}_2$ is large when there exist very short queues (because the congestion function (6) changes rapidly only for short queue lengths), and $\mathcal{T}_4$ is non-zero only when some queues are empty. Helpfully, it turns out that $\mathcal{T}_3$ is more negative in these same cases; we show this by exploiting the structure of the dual problem (14).

In Lemma 2 we provide an upper bound for $\mathcal{T}_3$ that becomes more negative as the shortest queue length decreases.

**Lemma 2.** *Consider the congestion function* (6)*, and any* $\boldsymbol{\phi}$ *with connectedness* $\alpha(\boldsymbol{\phi}) > 0$*. We have*

$$\mathcal{T}_3 \leq \; -\alpha(\boldsymbol{\phi}) \cdot \left[ f\left(\frac{1}{m}\right) - f\left(\min_{j \in V} \bar{q}_j\right) - 2m \right]^+ .$$

We prove Lemma 2 in Appendix C by utilizing complementary slackness for the SPP (10)-(12).

We are now ready to bound $\mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4$.

**Lemma 3.** *Consider the congestion function* (6)*, and any* $\boldsymbol{\phi}$ *with connectedness* $\alpha(\boldsymbol{\phi}) > 0$*. Then there exists* $K_1 = \mathrm{poly}\left(m, \frac{1}{\alpha(\boldsymbol{\phi})}\right)$ *such that for* $K \geq K_1$*,*

$$\mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4 \leq M_2 \cdot \frac{1}{\tilde{K}}$$

*for* $M_2 = Cm^2$ *where* $C > 0$ *is a universal constant. Here* $\tilde{K}$ *was defined in* (7)*.*

## 5.3 Proof of Theorem 1: Optimality Gap of MBP

Putting Lemma 1 and Lemma 3 together leads to the following proof of Theorem 1. The main idea is to use the so-called *Lyapunov drift argument* of Neely (2010), namely, to sum the expectation of (19) (the bound in Lemma 1) over the first $T$ time steps. The terms $\mathcal{T}_1$ form a telescoping sum.

*Proof of Theorem 1.* Plugging in Lemma 3 into (19) in Lemma 1 and taking expectation, we obtain

$$W^* - \mathbb{E}[v^{\mathrm{MBP}}[t]] \leq \tilde{K} \left( \mathbb{E}[F(\bar{\mathbf{q}}[t])] - \mathbb{E}[F(\bar{\mathbf{q}}[t+1])] \right) + M_2 \frac{1}{\tilde{K}} \qquad \text{for } K \geq K_1 \,. \tag{20}$$

Take the sum of both sides of the inequality (20) from $t = 0$ to $t = T - 1$, and divide the sum by $T$. This yields

$$L_T^{\mathrm{MBP}} = W^* - W_T^{\mathrm{MBP}} \leq \frac{\tilde{K}}{T} \left( \mathbb{E}[F(\bar{\mathbf{q}}[0])] - \mathbb{E}[F(\bar{\mathbf{q}}[T])] \right) + M_2 \frac{1}{\tilde{K}}$$

$$\leq \frac{\tilde{K}}{T} \sup_{\bar{\mathbf{q}}_1, \bar{\mathbf{q}}_2 \in \Omega} \big(F(\bar{\mathbf{q}}_1) - F(\bar{\mathbf{q}}_2)\big) + M_2 \frac{1}{\tilde{K}} .$$

Let $M_1 \triangleq \sup_{\bar{\mathbf{q}}_1, \bar{\mathbf{q}}_2 \in \Omega} \big(F(\bar{\mathbf{q}}_1) - F(\bar{\mathbf{q}}_2)\big)$. Observe that the function $F(\bar{\mathbf{q}})$ given in (18) is negative $F(\bar{\mathbf{q}}) \leq 0$ for all $\bar{\mathbf{q}} \in \Omega$, and is a convex function which achieves its minimum at $\bar{\mathbf{q}} = \frac{1}{m}\mathbf{1}$. Therefore we have

$$M_1 \leq -\inf_{\bar{\mathbf{q}} \in \Omega} F(\bar{\mathbf{q}}) \leq -F\left(\frac{1}{m}\mathbf{1}\right) = m .$$

Hence the finite-horizon optimality gap of MBP is upper bounded by $M_1 \frac{\tilde{K}}{T} + M_2 \frac{1}{\tilde{K}}$ where $M_1 = O(m)$, $M_2 = O(m^2)$ and $M_1$, $M_2$ do not depend on $\alpha(\boldsymbol{\phi})$. Moreover, $\tilde{K} = K + m\sqrt{K} \in [K, 2K]$ taking $K_1 \geq m^2$. This concludes the proof. $\square$

# 6 Pricing and Assignment Control

In this section, we allow the platform to have additional control levers beyond entry control and consider two general settings, namely, joint entry-assignment control (JEA) and joint pricing-assignment control (JPA). We also allow the queues to have finite buffers. We show that the extended models enjoy similar performance guarantees to that in Theorem 1 under mild conditions on the model primitives.

## 6.1 Congestion Functions for Finite Buffer Queue

Suppose the queue at node $j$ has a finite buffer size[11] $d_j \triangleq \bar{d}_j K$ for some $\bar{d}_j > 0$; here $(\bar{d}_j)_{j \in V}$ are model primitives. Throughout Section 6, the normalized state space is

$$\Omega \triangleq \left\{ \bar{\mathbf{q}} : \mathbf{1}^\top \bar{\mathbf{q}} = 1,\ \mathbf{0} \leq \bar{\mathbf{q}} \leq \bar{\mathbf{d}} \right\} , \quad \text{where } \bar{d}_j \triangleq d_j / K .$$

Similar to the case of entry control, we need to keep $\bar{\mathbf{q}}$ in the interior of $\Omega$, which is achieved by defining the normalized queue lengths $\bar{\mathbf{q}}$ as

$$\bar{q}_j \triangleq \frac{q_j + \bar{d}_j \delta_K}{\tilde{K}} \quad \text{for} \quad \delta_K = \sqrt{K} \quad \text{and} \quad \tilde{K} \triangleq K + \left( \sum_{j \in V} \bar{d}_j \right) \delta_K . \tag{21}$$

One can verify that $\bar{\mathbf{q}} \in \Omega$ for any feasible state $\mathbf{q}$. To avoid the infeasible case where the buffers are too small to accommodate all supply units, we assume that $\sum_{j \in V} \bar{d}_j > 1$. The congestion function $f(\cdot)$ is a monotone increasing function that maps each queue length to a congestion cost. Here we

[11] If the buffer size is infinite, we simply set $d_j = K$, i.e., $\bar{d}_j = 1$, since the queue length never exceeds $K$.

will state our main results for the congestion function vector

$$f_j(\bar{q}_j) = \sqrt{m} \cdot \left( (\bar{d}_j - \bar{q}_j)^{-\frac{1}{2}} - \bar{q}_j^{-\frac{1}{2}} \right) , \qquad \forall j \in V \, , \tag{22}$$

and state the results for other choices of congestion functions in Appendix D. The intuition behind (22) is to enable MBP to focus on queues which are currently either almost empty or almost full (the congestion function values for those queues take on their smallest and largest values, respectively), and use the control levers available to make the queue lengths for those queues trend strongly away from the boundary they are close to.

## 6.2 Joint Entry-Assignment Setting

We first generalize the entry control setting introduced in Section 2 by allowing the system to choose a flexible pickup and dropoff node for each demand. Formally, each node $j$ has a pick-up neighborhood $\mathcal{P}(j) \subset V, \mathcal{P}(j) \neq \emptyset$ and drop-off neighborhood $\mathcal{D}(j) \subset V, \mathcal{D}(j) \neq \emptyset$. The sets $(\mathcal{P}(j))_{j \in V}$ and $(\mathcal{D}(j))_{j \in V}$ are model primitives.

The platform control and payoff in this setting are as follows. At time $t$, after observing the demand type $(o[t], d[t]) = (j, k)$, the system makes a decision

$$(x_{ijkl}[t])_{i \in \mathcal{P}(j), l \in \mathcal{D}(k)} \in \{0,1\}^{|\mathcal{P}(j)| \cdot |\mathcal{D}(k)|} \quad \text{such that} \qquad \sum_{i \in \mathcal{P}(j), l \in \mathcal{D}(k)} x_{ijkl}[t] \leq 1 \, . \tag{23}$$

Here $x_{ijkl}[t] = 1$ stands for the platform choosing pick-up node $i \in \mathcal{P}(j)$ and drop-off node $l \in \mathcal{D}(k)$, causing a supply unit to be relocated from $i$ to $l$. The constraint in (23) captures that each demand unit is either served by one supply unit, or not served. With $x_{ijkl}[t] = 1$, the system collects payoff $v[t] = w_{ijkl}$. Without loss of generality, let

$$\max_{j,k \in V, i \in \mathcal{P}(j), k \in \mathcal{D}(k)} w_{ijkl} = 1 \, .$$

Because the queue lengths are non-negative and upper bounded by buffer sizes, we require the following constraint to be met at any $t$:

$$x_{ijkl}[t] = 0 \quad \text{if} \quad q_i[t] = 0 \text{ or } q_l[t] = d_l \, .$$

As a convention, let $x_{ij'k'l} = 0$ if $(o[t], d[t]) \neq (j', k')$. The dynamics of system state $\mathbf{q}[t]$ is as follows:

$$\mathbf{q}[t+1] = \mathbf{q}[t] + \sum_{j,k \in V^2, i \in \mathcal{P}(j), l \in \mathcal{D}(k)} (-\mathbf{e}_i + \mathbf{e}_l) x_{ijkl}[t] \, . \tag{24}$$

The definition of a feasible policy is similar to the case of entry control, hence we skip the details. We once again define the transient and steady state optimality gaps $L_T^\pi$ and $L^\pi$ as in Section 2 via

(3)-(5).

The static fluid problem (SPP) in this setting is the following:

$$\text{maximize}_{\mathbf{x}} \sum_{j,k\in V} \phi_{jk} \sum_{i\in\mathcal{P}(j),l\in\mathcal{D}(k)} w_{ijkl}\, x_{ijkl} \tag{25}$$

$$\text{s.t.} \sum_{j,k\in V} \phi_{jk} \sum_{i\in\mathcal{P}(j),l\in\mathcal{D}(k)} x_{ijkl}(\mathbf{e}_i - \mathbf{e}_l) = \mathbf{0} \qquad \text{(flow balance)}\,, \tag{26}$$

$$\sum_{i\in\mathcal{P}(j),l\in\mathcal{D}(k)} x_{ijkl} \le 1 \quad \forall j,k\in V\,,\ x_{ijkl}\ge 0 \quad \forall i,j,k,l\in V\ \text{(demand constraint)}\,. \tag{27}$$

The dual problem to the above SPP is

$$\text{minimize}_{\mathbf{y}}\, g_{\mathrm{JEA}}(\mathbf{y}), \quad \text{for } g_{\mathrm{JEA}}(\mathbf{y}) = \sum_{j,k\in V} \phi_{jk} \max_{i\in\mathcal{P}(j),l\in\mathcal{D}(k)} \left(w_{ijkl} + y_i - y_l\right)^+ \,. \tag{28}$$

MBP naturally generalizes to this setting, as described in Algorithm 2 below.

**ALGORITHM 2:** Mirror Backpressure (MBP) Policy for Joint Entry-Assignment

At the start of period $t$, the system observes $(o[t], d[t]) = (j,k)$.

$(i^*, l^*) \leftarrow \mathrm{argmax}_{i\in\mathcal{P}(j),l\in\mathcal{D}(k)} w_{ijkl} + f_i(\bar{q}_i[t]) - f_l(\bar{q}_l[t])$

**if** $w_{i^*jkl^*} + f_{i^*}(\bar{q}_{i^*}[t]) - f_{l^*}(\bar{q}_{l^*}[t]) \ge 0$ **and** $q_{i^*}[t] > 0$, $q_{l^*}[t] < d_{l^*}$ **then**

    $x_{i^*jkl^*}[t] \leftarrow 1$, i.e., serve the incoming demand using a supply unit from $i^*$ and relocate it to $l^*$ ;

**else**

    $x_{ijkl}[t] \leftarrow 0$ for all $i\in\mathcal{P}(j), l\in\mathcal{D}(k)$, i.e., drop the incoming demand;

**end**

The queue lengths update as $\bar{\mathbf{q}}[t+1] = \bar{\mathbf{q}}[t] - \frac{1}{K}\sum_{i\in\mathcal{P}(j),l\in\mathcal{D}(k)} x_{ijkl}[t](\mathbf{e}_i - \mathbf{e}_l)$.

Our performance guarantee requires the following connectivity assumption on the primitives $(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})$.

**Condition 2** (Strong Connectivity of $(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})$)**.** *Define the connectedness of triple* $(\boldsymbol{\phi},\mathcal{P},\mathcal{D})$ *as*

$$\alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D}) \triangleq \min_{S\subsetneq V, S\neq\emptyset} \sum_{j\in\mathcal{P}^{-1}(S),\ k\in\mathcal{D}^{-1}(V\backslash S)} \phi_{jk}\,. \tag{29}$$

*Here* $\mathcal{P}^{-1}(S) \triangleq \{j\in V : \mathcal{P}(j)\cap S \neq \emptyset\}$ *is the set of origins that locations* $S$ *can serve;* $\mathcal{D}^{-1}(\cdot)$ *is defined similarly. We assume that* $(\boldsymbol{\phi},\mathcal{P},\mathcal{D})$ *is strongly connected, namely,* $\alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D}) > 0$.

Note that Condition 2 is equivalent to requiring that for every ordered pair of locations $(i,l)$, there is a sequence of demand types with positive arrival rates and corresponding pick-up, drop-off nodes that would take a supply unit from $i$ eventually to $l$. If $\mathcal{P}(j) = \{j\}$, $\mathcal{D}(k) = \{k\}$, then the JEA setting reduces to entry control model in Section 2 and $\alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D}) = \alpha(\boldsymbol{\phi})$ for $\alpha(\boldsymbol{\phi})$ defined in (9).

We show the following performance guarantee, analogous to Theorem 1.

**Theorem 2.** *Consider a set of $m \in \mathbb{N}$ locations, scaled buffer sizes $\bar{\mathbf{d}} = (\bar{d}_j)_{j\in V}$ with $\bar{d}_j > 0$ and $\sum_{j\in V} \bar{d}_j > 1$, and any $(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})$ that satisfies Condition 2. Then there exists $\rho = \rho(\bar{\mathbf{d}}) < \infty$, $K_1 = \rho \cdot \mathrm{poly}\left(m, \frac{1}{\alpha(\boldsymbol{\phi},\mathcal{P},\mathcal{D})}\right)$, $M_1 = O(m)$ and $M_2 = \rho \cdot O(m^2)$, such that the following holds.*[12] *For the congestion functions $(f_j(\cdot))_{j\in V}$ defined in (22), for any $K \geq K_1$, the following guarantee holds for Algorithm 2*

$$L_T^{\mathrm{MBP}} \leq M_1 \cdot \frac{K}{T} + M_2 \cdot \frac{1}{K}, \qquad \text{and} \qquad L^{\mathrm{MBP}} \leq M_2 \cdot \frac{1}{K}.$$

In Appendix D we generalize Theorem 2 by showing that a similar performance guarantee holds for a large class of congestion functions.

### 6.3 Joint Pricing-Assignment Setting

In this section, we consider the joint pricing-assignment (JPA) setting and design the corresponding MBP policy. The platform's control problem is to set a price for each demand origin-destination pair, and decide an assignment plan at each period to maximize payoff. The proposed algorithm is a generalization of backpressure based joint-rate-scheduling control policies (see, e.g., Lin and Shroff 2004, Eryilmaz and Srikant 2006). For readability, here we only consider flexible pick-up locations, i.e., $\mathcal{D}(j) = \{j\}$, although the result can be easily generalized to allow flexible drop-off locations.

The platform control and payoff in this setting are as follows. At time $t$, after observing the demand type $(o[t], d[t]) = (j, k)$, the system chooses a *price* $p_{jk}[t] \in [p_{jk}^{\min}, p_{jk}^{\max}]$ and an assignment decision

$$(x_{ijk}[t])_{i\in\mathcal{P}(j)} \in \{0,1\}^{|\mathcal{P}(j)|} \quad \text{such that} \quad \sum_{i\in\mathcal{P}(j)} x_{ijk}[t] \leq 1. \tag{30}$$

As before we require

$$x_{ijk}[t] = 0 \quad \text{if} \quad q_i[t] = 0 \text{ or } q_k[t] = d_k.$$

The result of the platform control is as follows:

1. Upon seeing the price, the arriving demand unit will abandon with probability $F_{jk}(p_{jk}[t])$, where $F_{jk}(\cdot)$ is the cumulative distribution function of type $(j,k)$ demand's willingness-to-pay.

2. If the demand stays (i.e., buys) and $x_{ijk} = 1$, the system dynamics follows (24). Meanwhile, the platform collects payoff $v[t] = p_{jk}[t] - c_{ijk}$ where $c_{ijk}$ is the "cost" of serving a demand unit of type $(j,k)$ using pickup location $i$.

[12] Here the poly$(\cdot)$ term in $K_1$ does not depend on $\bar{\mathbf{d}}$, and "poly" indicates a polynomial. Also, $M_1$ and $M_2$ do not depend on $\alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})$ and $\bar{\mathbf{d}}$.

3. If the demand unit abandons, the supply units do not move and $v[t] = 0$.

We assume the following regularity conditions to hold for demand functions $(F_{jk}(p_{jk}))_{j,k}$. These assumptions are quite standard in the revenue management literature, (see, e.g., Gallego and Van Ryzin 1994).

**Condition 3.** *1. Assume*[13] $F_{jk}(p_{jk}^{\min}) = 0$ *and that* $F_{jk}(p_{jk}^{\max}) = 1$.

*2. Each demand type's willingness-to-buy is non-atomic with support* $[p_{jk}^{\min}, p_{jk}^{\max}]$ *and positive density everywhere on the support; hence* $F_{jk}(p_{jk})$ *is differentiable and strictly increasing on* $(p_{jk}^{\min}, p_{jk}^{\max})$. *(If the support is a subinterval of* $[p_{jk}^{\min}, p_{jk}^{\max}]$*, we can simply redefine* $p_{jk}^{\min}$ *and* $p_{jk}^{\max}$ *to be the boundaries of this subinterval.)*

*3. The revenue functions* $r_{jk}(\mu_{jk}) \triangleq \mu_{jk} \cdot p_{jk}(\mu_{jk})$ *are concave and twice continuously differentiable.*

As a consequence of Condition 3 parts 1 and 2, the willingness to pay distribution $F_{jk}(\cdot)$ has an inverse denoted as $p_{jk}(\mu_{jk}) : [0,1] \to [p_{jk}^{\min}, p_{jk}^{\max}]$ which gives the price leading to any *net demand fraction* $\mu_{jk} \in [0,1]$. In other words, Condition 3 ensures that the platform has sufficient flexibility in choosing prices $p_{jk} \in [p_{jk}^{\min}, p_{jk}^{\max}]$ to achieve any desired fraction of net (realized) demand. Without loss of generality, let $\max_{j,k\in V} p_{jk}^{\max} + \max_{i,j,k\in V} |c_{ijk}| = 1$.

In the JPA setting, the net demand $\phi_{jk}\mu_{jk}$ plays a role in myopic revenues but also affects the distribution of supply, and the chosen prices need to balance myopic revenues with maintaining a good spatial distribution of supply. Intuitively, when sufficiently flexible pricing is available as a control lever, the system will modulate the quantity of demand *purely* through changing the prices (and serving all the demand which is then realized) rather than apply entry control (i.e., dropping some demand proactively).

The dual problem to the SPP in the JPA setting is[14]

$$\begin{aligned} &\text{minimize}_{\mathbf{y}}\ g_{\text{JPA}}(\mathbf{y}), \\ &\text{for } g_{\text{JPA}}(\mathbf{y}) = \sum_{j,k\in V} \phi_{jk} \max_{\{0\le\mu_{jk}\le 1\}} \left( r_{jk}\left(\mu_{jk}\right) + \mu_{jk} \max_{i\in\mathcal{P}(j)} \left(-c_{ijk} + y_i - y_k\right) \right). \end{aligned} \tag{31}$$

Parallel to the derivation of Algorithm 1, we design the MBP policy by making it execute stochastic mirror descent on $g_{\text{JPA}}(\mathbf{y})$: The mean fraction of demand realized under the policy will be the outer argmax in the definition (31) of $g_{\text{JPA}}(\mathbf{y})$, and the assignment decision achieves the inner argmax in the definition (31) of $g_{\text{JPA}}(\mathbf{y})$.

[13] The assumption $F_{jk}(p_{jk}^{\min}) = 0$ is without loss of generality, since if a fraction of demand is unwilling to pay $p_{jk}^{\min}$, that demand can be excluded from $\boldsymbol{\phi}$ itself.

[14] The derivation of the dual objective is in Appendix C.

A key feature of MBP policy (see Algorithm 3 below) is that it *does not require any prior knowledge of gross demand* $\boldsymbol{\phi}$, in contrast to the fluid-based policies as in Banerjee et al. (2016) which need to know $\boldsymbol{\phi}$ exactly. This is an attractive feature in many applications. Here we assume that the willingness-to-pay distributions $F_{jk}(\cdot)$s are exactly known to the platform; it may be possible to relax this assumption via a modified policy which "learns" the $F_{jk}(\cdot)$s, however, this is beyond the scope of this paper.

**ALGORITHM 3:** Mirror Backpressure (MBP) Policy for Joint Pricing-Assignment

At the start of period $t$, the system observes $(o[t], d[t]) = (j, k)$.

$i^* \leftarrow \arg\max_{i \in \mathcal{P}(j)} \left\{ -c_{ijk} + f_i(\bar{q}_i[t]) - f_k(\bar{q}_k[t]) \right\}$;

**if** $q_{i^*}[t] > 0$, $q_k[t] < d_k$ **then**

  $\mu_{jk}[t] \leftarrow \text{argmax}_{\mu_{jk} \in [0,1]} \left\{ r_{jk}(\mu_{jk}) + \mu_{jk} \cdot (-c_{i^*jk} + f_i(\bar{q}_{i^*}[t]) - f_k(\bar{q}_k[t])) \right\}$;

  $p_{jk}[t] \leftarrow F^{-1}(\mu_{jk}[t])$;

  $x_{i^*jk}[t] \leftarrow 1$, i.e., serve the incoming demand by pick up from $i^*$;

**else**

  $x_{ijk}[t] \leftarrow 0$ for all $i \in \mathcal{P}(j)$, i.e., drop the incoming demand;

**end**

The queue lengths update as $\bar{\mathbf{q}}[t+1] = \bar{\mathbf{q}}[t] - \frac{1}{K} \sum_{i \in \mathcal{P}(j)} x_{ijk}[t](\mathbf{e}_i - \mathbf{e}_k)$.

Condition 3 ensures that Algorithm 3 has two key desirable properties:

1. The computed prices satisfy $p_{jk}[t] \in [p_{jk}^{\min}, p_{jk}^{\max}]$ (by the observation which immediately follows Condition 3).
2. The optimization problem for computing $\mu_{jk}[t]$ is a concave maximization problem (Condition 3 part 3), hence $\mu_{jk}[t]$ can be efficiently computed.

We have the following performance guarantee for Algorithm 3, analogous to Theorem 1.

**Theorem 3.** *Consider a set of* $m \in \mathbb{N}$ *locations,* $\bar{\mathbf{d}} = (\bar{d}_j)_{j \in V}$ *with* $\bar{d}_j > 0$ *and* $\sum_{j \in V} \bar{d}_j > 1$, *minimum and maximum allowed prices* $(p_{jk}^{\min}, p_{jk}^{\max})_{j \in V, k \in V}$, *any* $(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})$ *(where* $\mathcal{D}$ *is identity) that satisfy Condition 2, and willingness-to-pay distributions* $(F_{jk})_{j \in V, k \in V}$ *that satisfy Condition 3. Then there exist* $K_1 < \infty$, $M_1 < \infty$ *and* $M_2 < \infty$ *such that for the congestion functions* $(f_j(\cdot))$*s defined in* (22)*, for any* $K \geq K_1$*, the following guarantee holds for Algorithm 3*

$$L_T^{\text{MBP}} \leq M_1 \cdot \frac{K}{T} + M_2 \cdot \frac{1}{K}, \qquad \text{and} \qquad L^{\text{MBP}} \leq M_2 \cdot \frac{1}{K}.$$

See Appendix D for a general version of Theorem 3, where a similar performance guarantee is shown for a family of mirror maps, e.g., for logarithmic and linear mirror maps. The proof shows that, once again, $M_1$ grows linearly with $m$ and $M_2$ is quadratic in $m$ and both these constants do not depend on $\alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})$, and $K_1$ is polynomial in both $m$ and $\frac{1}{\alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})}$.

# 7 Application to Scrip Systems

In this section, we illustrate the application of our model to scrip systems (see Application 3 in Section 1.1, where real world examples of scrip systems such as kidney exchanges and babysitting co-ops are also discussed). We consider a service exchange with multiple agents and heterogeneous services, where agents exchange scrips (i.e., artificial currency) for services. There is a central planner who tries to maximize social welfare by making decisions over whether a trade should occur when a service request arises, and if so, who should be the service provider. The setting is closely related to the joint entry-assignment (JEA) setting introduced in Section 6. Using the methodology we developed earlier, we introduce a simple MBP control rule that asymptotically maximizes social welfare.

## 7.1 Model of Scrip Systems

We now describe a model of a scrip system. As will become clear below, scrip systems (with a central planner) can be modeled in a manner very similar to the JEA setting introduced in Section 6. To emphasize the connection, we point out the corresponding elements of the JEA model in parentheses.

**Service exchange.** We study an economy with a finite number of agents indexed by $i \in V_D$. There are finitely many types of services indexed by $j' \in V_S$. (Here $V_D \cup V_S$ are the *nodes* in JEA. Note that $j' \in V_S$ will serve as virtual "destinations" but the drop-off neighborhood of $j'$ will be a subset of $V_D$ and in particular, will not include $j'$ itself. As a result, there will be no scrips at nodes in $V_S$ at any time.) Consider a slotted time model, where each period exact one service request arises, and with probability $\phi_{ij'}$ it is of type $(i, j')$ (*demand types* in JEA), i.e., it comes from agent $i$ and requests type $j'$ service. Each agent has a skill set, i.e., the service types he[15] can provide. The skill set structure is modeled by the *skill compatibility graph*, which is an undirected bipartite graph $\mathcal{G} = (V_D \cup V_S, E)$ (see Figure 3 for an illustration). The neighborhood of $i \in V_D$ in $\mathcal{G}$ is his skill set, which is denoted by $\mathcal{N}(i) \subseteq V_S$. The neighborhood of $j' \in V_S$ in $\mathcal{G}$ consists of the providers of type $j'$ service, which is denoted by $\mathcal{N}(j') \subseteq V_D$. The payoff from serving a request from agent $i$ for type $j'$ service is $w_{ij'}$. (The planner can choose a suitable value of $w_{ij'}$ based on the estimated utility that agent $i$ has for service $j'$, the cost incurred by the service provider, and other relevant considerations.) Assume that an agent cannot fulfill his own service request. (For request type $(i, j')$, $\{i\}$ is the *pickup neighborhood*, $\mathcal{N}(j') \backslash \{i\}$ is the *dropoff neighborhood*[16].)

[15] For expository simplicity, we refer to an agent as "he" and the central planner as "she".

[16] This slightly generalizes the JEA model as the drop-off neighborhood depends on demand type, but it poses no

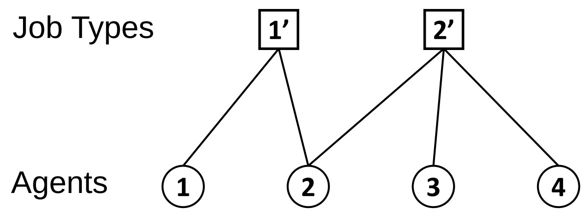


Figure 3: An example of skill compatibility graph in a service exchange with two job types and four agents.

**Scrips.** Denote the number of scrips (*queue lengths* in JEA) each agent has at time $t$ as $\mathbf{q}[t] = [q_1[t], \cdots, q_{|V_D|}[t]]$. There are a fixed number (denoted by $K$) of scrips in the system, hence $\mathbf{q}[t] \in \Omega_K$ where $\Omega_K$ is defined in Section 2.

**Central planner.** We consider a central planner who tries to maximize social welfare, which is defined as the per period average of collected payoff, over a finite/infinite horizon. The central planner's control levers include *acceptance rule* and *provider selection rule* (*entry control* and *assignment* in JEA): each time a type $(i, j')$ request arrives, she needs to decide whether to fulfill it, and if so, who should be the service provider. As is typical in scrip systems, an agent is required to have at least one scrip to request services (*no-underflow constraint* in JEA).

*Comparison with the model in Johnson et al. (2014).* The work of Johnson et al. (2014) consider the case where there is only one type of service which all agents can provide (i.e., $\mathcal{G}$ is a star graph), and $\phi_{ij'}$ is equal for all agents $i$. One one hand, we significantly generalize their model by considering heterogeneous services, asymmetric service request arrivals, and general skill compatibility graphs. They propose an optimal rule for the symmetric setting, whereas we develop an asymptotically optimal control rule for the general setting. On the other hand, we only focus on the central planner setting, and leave the incentives of agents for future work (see the remarks in Section 7.2).

### 7.2 MBP Control Rule and Asymptotic Optimality

Leveraging the similarity between the model introduced above and the JEA setting introduced earlier, we are easily able to define a simple control rule based on MBP that achieves asymptotic optimality in Algorithm 4 below. Here the congestion function $f(\cdot)$ can be chosen as in (6) since buffers are effectively infinite (alternate congestion functions such as a logarithm/linear function

technical difficulty.

may be used instead; see Remark 1). Denote the normalized number of scrips (defined in (7)) in the possession of agent $i$ by $\bar{q}_i$.

---

**ALGORITHM 4:** MBP-based Acceptance and Provider Selection Rule for Scrip Systems

---

At the start of period $t$, the central planner receives a request from agent $i$ for service $j'$.

**if** $w_{ij'} + f(\bar{q}_i[t]) - \min_{k\in\mathcal{N}(j')\backslash\{i\}} f(\bar{q}_k[t]) \geq 0$ **and** $\bar{q}_i[t] > 0$ **then**

  $k^* \leftarrow \text{argmin}_{k\in\mathcal{N}(j')\backslash\{i\}} f(\bar{q}_k[t])$,

  Let agent $k^*$ provide the service to $i$, and agent $i$ gives one scrip to agent $k^*$ ;

**else**

  Reject the service request from agent $i$;

**end**

---

A performance guarantee similar to Theorem 2 holds for Algorithm 4 under the following connectivity condition of $(\phi, \mathcal{G})$.

**Condition 4** (Strong Connectivity of $(\boldsymbol{\phi}, \mathcal{G})$)**.** *Define the connectedness of $(\boldsymbol{\phi}$,$\mathcal{G})$ as*

$$\alpha(\boldsymbol{\phi}, \mathcal{G}) \triangleq \min_{S\subsetneq V_D, S\neq\emptyset} \sum_{i\in S} \sum_{j'\in\mathcal{N}(V_D\backslash S)} \phi_{ij'} \,.$$

*We assume that $(\boldsymbol{\phi}, \mathcal{G})$ is strongly connected, namely, $\alpha(\boldsymbol{\phi}, \mathcal{G}) > 0$.*

We have the following performance guarantee.

**Proposition 2.** *Consider a set of $m$ agents and any $(\boldsymbol{\phi}, \mathcal{G})$ that satisfies Condition 4. Then there exists $K_1 = \text{poly}\left(m, \frac{1}{\alpha(\boldsymbol{\phi},\mathcal{G})}\right)$, $M_1 = O(m)$ and $M_2 = O(m^2)$, such that the following holds. For the congestion function $f(\cdot)$ defined in (6), for any $K \geq K_1$, the following guarantee holds for Algorithm 4*

$$L_T^{\text{MBP}} \leq M_1 \cdot \frac{K}{T} + M_2 \cdot \frac{1}{K}\,, \qquad \text{and} \qquad L^{\text{MBP}} \leq M_2 \cdot \frac{1}{K}\,.$$

The proof of Proposition 2 is very similar to that of Theorem 2; see Appendix D.

A few remarks on the model and results are in order:

1. *Necessity of declining trades.* By considering a more general setting than in Johnson et al. (2014), we obtain qualitatively different insights on the optimal control rule by central planner. In Johnson et al. (2014), it is optimal for the central planner to always approve trades, and let the agent with fewest scrips be the service provider. In our general setting, however, in many cases the central planner has to decline a non-trivial fraction of the trades to sustain flow balance of scrips in the system (constraint (11)).[17] When a trade is approved, our policy also chooses the compatible trade partner with the fewest scrips as service provider.

[17] For example, consider a setting with two agents $i_1$ and $i_2$, who each specialize in a different service type $\mathcal{N}(i_1) = \{j_1'\}$ and $\mathcal{N}(i_2) = \{j_2'\}$. Under the mild condition $\phi_{i_1j_2'} \neq \phi_{i_2j_1'}$, the planner will be forced to decline a positive fraction of requests.

2. *Incentives.* Our analysis of scrip systems is meant to illustrate the versatility of MBP control policies, hence we only focused on the central planner setting. It would be interesting to study the MBP control rule in the decentralized setting where the agents recommended to be potential trading partners can decide whether to trade, but that is beyond the scope of the current paper. (At a high level, we expect that agents will have an incentive to provide service whenever requested by the MBP policy as long as (i) agents are patient, and (ii) agents benefit from trading, i.e., derive more value from receiving service than the cost they incur from providing service.)

# 8 Numerical Experiments

In this section, we simulate the MBP policy in a realistic ridehailing environment using yellow cab data from NYC Taxi & Limousine Commission and Google Maps. We allow the platform two control levers: entry control and assignment/dispatch control, similar to the JEA setting in Section 6.2. Our theoretical model made the simplifying assumption that pickup and service of demand are *instantaneous*. We relax this assumption in our numerical experiments by adding realistic travel times. We consider the following two cases:

1. *Excess supply.* The number of cars in the system is slightly (5%) above the "fluid requirement" (see Section 8.1 for details on the "fluid requirement") to achieve the benchmark $W^{\mathrm{SPP}}$ (the value of the SPP (25)-(27)) in steady state. (Note that even with transit times it is still impossible to beat $W^{\mathrm{SPP}}$ in steady state since (26) and (27) must hold for the time average of $\mathbf{x}[t]$.)
2. *Scarce supply.* The number of cars fall short (by 25%) of the "fluid requirement", i.e., there aren't enough cars to realize the optimal solution of SPP defined in (25)-(27), hence $W^{\mathrm{SPP}}$ cannot be achieved.

**Summary of findings.** We make a natural modification of the MBP policy to account for finite travel times; specifically, we employ a *supply-aware MBP* policy which estimates and uses a shadow price of keeping a vehicle (supply unit) occupied for one unit of time. We find that in both the excess supply and the scarce supply cases, the MBP policy, which is given no information about the demand arrival rates, significantly outperforms static (fluid-based) policy, even when the latter is provided with prior knowledge of exact demand arrival rates. The MBP policy also vastly outperforms the greedy non-idling policy, which demonstrates the practical importance and value of proactively dropping demand.

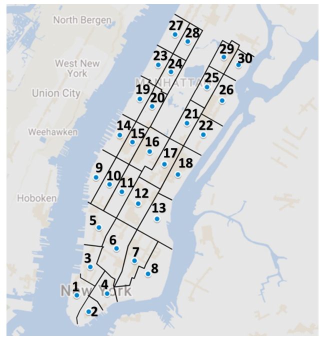


Figure 4: A 30 location model of Manhattan below 110-th street, excluding the Central Park. (Source: tessellation is based on Buchholz (2015), the figure is generated using Google Maps.)

## 8.1 Simulation Setup and Benchmark Policies

Throughout the numerical experiments, we use the following model primitives.

- *Payoff structure.* In many scenarios, ride-hailing platforms take a commission proportional to the trip fare, which increases with trip distance/duration. Motivated by this, we present results for $w_{ijk}$ set to be the travel time from $j$ to[18] $k$.
- *Graph topology.* We consider a 30-location model of Manhattan below 110-th street excluding Central Park (see Figure 4), as defined in Buchholz (2015). We let pairs of regions which share a non-trivial boundary be pickup compatible with each other, e.g., regions 23 and 24 are compatible but regions 23 and 20 are not.
- *Demand arrival process, and pickup/service times.* We consider a stationary demand arrival process, whose rate is the average decensored demand from 8 a.m. to 12 p.m. estimated in Buchholz (2015) (see Appendix E for a full description). This period includes the morning rush hour and has significant imbalance of demand flow across geographical locations (for many customers the destination is in Midtown Manhattan).[19] We estimate travel times between location pairs using Google Maps.[20]
- *Number of cars, and steady state upper bound.*
  — *Excess supply.* We use as a baseline the fluid requirement $K_{\text{fl}}$ on number of cars needed to achieve optimal payoff. A simple workload conservation argument (using Little's Law) gives

[18] We tested a variety of payoff structures, and found that our results are robust to the choice of $\mathbf{w}$. One set of tests was to generate 100 random payoff vectors $\mathbf{w}$, with each $w_{ijk}$ drawn i.i.d. from Uniform(0,1); we found that the results obtained are similar.

[19] We also simulated the MBP and greedy policy with time-varying demand arrival rates, where the demand arrival rate is estimated (from the real data) for every 5 min interval. Our MBP policy still significantly outperforms the greedy policy.

[20] The average travel time across all demand is 13.1 minutes, and the average pickup time is about 4 minutes (it is policy dependent).

the fluid requirement as follows. Applying Little's Law, if the optimal solution $\mathbf{x}^*$ of SPP (25)-(27) is realized as the average long run assignment, the mean number of cars who are currently occupied picking up customers is $\sum_{j,k\in V}\sum_{i\in\mathcal{P}(j)} d_{ijk}\phi_{jk}x^*_{ijk}$, for $d_{ijk} \triangleq \tilde{D}_{ij} + D_{jk}$, where $\tilde{D}_{ij}$ is the pickup time from $i$ to $j$ and $D_{jk}$ is the travel time from $j$ to $k$. In our case, it turns out that $K_{\text{fl}} = 7,307$. We use $1.05 \times K_{\text{fl}}$ as the total number of cars in the system to study the excess supply case, i.e., there are 5% extra (idle) cars in the system beyond the number needed to achieve the $W^{\text{SPP}}$ benchmark.

— *Scarce supply.* When the number of cars in the system is fewer than the fluid requirement, i.e., $K = \kappa K_{\text{fl}}$ for $\kappa < 1$, no policy can achieve a steady state performance of $W^{\text{SPP}}$. A tighter upper bound on the steady state performance is then the value of the SPP (25)-(27) with the additional supply constraint

$$\sum_{j,k\in V}\sum_{i\in\mathcal{P}(j)} d_{ijk}\phi_{jk}x_{ijk} \le K\,. \tag{32}$$

We denote the value of this problem for $K = \kappa K_{\text{fl}}$ by $W^{\text{SPP}}(\kappa)$. We study the case $\kappa = 0.75$ as an example of scarce supply. For our simulation environment, it turns out that $W^{\text{SPP}}(0.75) \approx 0.86W^{\text{SPP}}$, i.e., $0.86W^{\text{SPP}}$ is an upper bound on the per period payoff achievable in steady state.

We compare the performance of our MBP-based policy against the following two policies:

1. *Static (fluid-based) policy.* The fluid-based policy is a static randomization based on the solution to the SPP, given exactly correct demand arrival rates (see, e.g., Banerjee et al. 2016, Ozkan and Ward 2016). See Appendix E for details.
2. *Greedy non-idling policy.* For each demand type $(j,k)$, the greedy policy dispatches from supply location $i$ that has the highest payoff $w_{ijk}$ among all compatible neighbors of $j'$ which have at least one supply unit available. If there are ties (as is the case if the payoff $w_{ijk}$ does not depend on $i$), the policy prefers a supply location with shorter pickup time.

## 8.2 The Supply-Aware MBP Policy

We propose and study the following heuristic policy inspired by MBP, that additionally incorporates the supply constraint. We call it *supply-aware MBP*. Given a demand arrival with origin $j$ and destination $k$, the policy makes its decision as per:

$$i^* \leftarrow \arg\max_{i\in\mathcal{P}(j)} \left\{ w_{ijk} + f(\bar{q}_i[t]) - f(\bar{q}_k[t]) - v[t]d_{ijk} \right\}$$

**If** $w_{i^*jk} + f(\bar{q}_{i^*}[t]) - f(\bar{q}_{l^*}[t]) - v[t]d_{ijk} \ge 0$ **and** $q_{i^*}[t] > 0$, dispatch from $i^*$, **else** Drop, where $v[t]$ is the current estimate of the shadow price for a "tightened" version of supply constraint (32). We define the tightened supply constraint as

$$\sum_{j,k\in V}\sum_{i\in\mathcal{P}(j)} d_{ijk}\phi_{jk}x_{ijk} \le 0.95K\,, \tag{33}$$

where the coefficient of $K$ is the flexible "utilization" parameter, that we have set 0.95, meaning that we are aiming to keep 5% vehicles free on average, systemwide.[21] We update $v[t]$ as

$$v[t+1] = \left[v[t] + \frac{1}{K}\left(\sum_{j,k\in V}\sum_{i\in\mathcal{P}(j)} d_{ijk}a_{jk}[t]x_{ijk}[t] - 0.95K\right)\right]^{+}\,,$$

where $a_{jk}[t] = 1$ if and only if $(o[t], d[t]) = (j, k)$, otherwise $a_{jk}[t] = 0$. An iteration of supply-aware MBP is equivalent to executing a (dual) stochastic mirror descent step on the supply-aware SPP with objective (25) and constraints (26)(27)(33).

We use the congestion function $f(q_i) = c\log q_i$ in our numerical simulations, with $c = \max_{i,j,k\in V} w_{ijk}$.

### 8.3 The Excess Supply Case

We simulate the (stationary) system from 8 a.m. to 12 p.m. with 100 randomly generated initial states (see Appendix E for details on the initial state generation). The simulation results on performance are shown in Figure 5. The result confirms that the MBP policy significantly outperforms both the static policy and the greedy policy: the average payoff under MBP over 4 hours is about 103% of $W^{\mathrm{SPP}}$ (here $W^{\mathrm{SPP}}$ is again an upper bound on the steady state performance[22]), while the static policy and greedy policy only achieve 63% and 68% of $W^{\mathrm{SPP}}$, respectively. The performance of the static policy converges very slowly to $W^{\mathrm{SPP}}$, leading to poor transient performance.[23] The performance of the greedy policy quickly deteriorates over time because it ignores the flow balance constraints and creates huge geographical imbalances in supply availability.

### 8.4 The Scarce Supply Case

In the scarce supply case, e.g., $K = 0.75K_{\mathrm{fl}}$, no policy can achieve a stationary performance of $W^{\mathrm{SPP}}$; rather we have an steady state upper bound of $W^{\mathrm{SPP}}(0.75) \approx 0.86W^{\mathrm{SPP}}$. We use this as our benchmark.

[21] Keeping a small fraction of vehicles free is helpful in managing the stochasticity in the system. Note that the present paper does not study how to systematically choose the utilization parameter.

[22] $W^{\mathrm{SPP}}$ is still an upper bound on stationary performance when pickup and service times are included in our model. However, in this case a transient upper bound is difficult to derive. As a result, we use the ratio of average per period payoff to $W^{\mathrm{SPP}}$ as a performance measure, with the understanding that it may exceed 1 at early times.

[23] For example, after running for 20 hours, and the average payoff generated by static policy in the 20-th hour is $0.96W^{\mathrm{SPP}}$.

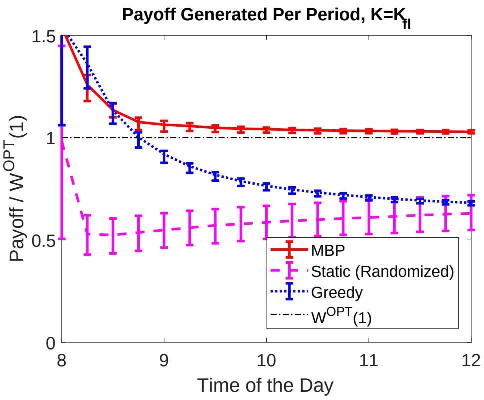


Figure 5: Payoff collected per period under the MBP policy, static fluid-based policy and greedy policy, relative to $W^{\mathrm{SPP}}$. We run 100 trials with random initial queue lengths; the error bars represent the performance between 95% and 5% quantiles, and the main line is the median.

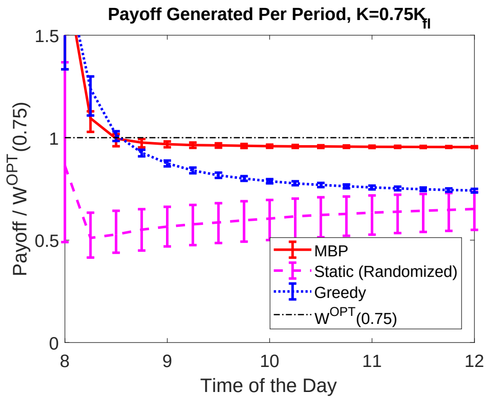


Figure 6: Payoff collected per period under the modified MBP policy, static fluid-based policy and greedy policy, relative to $W^{\mathrm{SPP}}(0.75)$, the value of SPP along with constraint (32) for $K = 0.75K_{\mathrm{fl}}$. We run 100 trials with random initial queue lengths; the error bars represent the performance between 95% and 5% quantiles, and the main line is the median.

Figure 6 shows that the MBP policy also vastly outperforms the static policy and greedy policy in the scarce supply case. MBP generates average per period payoff that is 95% of the benchmark $W^{\mathrm{SPP}}(0.75)$ over 4 hours, while the static policy and greedy policy only achieves 65% and 74% resp. of the benchmark over the same period. Reassuringly, the mean value of $v(t)$ in our simulations of supply-aware MBP is within 10% of the optimal dual variable to the tightened supply constraint (33) in the SPP along with (33) (both values are close to 0.50). Again, we observe that the average performance of static policy improves (slowly) as the time horizon gets longer, while the performance of greedy deteriorates.

# 9 Discussion

In this paper we consider the payoff maximizing dynamic control of a closed network of resources. We propose a novel family of policies called Mirror Backpressure (MBP), which generalize the celebrated backpressure policy such that it executes mirror descent with the desired mirror map. The MBP policy overcomes the challenge stemming from the no-underflow constraint and supply externalities as the policy accounts for the geometry of the problem, and it does not require any knowledge of demand arrival rates. We prove that the MBP policy is able to achieve good transient performance, losing at most an $O\left(\frac{K}{T} + \frac{1}{K}\right)$ fraction of the achievable payoff, where $K$ is the number of supply units and $T$ is the number of arrivals over the horizon of interest. Realistic numerical experiments in a ridehailing setting corroborate our theoretical findings.

Before closing, we point out several interesting directions for future research, many of which we are actively pursuing.

1. *Empty car repositioning.* Our model allows for the payoff to depend on dispatch location. As such, we may interpret our model in terms of empty car repositioning, especially if travel times can be incorporated.
2. *Time-varying demand arrivals.* Since our policy does not rely upon any statistical knowledge of the demand arrival rates, it is promising for the situation where demand arrival rate is time-varying.
3. *Improved performance via "centering" MBP based on demand arrival rates.* If the optimal shadow prices $\mathbf{y}^*$ are known (or learned by learning $\boldsymbol{\phi}$ via observing demand), we can modify the congestion function to $\tilde{f}_j(\bar{q}_j) = y_j^* + f(\bar{q}_j)$. For the resulting "centered" MBP policy, based on the result of Huang and Neely (2009) and the convergence of mirror descent, we are optimistic that the steady regret will be exponentially small in $K$ as $K$ increases.
4. *Other applications of MBP.* MBP appears to be a powerful and general approach to obtain near optimal performance despite no-underflow constraints in the control of queueing networks. It does not necessitate a heavy traffic assumption, and provides guarantees on both transient and steady state performance. For example, the matching queues problem studied by Gurvich and Ward (2014) is hard due to no-underflow constraints and to handle them that paper makes stringent assumptions on the network structure. MBP may be able to achieve provably near optimal performance for more general matching queue systems.

## References


Adan, Ivo, Gideon Weiss. 2012. A loss system with skill-based servers under assign to longest idle server policy. *Probability in the Engineering and Informational Sciences* **26**(3) 307–321.

Agarwal, Nikhil, Itai Ashlagi, Eduardo Azevedo, Clayton R Featherstone, Ömer Karaduman. 2019. Market failure in kidney exchange. *American Economic Review* **109**(11) 4026–70.

Balseiro, Santiago, David B Brown, Chen Chen. 2019. Dynamic pricing of relocating resources in large networks. *Available at SSRN 3313737* .

Banerjee, Siddhartha, Daniel Freund, Thodoris Lykouris. 2016. Pricing and optimization in shared vehicle systems: An approximation framework. *CoRR abs/1608.06819* .

Banerjee, Siddhartha, Yash Kanoria, Pengyu Qian. 2018. State dependent control of closed queueing networks with application to ride-hailing .

Beck, Amir, Marc Teboulle. 2003. Mirror descent and nonlinear projected subgradient methods for convex optimization. *Operations Research Letters* **31**(3) 167–175.

Braverman, Anton, Jim G Dai, Xin Liu, Lei Ying. 2019. Empty-car routing in ridesharing systems. *Operations Research* **67**(5) 1437–1452.

Buchholz, Nicholas. 2015. Spatial equilibrium, search frictions and efficient regulation in the taxi industry. Tech. rep., Technical report, University of Texas at Austin.

Bumpensanti, Pornpawee, He Wang. 2018. A re-solving heuristic for dynamic resource allocation with uniformly bounded revenue loss. *arXiv preprint arXiv:1802.06192* .

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

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

Chung, Hangil, Daniel Freund, David B Shmoys. 2018. Bike angels: An analysis of citi bike's incentive program. *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*. ACM, 5.

Dai, Jim G, Wuqin Lin. 2005. Maximum pressure policies in stochastic processing networks. *Operations Research* **53**(2) 197–218.

Dai, Jim G, Wuqin Lin. 2008. Asymptotic optimality of maximum pressure policies in stochastic processing networks. *The Annals of Applied Probability* **18**(6) 2239–2299.

Désir, Antoine, Vineet Goyal, Yehua Wei, Jiawei Zhang. 2016. Sparse process flexibility designs: is the long chain really optimal? *Operations Research* **64**(2) 416–431.

Eryilmaz, Atilla, R Srikant. 2006. Joint congestion control, routing, and mac for stability and fairness in wireless networks. *IEEE Journal on Selected Areas in Communications* **24**(8) 1514–1524.

Eryilmaz, Atilla, R Srikant. 2007. Fair resource allocation in wireless networks using queue-length-based scheduling and congestion control. *IEEE/ACM Transactions on Networking (TON)* **15**(6) 1333–1344.

Eryilmaz, Atilla, R Srikant. 2012. Asymptotically tight steady-state queue length bounds implied by drift conditions. *Queueing Systems* **72**(3-4) 311–359.

Gallego, Guillermo, Garrett Van Ryzin. 1994. Optimal dynamic pricing of inventories with stochastic demand over finite horizons. *Management science* **40**(8) 999–1020.

Georgiadis, Leonidas, Michael J Neely, Leandros Tassiulas, et al. 2006. Resource allocation and cross-layer control in wireless networks. *Foundations and Trends® in Networking* **1**(1) 1–144.

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

Harrison, J Michael. 2000. Brownian models of open processing networks: Canonical representation of workload. *Annals of Applied Probability* 75–103.

Harrison, J Michael. 2003. A broader view of brownian networks. *The Annals of Applied Probability* **13**(3) 1119–1150.

Huang, Longbo, Michael J Neely. 2009. Delay reduction via lagrange multipliers in stochastic network optimization. *Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, 2009. WiOPT 2009. 7th International Symposium on*. IEEE, 1–10.

Huang, Longbo, Michael J Neely. 2011. Utility optimal scheduling in processing networks. *Performance Evaluation* **68**(11) 1002–1021.

Jiang, Libin, Jean Walrand. 2009. Stable and utility-maximizing scheduling for stochastic processing networks. *2009 47th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 1111–1119.

Johnson, Kris, David Simchi-Levi, Peng Sun. 2014. Analyzing scrip systems. *Operations Research* **62**(3) 524–534.

Jordan, William C, Stephen C Graves. 1995. Principles on the benefits of manufacturing process flexibility. *Management Science* **41**(4) 577–594.

Lin, Xiaojun, Ness B Shroff. 2004. Joint rate control and scheduling in multihop wireless networks. *2004 43rd IEEE Conference on Decision and Control (CDC)(IEEE Cat. No. 04CH37601)*, vol. 2. IEEE, 1484–1489.

Maguluri, Siva Theja, R Srikant. 2016. Heavy traffic queue length behavior in a switch under the maxweight algorithm. *Stochastic Systems* **6**(1) 211–250.

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

Neely, Michael J. 2010. Stochastic network optimization with application to communication and queueing systems. *Synthesis Lectures on Communication Networks* **3**(1) 1–211.

Nemirovsky, Arkadii Semenovich, David Borisovich Yudin. 1983. Problem complexity and method efficiency in optimization. .

Ozkan, Erhun, Amy R Ward. 2016. Dynamic matching for real-time ridesharing .

Shi, Cong, Yehua Wei, Yuan Zhong. 2019. Process flexibility for multiperiod production systems. *Operations Research* **67**(5) 1300–1320.

Srikant, Rayadurgam. 2012. *The mathematics of Internet congestion control*. Springer Science & Business Media.

Stolyar, Alexander L. 2003. Control of end-to-end delay tails in a multiclass network: Lwdf discipline optimality. *Annals of Applied Probability* 1151–1206.

Stolyar, Alexander L. 2004. Maxweight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic. *The Annals of Applied Probability* **14**(1) 1–53.

Stolyar, Alexander L. 2005. Maximizing queueing network utility subject to stability: Greedy primal-dual algorithm. *Queueing Systems* **50**(4) 401–457.

Sweeney, Joan, Richard James Sweeney. 1977. Monetary theory and the great capitol hill baby sitting co-op crisis: comment. *Journal of Money, Credit and Banking* **9**(1) 86–89.

Talluri, Kalyan T, Garrett J Van Ryzin. 2006. *The theory and practice of revenue management*, vol. 68. Springer Science & Business Media.

Tassiulas, Leandros. 1992. Dynamic link activation scheduling in multihop radio networks with fixed or changing connectivity .

Tassiulas, Leandros, Anthony Ephremides. 1992. Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE transactions on automatic control* **37**(12) 1936–1948.

# A Overview of Proofs

In this paper, we proved performance guarantees for three settings: entry control (Theorem 1), joint entry-assignment control (Theorem 2) and joint pricing-assignment control (Theorem 3). For most parts of the proof, the proof of the entry control case can be easily extended to the other two cases. For particular lemmas/propositions, the proofs of the JEA and JPA settings are more involved. In the following, we rigorously prove all the intermediate results leading to Theorem 1, and skip some of the proofs of the JEA and JPA setting if the extension is immediate. For easier reading, we put analogous results together.

# B Finite Horizon Payoff Upper Bound: Proof of Proposition 1

In this section, we prove the finite horizon payoff upper bounds for entry control (Proposition 1), JEA (Proposition 3) and JPA setting (Proposition 4).

## B.1 Entry Control Setting

Recall Proposition 1.

**Proposition 1.** *For any horizon $T < \infty$ and any $K$, the finite and infinite horizon average payoff $W_T^*$ and $W^*$ are upper bounded as*

$$W_T^* \le W^{\mathrm{SPP}} + m \cdot \frac{K}{T}, \qquad W^* \le W^{\mathrm{SPP}}.$$

The idea behind Proposition 1 is as follows. As is typical in such settings, $W^{\mathrm{SPP}}$ is an upper bound to the long run expected payoff. We obtain a *finite horizon* upper bound in addition by slightly relaxing the flow constraint (11) in the SPP to

$$\left| \mathbf{1}_S^\top \left( \sum_{j,k\in V} \phi_{jk} \cdot x_{jk}(\mathbf{e}_j - \mathbf{e}_k) \right) \right| \leq \frac{K}{T} \qquad \forall\, S \subseteq V\,, \tag{34}$$

where $\mathbf{1}_S$ is the vector with 1s at nodes in $S$ and 0s at all other nodes.

We establish two key lemmas to facilitate the proof of Proposition 1. The first lemma (Lemma 4) shows that the expected payoff cannot exceed the value of the *finite horizon SPP* defined by (10), (34) and (12).

**Lemma 4.** *For any horizon $T < \infty$, any $K$ and any starting state $\mathbf{q}[0]$, the expected payoff generated by any feasible entry control policy $\pi$ is upper bounded by the value of the finite horizon SPP defined by* (10)*,* (34) *and* (12)*.*

*Proof.* Let $\pi$ be any feasible policy. For each pair $j, k \in V$, define

$$\bar{x}_{jk}[T] \triangleq \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[x_{jk}[t] | (o[t], d[t]) = (j,k)]\,.$$

In words, $\bar{\mathbf{x}}(T)$ is the average over $1 \leq t \leq T$ of the entry control decisions. We decompose the time-average of payoff collected in the first $T$ periods as:

$$\begin{aligned} W_T^\pi &= \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left[ \sum_{j,k\in V} \mathbb{1}\{(o[t], d[t]) = (j,k)\} \cdot w_{jk} \cdot x_{jk}[t] \right] \\ &= \frac{1}{T} \sum_{t=0}^{T-1} \sum_{j,k\in V} \phi_{jk} \cdot \mathbb{E}\left[ w_{jk} \cdot x_{jk}[t] \,\middle|\, (o[t], d[t]) = (j,k) \right] \\ &= \sum_{j,k\in V} \phi_{jk} \cdot w_{jk} \cdot \bar{x}_{jk}[T]\,, \end{aligned}$$

where we only used the basic properties of conditional expectation. Similarly, for the time-average of the change of queue length we have:

$$\frac{1}{T} \cdot \mathbb{E}[\mathbf{q}[T] - \mathbf{q}[0]] = \sum_{j,k\in V} \phi_{jk} \cdot \bar{x}_{jk}[T] \cdot (\mathbf{e}_j - \mathbf{e}_k)\,.$$

Because there are only $K$ resources circulating in the system, the net outflow from any subset of nodes $S \subseteq V$ should not exceed $K$. Since each individual $\mathbf{x}[t]$ satisfies (12), it must hold for $\bar{\mathbf{x}}[T]$. Optimize over $\bar{\mathbf{x}}[T]$, we have that $W_T^\pi$ is upper bounded by the optimization problem defined by (10), (34) and (12) regardless of the initial configuration $\mathbf{q}[0]$. This concludes the proof. ☐

In order to facilitate the second key lemma, we first prove a supporting lemma (Lemma 5). We call $\mathbf{x}$ a *directed acyclic assignment* if there is no sequence of node pairs $\mathcal{C} = ((j_1, j_2), (j_2, j_3), \ldots, (j_s, j_{s+1}))$ where $j_{s+1} \triangleq j_1$, $j_r \in V$ for $r = 1, 2, \ldots, s$, such that

$$\phi_{j_r j_{r+1}} \cdot x_{j_r j_{r+1}} > 0 \qquad \forall\, r = 1, 2, \ldots, s\,, \tag{35}$$

In words, there is no cycle $\mathcal{C}$ such that there is a positive flow along $\mathcal{C}$.

**Lemma 5.** *Any feasible solution* $\mathbf{x}^F$ *of the finite horizon SPP satisfying* (34) *and* (12) *can be decomposed as*

$$\mathbf{x}^{\mathrm{F}} = \mathbf{x}^{\mathrm{S}} + \mathbf{x}^{\mathrm{DAG}}\,, \tag{36}$$

*where* $\mathbf{x}^{\mathrm{S}}$ *is a feasible solution for the SPP satisfying* (11) *and* (12)*, and* $\mathbf{x}^{\mathrm{DAG}}$ *is a directed acyclic assignment satisfying* (34) *and* (12)*.*

*Proof.* The existence of such a decomposition can be established using a standard network flow argument: Start with $\mathbf{x}^{\mathrm{S}} = \mathbf{0}$ and $\mathbf{x}^{\mathrm{DAG}} = \mathbf{x}^{\mathrm{F}}$. Then, iteratively, if $\mathbf{x}^{\mathrm{DAG}}$ includes a cycle $\mathcal{C}$ with a positive flow along $\mathcal{C}$ as above, move a flow of $u(\mathcal{C}) \triangleq \min_{1\le r\le s} \phi_{j_r j_{r+1}} \cdot x_{j_r j_{r+1}}$ along $\mathcal{C}$ from $\mathbf{x}^{\mathrm{DAG}}$ to $\mathbf{x}^{\mathrm{S}}$, via the updates

$$x^{\mathrm{S}}_{j_r j_{r+1}} \leftarrow x^{\mathrm{S}}_{j_r j_{r+1}} + \frac{u(\mathcal{C})}{\phi_{j_r j_{r+1}}}\,, \qquad\qquad x^{\mathrm{DAG}}_{j_r j_{r+1}} \leftarrow x^{\mathrm{DAG}}_{j_r j_{r+1}} - \frac{u(\mathcal{C})}{\phi_{j_r j_{r+1}}}\,,$$

for all $r = 1, 2, \ldots, s$. This iterative process maintains the following invariants which hold at the end of each iteration:

- $\mathbf{x}^{\mathrm{S}}$ remains feasible for the SPP, in particular, it satisfies flow balance (11).
- $\mathbf{x}^{\mathrm{F}} = \mathbf{x}^{\mathrm{S}} + \mathbf{x}^{\mathrm{DAG}}$ remains true.
- It remains true that

  $$\sum_{j,k\in V} \phi_{jk} \cdot x^{\mathrm{DAG}}_{jk}(\mathbf{e}_j - \mathbf{e}_k) = \sum_{j,k\in V} \phi_{jk} \cdot x^{\mathrm{F}}_{jk}(\mathbf{e}_j - \mathbf{e}_k)\,.$$

  i.e., $\mathbf{x}^{\mathrm{DAG}}$ has the same net inflow/outflow from each supply node as $\mathbf{x}^{\mathrm{F}}$. In particular, $\mathbf{x}^{\mathrm{DAG}}$ satisfies approximate flow balance (34).

Moreover, the iterative process progresses monotonically: Observe that $\mathbf{x}^{\mathrm{S}}$ (weakly) increases monotonically, whereas $\mathbf{x}^{\mathrm{DAG}}$ (weakly) decreases monotonically (but preserves $\mathbf{x}^{\mathrm{DAG}} \ge \mathbf{0}$). Since we also know that $\mathbf{x}^{\mathrm{S}}$ is bounded, it follows that this iterative process converges. Moreover, when it converges, it must be that there is no remaining cycle with positive flow in $\mathbf{x}^{\mathrm{DAG}}$, else it contradicts the fact that the process has converged. Hence, $\mathbf{x}^{\mathrm{S}}, \mathbf{x}^{\mathrm{DAG}}$ at the end of the process provide the claimed decomposition. □

Using this supporting lemma, we now establish the second key lemma which shows that the value of the finite horizon SPP cannot be much larger than the value of the SPP.

**Lemma 6.** *Let* $W_T^{\mathrm{SPP}}$ *be the value of the finite horizon SPP. This value is upper bounded in terms of the value* $W^{\mathrm{SPP}}$ *of the SPP as*

$$W_T^{\mathrm{SPP}} \le W^{\mathrm{SPP}} + m \cdot \frac{K}{T}\,.$$

*Proof.* We appeal to the decomposition from Lemma 5 to decompose any feasible solution $\mathbf{x}^{\mathrm{F}}$ to the finite horizon fluid problem as

$$\mathbf{x}^{\mathrm{F}} = \mathbf{x}^{\mathrm{S}} + \mathbf{x}^{\mathrm{DAG}}\,,$$

where $\mathbf{x}^{\mathrm{S}}$ is feasible for the SPP and $\mathbf{x}^{\mathrm{DAG}}$ is a directed acyclic flow that is feasible for the finite horizon SPP, i.e., satisfying (34) and (12). Hence, the objective (10) of the finite horizon fluid problem can be written as the sum of two terms

$$\sum_{j,k \in V} \phi_{jk} \cdot w_{jk} \cdot x_{jk}^{\mathrm{F}} = \sum_{j,k \in V} \phi_{jk} \cdot w_{jk} \cdot (x_{jk}^{\mathrm{S}} + x_{jk}^{\mathrm{DAG}}) \,, \tag{37}$$

and each of the terms can be bounded from above. By definition of $W^{\mathrm{SPP}}$ we know that

$$\sum_{j,k \in V} \phi_{jk} \cdot w_{jk} \cdot x_{jk}^{\mathrm{S}} \leq W^{\mathrm{SPP}} \,.$$

We will now show that

$$\sum_{j,k \in V} \phi_{jk} \cdot w_{jk} \cdot x_{jk}^{\mathrm{DAG}} \leq m \cdot \frac{K}{T} \,.$$

The lemma will follow, since this will imply an upper bound of $W^{\mathrm{SPP}} + m \cdot \frac{K}{T}$ on the objective for any $\mathbf{x}^{\mathrm{F}}$ satisfying (34) and (12).

Consider $\mathbf{x}^{\mathrm{DAG}}$. Since it is a directed acyclic assignment, there is an ordering $(j_1, j_2, \ldots, j_m)$ of the nodes in $V$ such that all assignments move supply from an earlier nodes to a later node in this ordering. More precisely, it holds that

$$x_{j_l, j_r}^{\mathrm{DAG}} = 0 \qquad \forall\, l > r \,. \tag{38}$$

Now consider the subsets $A_\ell \triangleq \{j_1, j_2, \ldots, j_\ell\} \subset V$ for $\ell = 1, 2, \ldots, m-1$. Note that from (38), $x^{\mathrm{DAG}}$ does not move any supply from $V \backslash A_\ell$ to $A_\ell$. Hence we have

$$\begin{aligned} \mathbf{1}_{A_\ell}^\top \left( \sum_{j,k \in V} \phi_{jk} \cdot w_{jk} \cdot x_{jk}^{\mathrm{DAG}} (\mathbf{e}_j - \mathbf{e}_k) \right) &= \sum_{j \in A_\ell, k \in V \backslash A_\ell} \phi_{jk} \cdot x_{jk}^{\mathrm{DAG}} \\ &\leq \frac{K}{T} \qquad \forall\, l = 1, 2, \ldots, m-1 \,, \end{aligned} \tag{39}$$

where we made use of (34) to obtain the upper bound. Further, note that for each $x_{j_l, j_r}^{\mathrm{DAG}}$ with $l < r$, the term $\phi_{j_l j_r} \cdot x_{j_l, j_r}^{\mathrm{DAG}}$ is part of the above sum for $\ell = l$. Because $\max_{j,k \in V} |w_{jk}| \leq 1$, it follows that

$$\begin{aligned} \sum_{j,k \in V} \phi_{jk} \cdot w_{jk} \cdot x_{jk}^{\mathrm{DAG}} &\leq \sum_{j,k \in V} \phi_{jk} \cdot x_{jk}^{\mathrm{DAG}} \\ &\leq \sum_{1 \leq \ell < m} \sum_{j \in A_\ell, k \in V \backslash A_\ell} \phi_{jk} \cdot x_{jk}^{\mathrm{DAG}} \\ &\leq m \cdot \frac{K}{T} \,. \end{aligned}$$

This completes the proof. □

*Proof of Proposition 1.* The proposition follows immediately from Lemmas 4 and 6. □

### B.2 Joint Entry-Assignment Setting

**Proposition 3.** *For any horizon $T < \infty$, any $K$ and any starting state $\mathbf{q}[0]$, the finite and infinite horizon average payoff $W_T^*$ and $W^*$ in the JEA setting are upper bounded as*

$$W_T^* \leq W^{\text{SPP}} + m \cdot \frac{K}{T}\,, \qquad W^* \leq W^{\text{SPP}}\,.$$

*Here $W^{\text{SPP}}$ is the optimal value of SPP* (25)-(27).

*Proof Sketch of Proposition 3.* The only twist in this proof comparing to Proposition 1 is the definition of *directed acyclic assignment.* Here we call $\mathbf{x}$ a directed acyclic assignment if there is no sequence of node pairs

$$\mathcal{C} = ((j_1, j_2), (j_2, j_3), \cdots, (j_s, j_{s+1}))$$

where $j_{s+1} \triangleq j_1$, $j_r \in V$ for $r = 1, 2, \cdots, s$, such that for any $r = 1, \cdots, s$,

$$\exists j, k \in V\,, \text{ s.t. } j_r \in \mathcal{P}(j), j_{r+1} \in \mathcal{D}(k) \text{ and } \phi_{jk} \cdot x_{j_r, j, k, j_{r+1}} > 0\,.$$

The rest of the proof proceeds exactly the same as in Proposition 1. □

### B.3 Joint-Pricing-Assignment Setting

The static fluid problem (SPP) in the JPA setting is

$$\text{maximize}_{\mathbf{x}} \sum_{j,k \in V} \phi_{jk} \left( r_{jk} \left( \sum_{i \in \mathcal{P}(j)} x_{ijk} \right) - \sum_{i \in \mathcal{P}(j)} c_{ijk} \cdot x_{ijk} \right) \tag{40}$$

$$\text{s.t.} \sum_{j,k \in V} \phi_{jk} \sum_{i \in \mathcal{P}(j)} x_{ijk}(\mathbf{e}_i - \mathbf{e}_k) = \mathbf{0} \qquad \text{(flow balance)} \tag{41}$$

$$\sum_{i \in \mathcal{P}(j)} x_{ijk} \leq 1 \quad \forall j, k \in V\,,\ x_{ijk} \geq 0 \quad \forall i, j, k \in V \qquad \text{(demand constraint)}\,. \tag{42}$$

**Proposition 4.** *For any horizon $T < \infty$, any $K$ and any starting state $\mathbf{q}[0]$, the finite and infinite horizon average payoff $W_T^*$ and $W^*$ in the JPA setting are upper bounded as*

$$W_T^* \leq W^{\text{SPP}} + m \cdot \frac{K}{T}\,, \qquad W^* \leq W^{\text{SPP}}\,.$$

*Here $W^{\text{SPP}}$ is the optimal value of SPP* (40)-(42).

The main twist of the proof comparing to Proposition 1 is that the objective function in (40) is no longer linear. We first prove a JPA version of Lemma 4.

**Lemma 7.** *For any horizon $T < \infty$, any $K$ and any starting state $\mathbf{q}[0]$, the expected payoff generated by any JPA policy $\pi$ is upper bounded by the value of the finite horizon SPP:*

$$\text{maximize}_{\mathbf{x}} \sum_{j,k \in V} \phi_{jk} \cdot \left( r_{jk} \left( \sum_{i \in \mathcal{P}(j)} x_{ijk} \right) - \sum_{i \in \mathcal{P}(j)} c_{ijk} \cdot x_{ijk} \right)$$

$$\text{s.t. } \mathbb{1}_S^\top \left( \sum_{j,k \in V} \phi_{jk} \sum_{i \in \mathcal{P}(j)} x_{ijk} (\mathbf{e}_i - \mathbf{e}_k) \right) \le \frac{K}{T} \quad \forall S \subseteq V$$

$$\sum_{i \in \mathcal{P}(j)} x_{ijk} \le 1 \quad \forall j,k \in V\,, \;\; x_{ijk} \ge 0 \quad \forall i,j,k \in V\,.$$

*Proof.* Let $\pi$ be any feasible JPA policy. For each pair $j,k \in V$ and $i \in \mathcal{P}(j)$, define

$$\bar{x}_{ijk}[T] \triangleq \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}[\bar{F}_{jk}(p_{jk}[t]) \cdot x_{ijk}[t] | (o[t], d[t]) = (j,k)]\,.$$

In words, $\bar{x}_{ijk}[T]$ is the average rate over the first $T$ periods of picking up type $(j,k)$ demands from node $i$.

Let $U_{jk}[t]$ be the willingness-to-pay of a type $(j,k)$ demand arriving at time $t$. We decompose the time-average of payoff collected in the first $T$ periods as:

$$\begin{aligned} W_T^\pi =& \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E} \left[ \sum_{j,k \in V} \mathbb{1}\{(o[t], d[t]) = (j,k), U_{jk}[t] \ge p_{jk}[t]\} \sum_{i \in \mathcal{P}(j)} (p_{jk}[t] - c_{ijk}) \cdot x_{ijk}[t] \right] \\ =& \frac{1}{T} \sum_{t=0}^{T-1} \sum_{j,k \in V} \phi_{jk} \cdot \mathbb{E} \left[ \mathbb{1}\{U_{jk}[t] \ge p_{jk}[t]\} \sum_{i \in \mathcal{P}(j)} (p_{jk}[t] - c_{ijk}) \cdot x_{ijk}[t] \,\big|(o[t], d[t]) = (j,k) \right]. \end{aligned}$$

Because $U_{jk}[t]$ is independent of $p_{jk}[t]$ and $x_{ijk}[t]$, we have

$$W_T^\pi = \frac{1}{T} \sum_{t=0}^{T-1} \sum_{j,k \in V} \phi_{jk} \cdot \mathbb{E} \left[ \bar{F}_{jk}(p_{jk}[t]) \sum_{i \in \mathcal{P}(j)} (p_{jk}[t] - c_{ijk}) \cdot x_{ijk}[t] \,\big|(o[t], d[t]) = (j,k) \right].$$

Let $\mu_{jk}[t] \triangleq \bar{F}_{jk}(p_{jk}[t])$, and let $\hat{x}_{ijk}[t] \triangleq \mu_{jk}[t] \cdot x_{ijk}[t]$, we have

$$\begin{aligned} W_T^\pi =& \frac{1}{T} \sum_{t=0}^{T-1} \sum_{j,k \in V} \phi_{jk} \cdot \mathbb{E} \left[ \left( \sum_{i \in \mathcal{P}(j)} \hat{x}_{ijk}[t] \right) \cdot \bar{F}_{jk}^{-1}(\mu_{jk}[t]) - \sum_{i \in \mathcal{P}(j)} c_{ijk} \cdot \hat{x}_{ijk}[t] \,\big|(o[t], d[t]) = (j,k) \right] \\ \le& \frac{1}{T} \sum_{t=0}^{T-1} \sum_{j,k \in V} \phi_{jk} \cdot \mathbb{E} \left[ \left( \sum_{i \in \mathcal{P}(j)} \hat{x}_{ijk}[t] \right) \cdot \bar{F}_{jk}^{-1} \left( \sum_{i \in \mathcal{P}(j)} \hat{x}_{ijk}[t] \right) - \sum_{i \in \mathcal{P}(j)} c_{ijk} \cdot \hat{x}_{ijk}[t] \,\big|(o[t], d[t]) = (j,k) \right] \\ =& \frac{1}{T} \sum_{t=0}^{T-1} \sum_{j,k \in V} \phi_{jk} \cdot \mathbb{E} \left[ r_{jk} \left( \sum_{i \in \mathcal{P}(j)} \hat{x}_{ijk}[t] \right) - \sum_{i \in \mathcal{P}(j)} c_{ijk} \cdot \hat{x}_{ijk}[t] \,\big|(o[t], d[t]) = (j,k) \right]. \end{aligned}$$

Here the first inequality follows from the fact that $\bar{F}_{jk}^{-1}(\cdot)$ is non-increasing, the last equality uses the definition of revenue function $r_{jk}(\cdot)$. Linearity of conditional expectation and conditional Jensen's inequality yields:

$$W_T^\pi \le \sum_{j,k \in V} \phi_{jk} \cdot \left( - \sum_{i \in \mathcal{P}(j)} c_{ijk} \cdot \bar{x}_{ijk}[T] + \frac{1}{T} \sum_{t=0}^{T-1} r_{jk} \left( \mathbb{E} \left[ \sum_{i \in \mathcal{P}(j)} \hat{x}_{ijk}[t] \,\big|(o[t], d[t]) = (j,k) \right] \right) \right).$$

Use Jensen's inequality again, we have

$$W_T^\pi \leq \sum_{j,k\in V} \phi_{jk} \cdot \left( - \sum_{i\in\mathcal{P}(j)} c_{ijk} \cdot \bar{x}_{ijk}[T] + r_{jk}\left( \sum_{i\in\mathcal{P}(j)} \bar{x}_{ijk}[T] \right) \right).$$

For the time-average of the change of queue length we have:

$$\frac{1}{T} \cdot \mathbb{E}[\mathbf{q}[T] - \mathbf{q}[0]] = \sum_{j,k\in V} \phi_{jk} \sum_{i\in\mathcal{P}(j)} \bar{x}_{ijk}(T) \cdot (\mathbf{e}_i - \mathbf{e}_k).$$

Because there are only $K$ resources in the system, the net outflow from any subset of nodes should not exceed $K$. Note that $\bar{\mathbf{x}}$ must satisfy constraint (42). Optimize over $\bar{\mathbf{x}}(T)$, we obtain the desired result. □

*Proof Sketch of Proposition 4.* The rest of the proof proceeds almost exactly the same as in Proposition 1. The only caveat is that the equation (37) should be replaced by inequality

$$\begin{aligned}
&\sum_{j,k\in V} \phi_{jk} \left( - \sum_{i\in\mathcal{P}(j)} c_{ijk} \cdot x_{ijk}^{\mathrm{F}} + r_{ij}\left( \sum_{i\in\mathcal{P}(j)} x_{ijk}^{\mathrm{F}} \right) \right) \\
\leq &\sum_{j,k\in V} \phi_{jk} \left( - \sum_{i\in\mathcal{P}(j)} c_{ijk} \cdot x_{ijk}^{\mathrm{S}} + r_{ij}\left( \sum_{i\in\mathcal{P}(j)} x_{ijk}^{\mathrm{S}} \right) \right) \\
&+ \sum_{j,k\in V} \phi_{jk} \left( - \sum_{i\in\mathcal{P}(j)} c_{ijk} \cdot x_{ijk}^{\mathrm{DAG}} + r_{ij}\left( \sum_{i\in\mathcal{P}(j)} x_{ijk}^{\mathrm{DAG}} \right) \right).
\end{aligned}$$

Here the inequality follows from the subadditivity of non-negative concave functions. □

# C Lyapunov Analysis: Proof of Lemmas 1, 2

## C.1 Decomposition of Optimality Gap

### C.1.1 Entry Control Setting: Proof of Lemma 1

Recall Lemma 1.

**Lemma 1.**

$$\begin{aligned}
W^* - \mathbb{E}[v^{\mathrm{MBP}}[t]|\bar{\mathbf{q}}[t]] \leq &\underbrace{\tilde{K}\left( F(\bar{\mathbf{q}}[t]) - \mathbb{E}[F(\bar{\mathbf{q}}[t+1])|\bar{\mathbf{q}}[t]] \right)}_{\mathcal{T}_1} + \underbrace{\frac{1}{2\tilde{K}} \cdot \max_{j\in V} \left| f'(\bar{q}_j[t]) \right|}_{\mathcal{T}_2} \\
&+ \underbrace{\left( W^{\mathrm{SPP}} - g(\mathbf{f}(\bar{\mathbf{q}}[t])) \right)}_{\mathcal{T}_3} + \underbrace{\mathbb{1}\left\{ q_j[t] = 0, \exists j \in V \right\}}_{\mathcal{T}_4}.
\end{aligned}$$

*Proof.* For congestion function $f(\bar{q}_j)$ that is strictly increasing and continuous for each $j$, we consider the Lyapunov function $F(\bar{\mathbf{q}})$ which is the antiderivative of $\mathbf{f}(\bar{\mathbf{q}})$. The Bregman divergence associated with $\mathbf{f}(\bar{\mathbf{q}})$ is defined as:

$$D_F(\bar{\mathbf{q}}_1, \bar{\mathbf{q}}_2) = F(\bar{\mathbf{q}}_1) - F(\bar{\mathbf{q}}_2) - \langle \mathbf{f}(\bar{\mathbf{q}}_1), \bar{\mathbf{q}}_1 - \bar{\mathbf{q}}_2 \rangle \,. \tag{43}$$

Plug $\bar{\mathbf{q}}_1 = \bar{\mathbf{q}}[t+1]$, $\bar{\mathbf{q}}_2 = \bar{\mathbf{q}}[t]$ in (43) and rearrange the terms, we have:

$$F(\bar{\mathbf{q}}[t+1]) - F(\bar{\mathbf{q}}[t]) = \langle \mathbf{f}(\bar{\mathbf{q}}[t]), \bar{\mathbf{q}}[t+1] - \bar{\mathbf{q}}[t] \rangle + D_F(\bar{\mathbf{q}}[t+1], \bar{\mathbf{q}}[t]) \,.$$

Subtracting $\frac{1}{\tilde{K}} \sum_{j,k \in V} w_{jk} \cdot \phi_{jk} \cdot x_{jk}[t]$ on both sides and take conditional expectation given $\bar{\mathbf{q}}[t]$, we have:

$$\begin{aligned} &\mathbb{E}[F(\bar{\mathbf{q}}[t+1])|\bar{\mathbf{q}}[t] - F(\bar{\mathbf{q}}[t]) - \frac{1}{\tilde{K}} \sum_{j,k\in V} w_{jk} \cdot \phi_{jk} \cdot \mathbb{E}[x_{jk}[t]|\bar{\mathbf{q}}[t] \\ =\ &\underbrace{-\frac{1}{\tilde{K}} \sum_{j,k\in V} w_{jk} \cdot \phi_{jk} \cdot \mathbb{E}[x_{jk}[t]|\bar{\mathbf{q}}[t]] + \langle \mathbf{f}(\bar{\mathbf{q}}[t]), \mathbb{E}[\bar{\mathbf{q}}[t+1]|\bar{\mathbf{q}}[t]] - \bar{\mathbf{q}}[t] \rangle}_{\text{(I)}} + \underbrace{\mathbb{E}[D_F(\bar{\mathbf{q}}[t+1], \bar{\mathbf{q}}[t])|\bar{\mathbf{q}}[t]}_{\text{(II)}} \,. \end{aligned} \tag{44}$$

Let $x_{jk}^{\text{NOM}}[t]$ be the “nominal” control that ignores the no-underflow constraint, i.e.

$$(x_{jk}^{\text{NOM}})[t] = \begin{cases} 1 & \text{if } w_{jk} + f(\bar{q}_j[t]) - f(\bar{q}_k[t]) \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{45}$$

It immediately follows that

$$(x_{jk}^{\text{MBP}})[t] = (x_{jk}^{\text{NOM}})[t] \cdot \mathbb{1}\{q_j[t] > 0\} \,. \tag{46}$$

With a slight abuse of notation, denote $\mathbf{x}^{\text{NOM}}$ as $\tilde{\mathbf{x}}$, $\mathbf{x}^{\text{MBP}}$ as $\mathbf{x}$. Rearrange the terms in (I) and plug in (46), we have

$$\begin{aligned} \text{(I)} =\ & -\frac{1}{\tilde{K}} \sum_{j,k\in V} \phi_{jk} \cdot \big(w_{jk} + f(\bar{q}_j[t]) - f(\bar{q}_k[t])\big) \cdot \mathbb{E}[x_{jk}[t]|\bar{\mathbf{q}}[t]] \\ =\ & -\frac{1}{\tilde{K}} \sum_{j,k\in V} \phi_{jk} \cdot \big(w_{jk} + f(\bar{q}_j[t]) - f(\bar{q}_k[t])\big) \cdot \mathbb{E}[\tilde{x}_{jk}[t]|\bar{\mathbf{q}}[t]] \\ & + \frac{1}{\tilde{K}} \sum_{j,k\in V} \phi_{jk} \cdot \big(w_{jk} + f(\bar{q}_j[t]) - f(\bar{q}_k[t])\big) \cdot \mathbb{1}\left\{q_j[t] = 0\right\} \,. \end{aligned}$$

By definition of MBP policy, we have:

$$\begin{aligned} & -\frac{1}{\tilde{K}} \sum_{j,k\in V} \phi_{jk} \cdot \big(w_{jk} + f(\bar{q}_j[t]) - f(\bar{q}_k[t])\big) \cdot \mathbb{E}[\tilde{x}_{jk}[t]|\bar{\mathbf{q}}[t]] \\ =\ & -\frac{1}{\tilde{K}} \sum_{j,k\in V} \phi_{jk} \cdot \big(w_{jk} + f(\bar{q}_j[t]) - f(\bar{q}_k[t])\big)^+ \\ =\ & -\frac{1}{\tilde{K}} \cdot g(\mathbf{f}(\bar{\mathbf{q}}[t])) \,. \end{aligned}$$

Using the fact that $f(\cdot)$ is increasing and that $\max_{j,k\in V}|w_{jk}|= 1$, we have

$$\frac{1}{\tilde{K}} \sum_{j,k\in V} \phi_{jk} \cdot \big(w_{jk} + f(\bar{q}_j[t]) - f(\bar{q}_k[t])\big) \cdot \mathbb{1}\left\{q_j[t] = 0\right\} \leq \frac{1}{\tilde{K}} \cdot \mathbb{1}\left\{q_j[t] = 0, \exists j\right\} \,.$$

Combining the above inequality and equality yields

$$\text{(I)} \leq -\frac{1}{\tilde{K}} \cdot g(\mathbf{f}(\bar{\mathbf{q}}[t])) + \frac{1}{\tilde{K}} \cdot \mathbb{1}\left\{q_j[t] = 0, \exists j\right\} .$$

Now we proceed to bound (II). By definition of Bregman divergence, (II) is the second order remainder of the Taylor series of $F(\cdot)$. Using the fact that $f(\cdot)$ is increasing, we have[24]

$$\text{(II)} \leq \frac{1}{2}\sum_{j\in V}\mathbb{E}\left[f'(\bar{q}_j[t])(\bar{q}_j[t]-\bar{q}_j[t+1])^2|\bar{\mathbf{q}}[t]\right] \leq \frac{1}{2\tilde{K}^2} \cdot \max_{j\in V} f'\left(\bar{q}_j[t]\right) .$$

Plug the above inequalities of (I) and (II) in (44), we have

$$\begin{aligned} &\mathbb{E}[F(\bar{\mathbf{q}}[t+1])|\bar{\mathbf{q}}[t]] - F(\bar{\mathbf{q}}[t]) - \frac{1}{\tilde{K}}\mathbb{E}[v^{\text{MBP}}[t]|\bar{\mathbf{q}}[t]] \\ \leq &-\frac{1}{\tilde{K}} \cdot g(\mathbf{f}(\bar{\mathbf{q}}[t])) + \frac{1}{2\tilde{K}^2} \cdot \max_{j\in V} f'(\bar{q}_j[t]) + \frac{1}{\tilde{K}} \cdot \mathbb{1}\left\{q_j[t] = 0, \exists j\right\} . \end{aligned}$$

Rearranging the terms yields:

$$\begin{aligned} -\mathbb{E}[v^{\text{MBP}}[t]|\bar{\mathbf{q}}[t]] \leq \tilde{K}\left(F(\bar{\mathbf{q}}[t]) - \mathbb{E}[F(\bar{\mathbf{q}}[t+1])|\bar{\mathbf{q}}[t]]\right) + \frac{1}{2\tilde{K}} \cdot \max_{j\in V} f'(\bar{q}_j[t]) \\ - g(\mathbf{f}(\bar{\mathbf{q}}[t])) + \mathbb{1}\left\{q_j[t] = 0, \exists j\right\} . \end{aligned}$$

Using Proposition 1, we have $W^* \leq W^{\text{SPP}}$. Adding these two inequalities together concludes the proof. □

#### C.1.2 Joint Entry-Assignment Setting

For JEA setting, we have the following lemma which is analogous to Lemma 1.

**Lemma 8.** *Consider congestion functions $f_j(\cdot)$s that are strictly increasing and continuously differentiable. We have the following decomposition:*

$$\begin{aligned} W^* - \mathbb{E}[v^{\text{MBP}}[t]|\bar{\mathbf{q}}[t]] \leq \tilde{K}\left(F(\bar{\mathbf{q}}[t]) - \mathbb{E}[F(\bar{\mathbf{q}}[t+1])|\bar{\mathbf{q}}[t]]\right) + \frac{1}{2\tilde{K}} \cdot \max_{j\in V} f'(\bar{q}_j[t]) \\ + \left(W^{\text{SPP}} - g_{\text{JEA}}(\mathbf{f}(\bar{\mathbf{q}}[t]))\right) + \mathbb{1}\left\{q_j[t] = 0 \text{ or } d_j, \exists j\right\} , \end{aligned} \tag{47}$$

*where $g_{\text{JEA}}(\mathbf{y})$ is defined in* (28)*.*

*Proof Sketch.* The proof is a direct extension of the proof of Lemma 1. The only difference is that the actual control deviates from the nominal control not only when $q_j[t] = 0$ for some $j$, but also when $q_j[t] = d_j$ for some $j$. □

#### C.1.3 Joint Pricing-Assignment Setting

For JPA setting, we have the following lemma which is analogous to Lemma 1.

**Lemma 9.** *Consider congestion functions $f_j(\cdot)$s that are strictly increasing and continuously differentiable. We have the following decomposition:*

$$W^* - \mathbb{E}[v^{\text{MBP}}[t]|\bar{\mathbf{q}}[t]] \leq \tilde{K}\left(F(\bar{\mathbf{q}}[t]) - \mathbb{E}[F(\bar{\mathbf{q}}[t+1])|\bar{\mathbf{q}}[t]]\right) + \frac{1}{2\tilde{K}} \cdot \max_{j\in V} f'(\bar{q}_j[t]) \tag{48}$$

[24] For exposition simplicity, we ignore the difference between $f'(\bar{q}_j[t])$ and $f'(\bar{q}_j[t+1])$ in the Taylor expansion.

$$+ \left(W^{\mathrm{SPP}} - g_{\mathrm{JEA}}(\mathbf{f}(\bar{\mathbf{q}}[t]))\right) + \mathbb{1}\left\{q_j[t] = 0 \text{ or } d_j, \exists j\right\},$$

*where* $g_{\mathrm{JPA}}(\mathbf{y})$ *is defined in* (31).

*Proof Sketch.* The proof is a direct extension of the proof of Lemma 1. To use the strong duality argument, we prove below that $g_{\mathrm{JPA}}(\cdot)$ is indeed the partial dual function of the SPP (40)-(42). Then because the primal problem is a concave optimization problem with linear constraint, strong duality must hold.

Let $\mathbf{y}$ be the Lagrange multipliers corresponding to constraints (41). We have

$$\begin{aligned}
g_{\mathrm{JPA}}(\mathbf{y}) = & \max_{\sum_{i\in\mathcal{P}(j)} x_{ijk}\le 1, x_{ijk}\ge 0} \sum_{j,k\in V} \phi_{jk}\left(r_{jk}\left(\sum_{i\in\mathcal{P}(j)} x_{ijk}\right) + \sum_{i\in\mathcal{P}(j)}\left(-c_{ijk} + y_i - y_k\right)x_{ijk}\right) \\
= & \sum_{j,k\in V} \phi_{jk} \max_{\sum_{i\in\mathcal{P}(j)} x_{ijk}\le 1, x_{ijk}\ge 0}\left(r_{jk}\left(\sum_{i\in\mathcal{P}(j)} x_{ijk}\right) + \sum_{i\in\mathcal{P}(j)}\left(-c_{ijk} + y_i - y_k\right)x_{ijk}\right) \\
= & \sum_{j,k\in V} \phi_{jk} \max_{0\le\mu_{jk}\le 1} \max_{\sum_{i\in\mathcal{P}(j)} x_{ijk}=\mu_{jk}, x_{ijk}\ge 0}\left(r_{jk}\left(\mu_{jk}\right) + \sum_{i\in\mathcal{P}(j)}\left(-c_{ijk} + y_i - y_k\right)x_{ijk}\right) \\
= & \sum_{j,k\in V} \max_{0\le\mu_{jk}\le 1}\left(r_{jk}\left(\mu_{jk}\right) + \mu_{jk}\max_{i\in\mathcal{P}(j)}\left(-c_{ijk} + y_i - y_k\right)\right).
\end{aligned}$$

□

## C.2 Geometry of the Dual Function

### C.2.1 Entry Control Setting: Proof of Lemma 2

Recall Lemma 2.

**Lemma 2.** *Consider the congestion function* (6), *and any* $\boldsymbol{\phi}$ *with connectedness* $\alpha(\boldsymbol{\phi}) > 0$. *We have*

$$\mathcal{T}_3 \le -\alpha(\boldsymbol{\phi}) \cdot \left[f\left(\frac{1}{m}\right) - f\left(\min_{j\in V} \bar{q}_j\right) - 2m\right]^+ .$$

*Proof.* Order the nodes in $V$ in decreasing order of $y_j$ as $y_{i_1} \ge y_{i_2} \ge \cdots y_{i_m}$. For $r = 1$ to $r = m-1$, we repeat the following procedure: if $y_{i_r} - y_{i_{r+1}} \le 2$, then do nothing and move on to $r+1$; if otherwise, perform the following update:

$$y_{i_k} \leftarrow y_{i_k} - \left(y_{i_r} - y_{i_{r+1}} - 2\right) \qquad \forall 1 \le k \le r.$$

Recall that $g(\mathbf{y}) = \sum_{j,k\in V} \phi_{jk}[w_{jk} + y_j - y_k]^+$. For the terms where $j,k \in \{i_1,\cdots,i_r\}$ or $j,k \in \{i_{r+1},\cdots,i_m\}$, their value are not affected by the update. Consider the terms where $j \in \{i_1,\cdots,i_r\}$, $k \in \{i_{r+1},\cdots,i_m\}$: If $y_{i_r} - y_{i_{r+1}} > 2$, then after the update

$$w_{jk} + y_j - y_k \ge w_{jk} + y_{i_r} - \left(y_{i_r} - y_{i_{r+1}} - 2\right) - y_{i_{r+1}} \ge 2 > 0,$$

hence the update decrease these terms each by $y_{i_r} - y_{i_{r+1}} - 2$. Finally, for the terms where $j \in \{i_{r+1}, \cdots, i_m\}, k \in \{i_1, \cdots, i_r\}$, it is easy to verify that their value stay at zero after the update. To sum up, such an update decreases $g(\mathbf{y})$ by at least

$$\left( \sum_{k \le r, k' \ge r+1} \phi_{i_k, i_{k'}} \right) \cdot \left[ y_{i_r} - y_{i_{r+1}} - 2 \right]^+ .$$

Note that the first term is lower bounded by $\alpha(\boldsymbol{\phi})$ defined in (9). As a result, after the finishing the procedure, $g(\mathbf{y})$ decreased by at least:

$$\alpha(\boldsymbol{\phi}) \cdot \sum_{r=1}^{m-1} \left[ y_{i_r} - y_{i_{r+1}} - 2 \right]^+ \ge \ \alpha(\boldsymbol{\phi}) \cdot \left[ y_{i_1} - y_{i_m} - 2m \right]^+ .$$

By strong duality we have $\min_{\mathbf{y}} g(\mathbf{y}) = W^{\text{SPP}}$, hence

$$g(\mathbf{y}) - W^{\text{SPP}} \ge \ \alpha(\boldsymbol{\phi}) \cdot \left[ \max_{j \in V} y_j - \min_{k \in V} y_k - 2m \right]^+ .$$

Recall that the congestion cost is strictly increasing. This concludes the proof. □

#### C.2.2 Joint Entry-Assignment Setting

**Lemma 10.** *We have*

$$g_{\text{JEA}}(\mathbf{y}) - W^{\text{SPP}} \ge \ \alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D}) \cdot \left[ \max_{j \in V} y_j - \min_{k \in V} y_k - 2m \right]^+ ,$$

*where $W^{\text{SPP}}$ is the value of SPP* (25)-(27)*, and $\alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})$ is defined in* (29)*.*

*Proof Sketch.* The proof is a direct extension of the proof of Lemma 2. □

#### C.2.3 Joint Pricing-Assignment Setting

**Lemma 11.** *We have*

$$g_{\text{JPA}}(\mathbf{y}) - W^{\text{SPP}} \ge \ \alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D}) \cdot \left[ \max_{j \in V} y_j - \min_{k \in V} y_k - 2m \right]^+ ,$$

*where $W^{\text{SPP}}$ is the value of SPP* (40)-(42)*, and $\alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})$ is defined in* (29)*.*

*Proof Sketch.* The proof is a direct extension of the proof of Lemma 2. The key observation is that: if $y_i - y_k \ge 2 \ge 2 \max_{i,j,k \in V} |c_{ijk}| + \bar{p}$, then for any $j \in \mathcal{P}^{-1}(i)$ we have

$$\text{argmax}_{\{0 \le \mu_{jk} \le 1\}} \left( r_{jk}(\mu_{jk}) + \mu_{jk} \cdot \max_{i \in \mathcal{P}(j)} \left( -c_{ijk} + y_i - y_k \right) \right) = 1 ,$$

for any $j \in \mathcal{P}^{-1}(k)$ we have:

$$\text{argmax}_{\{0 \le \mu_{ji} \le 1\}} \left( r_{ji}(\mu_{ji}) + \mu_{ji} \cdot \max_{k \in \mathcal{P}(j)} \left( -c_{kji} + y_k - y_i \right) \right) = 0 .$$

□

## D The General Result

Instead of directly proving Lemma 3, we first state below a general lemma (Lemma 12) and Lemma 3 is its special case.

**Definition 1** (Growth condition of congestion function)**.** *Consider congestion function $f(\cdot)$ that is strictly increasing and continuously differentiable. Define*

$$\mathcal{B}(\mathbf{f}) \triangleq \left\{ \bar{\mathbf{q}} \in \Omega : \ f\left(\frac{1}{m}\right) - \min_{j \in V} f(\bar{q}_j) \leq 4m \right\} .$$

*Denote $\bar{\mathcal{B}}(\mathbf{f}) \triangleq \Omega \backslash \mathcal{B}(\mathbf{f})$. We say congestion function $f(\cdot)$s satisfies* growth condition *if the following holds.*

- *For fixed $\alpha > 0$, there exists $K_1 = K_1(\alpha, \mathbf{f}) > 0$ such that $\forall K > K_1$, $\forall \bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$,*

$$\alpha \left( f\left(\frac{1}{m}\right) - \min_{j \in V} f(\bar{q}_j) - 2m \right)^+ \geq \frac{1}{2\tilde{K}} \cdot \max_{j \in V} f'(\bar{q}_j) + \mathbb{1}\{q_j = 0, \exists j\} .$$

- *There exists a universal constant $C < \infty$ such that $\sup_{\bar{\mathbf{q}} \in \mathcal{B}(\mathbf{f})} \max_{j \in V} f'(\bar{q}_j) \leq C$.*
- *For any $\bar{\mathbf{q}} \in \mathcal{B}(\mathbf{f})$, we have $\mathbf{q} > \mathbf{0}$.*

**Lemma 12.** *In the entry control setting, if a strictly increasing and continuously differentiable congestion function $f(\cdot)$ satisfies the growth condition given in Definition 1, then*

$$\mathcal{T}_2 + \mathcal{T}_3 + \mathcal{T}_4 \leq C \frac{1}{\tilde{K}} ,$$

*where $\mathcal{T}_2, \mathcal{T}_3, \mathcal{T}_4$ are defined in* (19)*.*

**Theorem 4** (General result for entry control setting)**.** *Consider a set of $m \in \mathbb{N}$ locations and any demand arrival rates $\boldsymbol{\phi}$ that satisfies Condition 1, and congestion function $f(\cdot)$ that satisfies the growth condition given in Definition 1. Then there exists $K_1 = K\big(\alpha(\boldsymbol{\phi}), \mathbf{f}\big)$ such that the following guarantees hold for Algorithm 1*

$$L_T^{\mathrm{MBP}} \leq M_1 \cdot \frac{K}{T} + M_2 \cdot \frac{1}{K} ,$$

*where*

$$M_1 = \sup_{\bar{\mathbf{q}}_1, \bar{\mathbf{q}}_2 \in \Omega} \left( F(\bar{\mathbf{q}}_1) - F(\bar{\mathbf{q}}_2) \right) , \qquad M_2 = C .$$

### D.1 Verifying the Growth Condition

To finish the proof of Lemma 3 and Theorem 1, the only remaining step to specify the dependence of $K_1, M_1, M_2$ on $m$ and $\boldsymbol{\phi}$. Note that the following Corollary together with Lemma 12 and Theorem 4 implies Lemma 3 and Theorem 1.

**Corollary 1.** *Let $\delta_K = \sqrt{K}$. We have the following explicit dependence of $K_1, M_1, M_2$ on $m$ and $\boldsymbol{\phi}$.*

- *For logarithmic mirror map, i.e.* $f(\bar{q}) = c \cdot \log(\bar{q})$, *if* $c \geq \max\left\{8m, \frac{2}{\alpha(\boldsymbol{\phi})}\right\}$, *Theorem 4 holds for*

$$K_1 = \Theta\left(\max\left\{\frac{1}{\alpha^2(\boldsymbol{\phi})}, m^2, \frac{c}{\alpha(\boldsymbol{\phi})}\right\}\right), \quad M_1 = \Theta\left(m + c\log m\right), \quad M_2 = \Theta(cm).$$

- *For inverse square root mirror map* $f(\bar{q}) = \sqrt{m} \cdot \bar{q}^{-\frac{1}{2}}$, *Theorem 4 holds for*

$$K_1 = \Theta\left(\max\left\{\frac{m}{\alpha(\boldsymbol{\phi})}, \frac{1}{m^2 \cdot \alpha^4(\boldsymbol{\phi})}, m^2\right\}\right), \ M_1 = \Theta(m), \ M_2 = \Theta(m^2).$$

- *For linear mirror map, i.e.* $f(\bar{q}) = c \cdot \bar{q}$, *if* $c \geq 2m^2 + \frac{2}{\alpha(\boldsymbol{\phi})}m$, *Theorem 4 holds for*

$$K_1 = \Theta\left(\frac{c}{\alpha(\boldsymbol{\phi})}\right), \quad M_1 = \Theta\left(\max\{c, m\}\right), \quad M_2 = \Theta(c).$$

*Proof.* **Logarithmic function.** Let $\delta_K = \sqrt{K}$. Let $\bar{q}_{\min}$ be the solution of:

$$f\left(\frac{1}{m}\right) - f(\bar{q}_{\min}) = 4m.$$

We have $\bar{q}_{\min} = \frac{1}{m} \cdot e^{-\frac{4m}{c}}$. It is not hard to see that $\mathcal{B}(\mathbf{f}) = \{\bar{\mathbf{q}} \in \Omega : \min_{j\in V} f(\bar{q}_j) \geq \bar{q}_{\min}\}$. The growth condition given by Definition 1 can be written equivalently as the following inequalities:

$$\alpha\left(c \cdot \log\left(\frac{\frac{K}{m} + \sqrt{K}}{\sqrt{K}}\right) - 2m\right) \geq \frac{c}{2} \cdot \frac{1}{\sqrt{K}} + 1 \tag{49}$$

$$\alpha\left(c \cdot \log\left(\frac{1}{m\bar{q}}\right) - 2m\right) \geq \frac{c}{2(K + m\sqrt{K})} \cdot \frac{1}{\bar{q}} \qquad \forall \bar{q} \in \left(\frac{\sqrt{K}}{K + m\sqrt{K}}, \bar{q}_{\min}\right) \tag{50}$$

$$\frac{1}{m} \cdot e^{-\frac{4m}{c}} > \frac{\sqrt{K}}{K + m\sqrt{K}}. \tag{51}$$

It is easy to verify that the function obtained by subtracting the RHS of (50) from the LHS of (50) has the only stationary point at $\bar{q} = \frac{1}{2\alpha(K+m\sqrt{K})}$. As a result, it suffices for the inequality to hold for the endpoints of the interval and the stationary point: the inequality for the left endpoint is subsumed by (49), hence we can replace (50) by:

$$\alpha\left(c \cdot \log\left(\frac{1}{m\bar{q}_{\min}}\right) - 2m\right) \geq \frac{c}{2(K + m\sqrt{K})} \cdot \frac{1}{\bar{q}_{\min}}, \tag{52}$$

$$\alpha\left(c \cdot \log\left(\frac{2\alpha(K + m\sqrt{K})}{m}\right) - 2m\right) \geq \alpha c. \tag{53}$$

For the mirror map to satisfy the growth condition given by Definition 1, it suffices to find a $c$ and $K_1$ such that (49)(51)(52)(53) hold. A sufficient condition for (49) to hold is:

$$K_1 \geq m^2 e^{\frac{1}{\alpha\sqrt{K_1}} + \frac{2}{c}\left(\frac{1}{\alpha} + 2m\right)}.$$

A sufficient condition for (51) to hold is:

$$K_1 > m^2 e^{\frac{8m}{c}}.$$

A sufficient condition for (52) to hold is:

$$K_1 \geq \frac{c}{4\alpha} e^{\frac{4m}{c}}\,.$$

A sufficient condition for (53) to hold is:

$$K_1 \geq \frac{2m}{\alpha} e^{\frac{2m}{c}}\,.$$

Combined, a sufficient condition for (49)(51)(52)(53) to hold is:

$$K_1 \geq 2e^{\frac{8m}{c}} \cdot \max\left\{ m^2 e^{\frac{1}{\alpha(\boldsymbol{\phi})\sqrt{K_1}} + \frac{2}{\alpha(\boldsymbol{\phi})c}}\,,\ \frac{m}{\alpha(\boldsymbol{\phi})}\,,\ \frac{c}{\alpha(\boldsymbol{\phi})} \right\}$$

Thus, for

$$\begin{aligned} c &\geq \mathrm{poly}\left(m, \frac{1}{\alpha(\boldsymbol{\phi})}\right) \triangleq \max\left\{8m, \frac{2}{\alpha(\boldsymbol{\phi})}\right\} \\ K_1 &= \mathrm{poly}\left(c, m, \frac{1}{\alpha(\boldsymbol{\phi})}\right) \triangleq 6\max\left\{\frac{1}{\alpha^2(\boldsymbol{\phi})}, \frac{m}{\alpha(\boldsymbol{\phi})}, \frac{c}{\alpha(\boldsymbol{\phi})}, m^2\right\}\,, \end{aligned}$$

Condition 1 holds.

We proceed to bound $M_1$ and $M_2$. Note that

$$\begin{aligned} \sup_{\mathbf{q},\mathbf{q}'\in\Omega^K} \left(F(\bar{\mathbf{q}}) - F(\bar{\mathbf{q}}')\right) &= c \cdot \sup_{\mathbf{q},\mathbf{q}'\in\Omega^K} \left(\sum_j \bar{q}_j \log \bar{q}_j - \sum_j \bar{q}'_j \log \bar{q}'_j\right) \\ &\leq\ -c \cdot \min_{\mathbf{q}'\in\Omega^K} \sum_j \bar{q}'_j \log \bar{q}'_j \\ &= c \log m\,, \end{aligned}$$

where the inequality follows from the fact that $\bar{q}_j \in (0,1)$. Hence

$$M_1 = \mathrm{poly}(c,m) = m + c\log m\,.$$

For $M_2$ we have

$$M_2 = \max_{\bar{\mathbf{q}}\in\mathcal{B}_{\mathbf{f}}} \max_j |f'(\bar{q}_j)| = f'(\bar{q}_{\min}) \leq \mathrm{poly}(c,m) = 2cm\,,$$

where the last inequality follows from the definition of $\bar{q}_{\min}$.

**Inverse square root function.** Now we consider the inverse square root function. Similar to the analysis above, we have the following inequalities:

$$\alpha\left(c\cdot\sqrt{m+\sqrt{K}} - c\cdot\sqrt{m} - 2m\right) \geq \frac{c}{4} K^{-\frac{1}{2}} \sqrt{m+\sqrt{K}} + 1\,, \tag{54}$$

$$\alpha\left(c\cdot\bar{q}^{-\frac{1}{2}} - c\cdot\sqrt{m} - 2m\right) \geq \frac{c}{4}\left(m\sqrt{K}+K\right)^{-1} \bar{q}^{-\frac{3}{2}}\,, \quad \forall \bar{q} \in \left(\frac{\sqrt{K}}{K+m\sqrt{K}},\ \bar{q}_{\min}\right) \tag{55}$$

$$\bar{q}_{\min} \geq \frac{1}{m+\sqrt{K}}\,. \tag{56}$$

Where $c\cdot\bar{q}_{\min}^{-\frac{1}{2}} - c\cdot\sqrt{m} = 4m$. Let $c=\sqrt{m}$, hence $\bar{q}_{\min} = \frac{1}{25m}$, *and* a sufficient condition for the above inequalities to hold is:

$$K_1 \geq 60\left(\frac{1}{2\alpha\sqrt{m}} + \sqrt{m}\right)^4\,,$$

$$K_1 \geq \frac{125m}{8\alpha},$$
$$K_1 \geq 625m^2,$$
$$K_1 \geq 360\frac{m^2}{\alpha^2}.$$

To sum up, the above condition is:

$$K_1 = \text{poly}\left(m, \frac{1}{\alpha(\boldsymbol{\phi})}\right) \triangleq \Theta\left(\max\left\{\frac{m^2}{\alpha^2(\boldsymbol{\phi})}, \frac{1}{m^2 \cdot \alpha^4(\boldsymbol{\phi})}\right\}\right).$$

Now we proceed to bound $M_1$ and $M_2$. For $M_1$ we have

$$\begin{aligned}
\sup_{\mathbf{q},\mathbf{q}' \in \Omega^K} \left(F(\bar{\mathbf{q}}) - F(\bar{\mathbf{q}}')\right) &= 2c \cdot \sup_{\mathbf{q},\mathbf{q}' \in \Omega^K} \left(\sum_j \sqrt{\bar{q}_j} - \sum_j \sqrt{\bar{q}'_j}\right) \\
&\leq 2cm\sqrt{\frac{1}{m}} \\
&= 2m,
\end{aligned}$$

Hence $M_1 = O(m)$. Similar to the logarithmic function case we have

$$M_2 = f'(\bar{q}_{\min}) \leq \text{poly}(m) = 70m^2.$$

**Linear Function.** Similar to the analysis above, we have:

$$\alpha\left(c \cdot \left(\frac{1}{m} - 0\right) - 2m\right) \geq \frac{c}{2K} + 1,$$
$$\alpha\left(c \cdot \left(\frac{1}{m} - c\bar{q}_{\min}\right) - 2m\right) \geq \frac{c}{2K},$$
$$c \cdot \left(\frac{1}{m} - \bar{q}_{\min}\right) = 4m, \; \bar{q}_{\min} > 0.$$

A sufficient condition for the above inequalities to hold is:

$$K_1 \geq \frac{c}{2} \cdot \frac{1}{\alpha\left(\frac{c}{m} - 2m\right) - 1}$$
$$K_1 \geq \frac{c}{2} \cdot \frac{1}{\alpha\left(c\left(\frac{1}{m} - \bar{q}_{\min}\right) - 2m\right)}$$
$$\bar{q}_{\min} = \frac{1}{m} - \frac{4m}{c} > 0.$$

A sufficient condition is:

$$K_1 \geq \text{poly}\left(m, \frac{1}{\alpha(\boldsymbol{\phi})}\right) \triangleq \frac{c}{4\alpha(\boldsymbol{\phi})m}, \qquad c \geq 4m^2 + \frac{2m}{\alpha(\boldsymbol{\phi})}.$$

In this case we have

$$M_1 = m + c, \qquad M_2 = c.$$

□

*Proof of Theorem 1.* Lemma 3 is implied by Corollary 1 and Lemma 12. Given Lemma 1, 2, and

3, the proof of Theorem 1 follows using the argument given in Section 5.3. □

*Proof Sketch of Theorem 2.* The proof is a direct extension of the proof of Theorem 1: by plugging Lemma 10 in Lemma 8, then showing the growth condition analogous to the above, the result naturally follows. □

*Proof Sketch of Theorem 3.* The proof is a direct extension of the proof of Theorem 1: by plugging Lemma 11 in Lemma 9, then showing the growth condition analogous to the above, the result naturally follows. □

# E Simulation Settings

**Model Primitives.**

- *Demand arrival process.* Using the estimation in Buchholz (2015), which is based on Manhattan's taxi trip data during August and September in 2012, we obtain the (average) demand arrival rates for each origin-destination pair during the day (7 a.m. to 4 p.m.).
- *Pickup/service times.* We extract the pairwise travel time between region centroids (marked by the dots in Figure 4) using Google Maps, denoted by $D_{ij}$'s ($i, j = 1, \cdots, 30$). We use $D_{jk}$ as service time for customers traveling from $j$ to $k$. For each customer at $j$ who is picked up by a supply from $i$ we add a pickup time [25] of $\tilde{D}_{ij} = \max\{D_{ij}, 2 \text{ minutes}\}$.

**Benchmark policy: static fluid-based policy.** We consider the fluid-based randomized policy (Banerjee et al. 2016, Ozkan and Ward 2016) as a benchmark. Let $\mathbf{x}^*$ be a solution of SPP. When a type $(j, k)$ demand arrives at location $j$, the randomized fluid-based policy dispatches from location $i \in \mathcal{N}(j)$ with probability $x^*_{ijk}$.

**Initial state generation.** We first uniformly sample 100 points from the simplex $\{\mathbf{q} : \sum_{i \in V_S} q_i = K\}$, which are used as the system's initial states at 6 a.m. (note that all the cars are free). Then we "warm-up" the system by employing the static policy from 6 a.m. to 8 a.m., assuming the demand arrival process during this period to be stationary (with the average demand arrival rate during this period as mean). Finally, we use the system's states at 8 a.m. as the initial states for the simulations in Section 8.

[25] We use the inflated $D_{ij}$'s as pickup times to account for delays in finding or waiting for the customer.