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 a finite, fixed number of supply units which circulate in the system. Demand arrives stochastically, and serving a demand unit (customer) causes a supply unit to relocate from the "origin" to the "destination" of the customer. The key challenge is to manage the distribution of supply in the network. We consider general controls including customer 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 and rich 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 (customer) arrival rates (these rates are permitted to vary slowly in time). Under mild conditions, we propose MBP policies that are provably near optimal. Specifically, our policies lose at most $O(\frac{K}{T} + \frac{1}{K} + \sqrt{\eta K})$ payoff per customer relative to the optimal policy that knows the demand arrival rates, where K is the number of supply units, T is the total number of customers over the time horizon, and η is the maximum change in demand arrival rates per period (i.e., per customer arrival). A natural model of a scrip system is a special case of our setup. An adaptation of MBP is found to perform well in a realistic ride-hailing environment.

Keywords: control of queueing networks; backpressure; mirror descent.

1 Introduction

The control of complex systems with circulating resources such as shared transportation platforms and scrip systems has been heavily studied in recent years. The hallmark of such systems is that

*Decision, Risk and Operations Division, Columbia Business School, Email: `ykanoria@columbia.edu`

†Decision, Risk and Operations Division, Columbia Business School, Email: `PQian20@gsb.columbia.edu`

serving a demand unit causes a (reusable) supply unit to be relocated. Closed queueing networks (i.e., networks where a fixed number of supply units circulate in the system) provide a powerful abstraction for these applications (Banerjee, Freund, and Lykouris 2016, Braverman et al. 2019). The key challenge is *managing the distribution of supply* in the network. 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 control policy is near-optimal in a certain asymptotic regime. However, this approach only works under the restrictive assumption that *(1) the system parameters (demand arrival rates) are precisely known.* Furthermore, previous papers (Banerjee, Freund, and Lykouris 2016, Braverman et al. 2019) assume that *(2) the system is in steady state.* As is pointed out by Banerjee, Freund, and Lykouris (2016), relaxing either of these assumptions has been of interest.

In this paper, we relax *both* assumptions.[1] We propose a family of simple, practical control policies that are *blind* in that they use *no* prior knowledge of demand arrival rates, and prove strong transient and steady state performance guarantees for these policies, for demand arrival rates that are stationary or vary slowly in time. In simulations, our policies achieve excellent performance that beats the state-of-the-art policies even in an unequal contest where the latter policies are provided exact demand arrival rates 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, and demand is stationary. Later we allow other controls including dynamic pricing, and flexible assignment of resources, and moreover allow for time-varying demand arrival rates, and show that our machinery and guarantees extends seamlessly. In our baseline entry control model, we consider a closed queueing network consisting of m nodes (locations), 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 with some stationary arrival rates which are unknown to the controller. The controller dynamically decides whether to admit each 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

[1]The paper Banerjee, Kanoria, and Qian (2018) is similarly motivated, but restricts attention to a narrow special case: assignment control in networks satisfying a strong *complete resource pooling* (CRP) assumption, and conducts a sharp large deviations analysis. In particular, non-idling/greedy policies suffice to achieve asymptotic optimality under CRP. In contrast, the present work is general: e.g., the JEA setting we solve in Section 6 generalizes the model of that paper by dropping the CRP assumption, necessitating a completely different non-greedy approach to control; already under our illustrative model (Section 2) the CRP assumption of Banerjee, Kanoria, and Qian (2018) is automatically violated and the greedy policy fails to achieve asymptotic optimality.

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.

Notably, the greedy policy, which admits a demand unit if a supply unit is available, is generically far from optimal: even as $K \to \infty$, the optimality gap per demand unit of this policy is $\Omega(1)$ even in steady state; see Remark 1 in Section 2. The intuition is that some nodes have no available supply an $\Omega(1)$ fraction of the time in steady state under the greedy policy, and so the policy is forced to drop a significant proportion of the demand which would have been served under the optimal policy. Furthermore, if demand arrival rates are imperfectly known, any state independent policy (such as that of Banerjee, Freund, and Lykouris 2016) generically suffers a steady state optimality gap per demand unit of $\Omega(1)$; see Banerjee, Kanoria, and Qian (2018, Proposition 4).

Preview of our main result. 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 policies are near optimal. Specifically, we show that our policies lose payoff (per demand unit) at most $O\left(\frac{K}{T} + \frac{1}{K}\right)$ relative to the optimal policy that knows the demand arrival rates, where K is the number of supply units, T is the number of demand units that arrive during the period of interest. Our result is *non-asymptotic*, i.e., our performance guarantee holds for finite K and T, and thus covers both transient and steady state performance. In particular, taking $T \to \infty$, we obtain a steady state optimality gap of $O(\frac{1}{K})$, matching that of the state-of-the-art policy of Banerjee, Freund, and Lykouris (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. Notably, 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 (see the discussion after Theorem 1): here the number of arrivals $T = \Theta(K \cdot T^{\text{real}})$, where T^{real} is the time horizon measured in physical time, and we can rewrite our bound on the optimality gap as $O\left(\frac{1}{T^{\text{real}}} + \frac{1}{K}\right) \xrightarrow{K\to\infty} O\left(\frac{1}{T^{\text{real}}}\right)$.

Our policies retain their good performance if demand arrival rates vary slowly over time: We show (see Section 6.2) that the loss in payoff per customer under MBP is bounded by $O\left(\frac{K}{T} + \frac{1}{K} + \sqrt{\eta K}\right)$, where is η is the maximum change in demand arrival rates per customer arrival. In the aforementioned large market regime, the optimality gap per customer can be expressed as $O\left(\frac{1}{T^{\text{real}}} + \frac{1}{K} + \sqrt{\zeta}\right)$ where $\zeta \triangleq \eta K$ is rate of change of $\boldsymbol{\phi}$ with respect to physical time.

We now motivate and introduce our control policies. First, we describe 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 node 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. (We emphasize the reversal of the usual mapping: in our setup supply units are "jobs" and demand units act as service tokens.) 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.[2] (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) for the control of queueing networks. Backpressure simply uses queue lengths as congestion costs (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 congestion costs. Concretely, in our baseline entry control setting, backpressure admits a demand if and only if the payoff of serving the demand plus the origin queue length exceeds the destination queue length. This simple approach has been used very effectively in a range of settings arising in cloud computing, networking, etc.; see, e.g. Georgiadis et al. (2006). Backpressure is provably near-optimal (in the large market limit) in many settings where payoffs accrue from serving jobs, 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 appears to be the reason that backpressure has not yet been proposed as a control approach in such settings with circulating resources).

Main challenge: no-underflow constraint. The control policy must satisfy the *no-underflow*

[2] There are subtle differences between our model and "classical" closed queueing networks in the timing of when a job joins the destination queue, and when the "service" of a job is initiated. These differences are non-essential, see, e.g., Banerjee, Kanoria, and Qian (2018, Section 8).

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, in the well known "crossbar switch" problem in Maguluri and Srikant (2016), there are no "payoffs" apart from the shadow prices (the goal is only to prevent queues from building up), so backpressure only recommends to serve a queue with positive length (after all, 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.[3] 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 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 designed 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 continuum limit optimization problem,

[3]For example, Dai and Lin (2005) assume that the network satisfies a so-called Extreme Allocation Available (EAA) condition, which ensures that the no-underflow constraint does not bind; 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 variant of backpressure avoids underflow entirely under this assumption. Gurvich and Ward (2014) assume that the network satisfies a so-called Dedicated Item (DI) condition.

with the chosen mirror map.[4] 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). In case of finite buffers, we find it beneficial to use congestion functions which moreover increase steeply as the queue length approaches buffer capacity, to prevent buffer overflow (Section 6.1).

We develop a general machinery to prove performance guarantees for MBP, which draws inspiration from two distinct toolkits: the machinery for proving convergence of mirror descent from the optimization literature, and the Lyapunov drift method from the network control literature. We provide a ready Lyapunov function for any MBP policy. Furthermore, we improve upon the Lyapunov drift method to obtain a sharp bound on the suboptimality caused by the no-underflow constraint. Our analysis exploits the structure of the platform's continuum limit optimization problem in a novel way (see Section 5).

Our work fits into the broad literature on the control of stochastic processing networks (Harrison 2000). Our MBP methodology for designing blind control policies with provable guarantees applies to open queueing networks as well. We are optimistic that MBP will prove broadly useful in the control of queueing networks.

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

(i) **Mirror Backpressure: a class of near-optimal control policies for queueing networks that are completely blind.** In general settings that consider entry control, pricing, and flexible assignment, we propose a family of dynamic control policies for queueing networks, the Mirror Backpressure policies, that have strong transient and steady state performance guarantees. The MBP policies are simple and practical, and do not require any prior knowledge of demand arrival rates (which are permitted to vary in time), making them promising for applications. Policy design boils down to choosing suitable congestion functions.

(ii) **A framework for systematic design and analysis of MBP control policies.** Our control framework has a tight connection with mirror descent, which makes the process of policy design and analysis both systematic and flexible, and allows us to handle the challenging no-underflow constraint. The general machinery we develop can be seamlessly leveraged to design policies with provable guarantees for a variety of settings. This is in contrast with various intricate approaches in the queueing literature that do not easily generalize.

[4] 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).

In Section 6 we generalize the baseline model (which allows entry control only) and include pricing and flexible assignment as control levers. We study joint entry-assignment control (JEA) in Section 6.2 and joint pricing-assignment control (JPA) in Section 6.3. Our control policies and performance guarantees extend seamlessly.

Applications. Our general model (Section 6) includes a number of key ingredients common to many applications. We illustrate its versatility by discussing the application to shared transportation systems (Section 7) and the application to scrip systems (Section 8). These applications and the relevant settings in the paper are summarized in Table 1.

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 systems	Admission & Provider selection	Joint entry-assignment

Table 1: Summary of applications of our model, the control levers therein and the corresponding settings in this paper. See Section 6 for the joint entry-assignment and joint pricing-assignment settings (which allow for finite buffers). For each setting, we design MBP policies that are near optimal.

Shared transportation systems include ride-hailing and bike sharing systems. Here the nodes in our model correspond to geographical locations, while supply units and demand units correspond to vehicles and customers, respectively. Bike sharing systems dynamically incentivize certain trips using point systems to minimize out-of-bike and out-of-dock events caused by demand imbalance. Our pricing setting is relevant for the design of a dynamic incentive program for bike sharing; in particular, it allows for a limited number of docks. Ride-hailing platforms make dynamic decisions to optimize their objectives (e.g., revenue, welfare, etc.). For ride-hailing, our pricing-assignment model is relevant in regions such as North America, and our entry-assignment control model is relevant in in regions where dynamic pricing is undesirable like in China. We perform realistic simulations of ride-hailing and find that our MBP policy, suitably adapted to account for positive travel times, performs well (Section 7.1 and Appendix D).

A scrip system is a nonmonetary trade economy where agents use scrips (tokens, coupons, artificial currency) to exchange services (because monetary transfer is undesirable or impractical), e.g., for babysitting or kidney exchange. A key challenge in these markets is the design of the admission-and-provider-selection rule: If an agent is running low on scrip balance, should they be allowed to request services? If yes, and if there are several possible providers for a trade, who should be selected as the service provider? In Section 8, we show that a natural model of a scrip system is a special case of our entry-assignment control setting, yielding a near optimal admission-and-provider-selection control rule.

1.1 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. Attractive features of MaxWeight/backpressure policies include their simplicity and provably good performance, and that arrival/service rate information is not required beforehand. The main challenge in using backpressure is the no-underflow constraints, as described earlier. Most of this literature considers the open queueing networks setting, where packets/jobs enter and leave, and there is much less work on closed networks. An exception is a recent paper on assignment control of closed networks by Banerjee, Kanoria, and Qian (2018), which shows the large deviations optimality of "scaled" MaxWeight policies. Importantly, in that paper the demand arrival rates are assumed to satisfy a strong near balance assumption ("complete resource pooling"), as a result of which it suffices to consider non-idling policies (i.e., a "greedy" policy with assignment control only). In the present paper, in contrast, we allow very general demand arrival rates, which makes it necessary to deploy idling policies (e.g., entry control, pricing) to achieve good performance. Indeed, already under our illustrative model (Section 2) the CRP assumption of Banerjee, Kanoria, and Qian (2018) is automatically violated and the greedy policy fails to achieve asymptotic optimality; see Remark 1 in Section 2.

While previous works use queue lengths or their power as congestion costs (Stolyar 2004), our MBP policies significantly generalize backpressure by allowing a general increasing function (e.g., the logarithm) of queue lengths as congestion costs. As with backpressure, MBP policies carry provable guarantees.

Mirror Descent. Mirror descent (MD) is a generalization of the gradient descent algorithm for optimization, which was proposed by Nemirovsky and Yudin (1983), see also Beck and Teboulle (2003). MD is much more flexible than gradient descent as one can freely choose a "mirror map" that captures the geometry of the problem (including its objective and its constraints). Recently, there have been several works that use MD to solve online decision-making problems (e.g., Gupta and Radovanović 2020). Notably, Bubeck et al. (2018) uses MD to obtain an improved approximation factor for a worst-case version of the so-called "k-server problem"; the k-server problem bears a certain resemblance to our setting in that the controller needs to manage the spatial distribution of supply. A key difference between our work and the existing works is that our proposed simple control policies remarkably have the *property* that they induce the queue lengths to follow MD dynamics, whereas the existing works actively run MD to solve their algorithmic problems.

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, Freund, and Lykouris (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 dynamic pricing for a specific network of star structure. (Ma et al. (2019) studied the somewhat different issue of ensuring that drivers have the incentive to accept dispatches by setting prices which are sufficiently smooth in space and time, in a model with no demand stochasticity.) Banerjee, Kanoria, and Qian (2018) which assumes a near balance condition on demands and equal pickup costs may be the only paper in this space that does not require knowledge of system parameters. Comparing with Banerjee, Freund, and Lykouris (2016) which obtains a steady state optimality gap of $O(\frac{1}{K})$ (in the absence of travel times) assuming *perfect* knowledge of demand arrival rates which are assumed to be *stationary*, our control policy achieves the same steady state optimality gap with *no* knowledge of demand arrival rates, and further achieves a *transient* optimality gap under *time-varying* demand arrival rates of $O(\frac{K}{T} + \frac{1}{K} + \sqrt{\eta K})$ for a finite number of arrivals T and changes of up to η per period (i.e., per arrival) in demand arrival rates. Some of these papers are able to formally handle travel delays: Braverman et al. (2019), Banerjee et al. (2016), and Banerjee, Kanoria, and Qian (2018) prove theoretical results for the setting with i.i.d. geometric/exponential travel delays; Ma et al. (2019) consider deterministic travel delays. On the other hand, Balseiro et al. (2019) ignores travel delays in their theory and later heuristically adapt their policy to accommodate travel delay (the present paper follows a similar approach). On the other hand, Ozkan and Ward (2016) is the only paper among these which (like the present paper) allows time-varying demand.

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 our policy, and our methodology leads to control rules in much more general settings as described in Section 8.

Other related work. A related stream of research studies 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. Network revenue management is a classical set of (open network) dynamic resource allocation problems, e.g., see Gallego and Van Ryzin (1994), Talluri and Van Ryzin (2006), and recent works, e.g., Bumpensanti and Wang (2018). 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. Again, this is an open network setting in that each supply unit can be used only once.

1.2 Organization of the Paper

The remainder of our paper is organized as follows. 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 provide MBP policies for joint entry-assignment and joint pricing-assignment control settings and allow for time-varying demand arrival rates, demonstrating the versatility of our approach. In Sections 7 and 8 we discuss the applications to shared transportation systems and scrip systems, respectively.

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 $t = 0, 1, 2, \dots$, where a fixed number (denoted by K) of identical *supply units* circulate among a set of *nodes* V (locations), with $m \triangleq |V| > 1$. In our model, t will capture the number of demand units (customers) who have arrived so far (minus 1).

Queues (system state). At each node $j \in V$, there is an infinite-buffer queue of supply units. (Section 6.1 shows how to seamlessly incorporate 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 $\Omega_K \triangleq \{\mathbf{q} : \mathbf{q} \in \mathbb{Z}_+^m, \mathbf{1}^\top \mathbf{q} = K\}$, 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 her type by $(o[t], d[t]) \in V \times V$, where $o[t]$ is her origin and $d[t]$ is her destination. With probability ϕ_{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 exact knowledge of demand arrival rates (Ozkan and Ward 2016, Banerjee, Freund, and Lykouris 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, we 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 $(o[t'], d[t'])_{t' \leq t}$ and states $(\mathbf{q}[t'])_{t' \leq t}$ 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 policies will be deterministic. The set of feasible

[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 (ϕ_{jk})s. The discrete time model considered is the embedded chain of the continuous time model.

policies is denoted by $\mathcal{U}$.

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

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

We use $v^\pi[t]$ to denote the payoff collected at time t under control policy π. Let W_T^π denote the average payoff per period (i.e., per customer) collected by policy π 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:

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

Define the infinite-horizon per period payoff W^π collected by policy π and the optimal per period payoff over all admissible policies W^* respectively as:

$$W^\pi \triangleq \liminf_{T\to\infty} W_T^\pi\,, \qquad W^* \triangleq \limsup_{T\to\infty} W_T^*\,. \tag{4}$$

We measure the performance of a control policy π by its per-customer *optimality gap* ("loss"):

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

Note that we consider the worst-case initial system state when evaluating a given policy, and the best initial state for the optimal benchmark; see (3). Such a definition of optimality gap provides a conservative bound on policy performance and avoids the (unilluminating) discussion of the dependence of performance on initial state.

We make the following mild connectivity assumption on the demand arrival rates $\boldsymbol{\phi}$.

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

$$\alpha(\boldsymbol{\phi}) \triangleq \min_{S\subsetneq V, S\neq\emptyset} \sum_{j\in S, k\in V\backslash S} \phi_{jk}\,. \tag{6}$$

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 nodes (j,k), there is a sequence of demand types with positive arrival rate that would take a supply unit from j eventually to k.

We conclude this section with an example which shows that the greedy policy typically has steady state optimality gap $\Omega(1)$ per period, followed by the observation that the main assumption of Banerjee, Kanoria, and Qian (2018) is automatically violated in our setting.

Example 1 (Greedy policy typically incurs $\Omega(1)$ loss)**.** *Consider a network with three nodes $V = \{1,2,3\}$, demand arrival probabilities $\phi_{12} = \epsilon$, $\phi_{23} = \frac{1}{3} + \epsilon$, $\phi_{21} = \phi_{32} = \frac{1}{3} - \epsilon$ (where $0 < \epsilon < \frac{1}{6}$), and payoffs $w_{23} = w > 0$, $w_{12} = w_{21} = w_{32} = \frac{w}{2}$. Let $\mathbf{x}^*$ be the optimal solution to the SPP* (10)-(12)*. By inspection, $\mathbf{x}^*$ should induce the maximum circulation in each of the two cycles* 1—2—1 *and* 2—3—2*, hence $x^*_{12} = x^*_{32} = 1$, $x^*_{21} = \frac{\epsilon}{\frac{1}{3}-\epsilon}$, $x^*_{23} = \frac{\frac{1}{3}-\epsilon}{\frac{1}{3}+\epsilon}$. We know that there exists a policy whose performance approaches the value of the SPP as $K \to \infty$ (Banerjee, Freund, and Lykouris 2016). We will prove by contradiction that the greedy policy incurs an $\Omega(1)$ loss for this example, by showing that its payoff per period is $\Omega(1)$ below the value of the SPP. Consider the steady state under the greedy policy. Suppose the loss is vanishing, i.e., all but an $o(1)$ fraction of type $(1,2)$ and type $(3,2)$ demand are served. Suppose a γ fraction of the time there is a supply unit present at node 2. As a result, since the greedy policy is being used, a γ fraction of demands of type $(2,1)$ are served, and a γ fraction of demands of type $(2,3)$ are served. Flow-balance at nodes 1 and 3, respectively, implies that we have $(\frac{1}{3} - \epsilon)\gamma = \epsilon - o(1)$, $(\frac{1}{3} + \epsilon)\gamma = \frac{1}{3} - \epsilon - o(1)$. However, these two equations cannot both be satisfied as $K \to \infty$ unless $\epsilon = \frac{1}{9}$. We infer that the greedy policy incurs an $\Omega(1)$ loss in this network for any $\epsilon \in (0, \frac{1}{6})$, $\epsilon \neq \frac{1}{9}$.*

Remark 1. *The complete resource pooling (CRP) condition imposed in Banerjee, Kanoria, and Qian (2018, Assumption 3) is automatically violated in the model we have defined in this section. Consider our setup including Condition 1. The CRP condition can be stated as follows: for each subset of nodes $S \subsetneq V, S \neq \emptyset$, the "net demand" $\mu_S \triangleq \sum_{i\in S}\sum_{j\in V\backslash S} \phi_{ij}$ is less than the "net supply" $\lambda_S \triangleq \sum_{j\in V\backslash S}\sum_{i\in S} \phi_{ji}$, i.e., $\mu_S < \lambda_S$. Clearly, any demand arrival rates $\boldsymbol{\phi}$ violate CRP, since if $\mu_S < \lambda_S$ for some $S \subsetneq V, S \neq \emptyset$ then this means that $\mu_{V\backslash S} > \lambda_{V\backslash S}$ (given that $\mu_{V\backslash S} = \lambda_S$ and $\lambda_{V\backslash S} = \mu_S$ by definition), i.e., CRP is violated. In Example 1, the subset $\{2\}$ (and the subset $\{1,2\}$) violates this constraint.*

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.

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 (normalized) queue length to a congestion cost at that node, based on which MBP will make its decisions. (We will defined normalized queue lengths $\bar{\mathbf{q}}$ below.)

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 algorithms, the choice of congestion function should depend on the constraints of the setting, leading 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

$$f(\bar{q}_j) \triangleq -\sqrt{m} \cdot \bar{q}_j^{-\frac{1}{2}}\,, \tag{7}$$

and postpone the results for other choices of congestion functions to Appendix C (see also Remark 2). For technical reasons, we need to keep $\bar{\mathbf{q}}$ in the *interior* of the normalized state space Ω, 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

$$\bar{q}_i \triangleq \frac{q_i + \delta_K}{\tilde{K}} \quad \text{for} \quad \delta_K \triangleq \sqrt{K} \quad \text{and} \quad \tilde{K} \triangleq K + m\delta_K\,. \tag{8}$$

Note that this definition leads to $\mathbf{1}^\top \bar{\mathbf{q}} = 1$ and 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*

$$w_{jk} + f(\bar{q}_j) - f(\bar{q}_k) \tag{9}$$

is nonnegative *and* the origin node j has at least one supply unit (see Figure 1 for illustration of the score). The score (9) 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; see Figures 1 and 2. The policy is not only completely blind, but also semi-local, i.e., it only uses the queue lengths at the origin and destination. Note that the congestion cost (7) increases with queue length (as required), and furthermore decreases

[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 $\mathbf{f}(\bar{\mathbf{q}}) \triangleq [f_1(\bar{q}_1), \cdots, f_m(\bar{q}_m)]^\top$ for some monotone increasing (f_j)s. Here it suffices to consider a single congestion function $f(\cdot)$, whereas in Section 6 we will employ queue-specific congestion functions $f_j(\cdot)$.

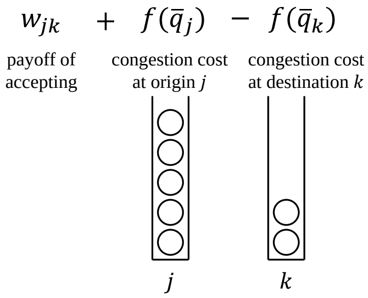

Figure 1: The score (9); MBP admits a demand unit only if the score is non-nonnegative,

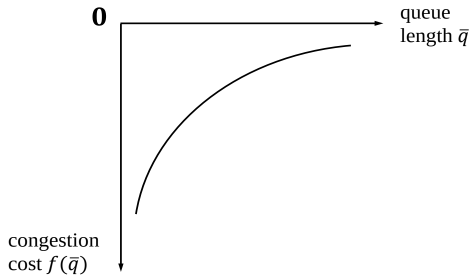

Figure 2: An example of a congestion function (a mapping from queue lengths to congestion costs) which aggressively protects supply units in near-empty queues.

sharply as queue length approaches zero. Observe that such a choice of congestion function makes 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 $\bar{\mathbf{q}}[t+1] = \bar{\mathbf{q}}[t] - \frac{1}{K} x_{jk}[t](\mathbf{e}_j - \mathbf{e}_k)$.

3.2 Performance Guarantee for 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.

Theorem 1. *Consider a set of m nodes and any demand arrival rates $\boldsymbol{\phi}$ that satisfy Condition 1. Then there exists $K_1 = \mathrm{poly}(m, \frac{1}{\alpha(\boldsymbol{\phi})})$, and a universal constant $C < \infty$, such that the following holds.*[7] *For the congestion function $f(\cdot)$ defined in (7), for any $K \geq K_1$, the following guarantees hold for Algorithm 1*

$$L_T^{\mathrm{MBP}} \leq M_1 \cdot \frac{K}{T} + M_2 \cdot \frac{1}{K}\,, \qquad \text{and} \qquad L^{\mathrm{MBP}} \leq M_2 \cdot \frac{1}{K}\,, \qquad \text{for } M_1 \triangleq Cm \text{ and } M_2 \triangleq Cm^2\,.$$

[7] Here "poly" indicates a polynomial. The constant C is universal in the sense that it does not depend on K, m or $\alpha(\boldsymbol{\phi})$.

Remark 2. *In Section 6 we obtain results similar to Theorem 1 in broader settings that allow pricing and flexible assignment (Theorem 2, 3), and moreover allow for time-varying demand arrival rates in Section 6.2. In Appendix C (Theorem 4), we generalize Theorem 1 by showing similar performance guarantees for a whole class of congestion functions that satisfy certain growth conditions. Informally, the congestion function needs to be steep enough near zero to protect the nodes from being drained of supply units. 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* $c_0 = \mathrm{poly}(m, \frac{1}{\alpha(\boldsymbol{\phi})})$*, the same guarantee as in Theorem 1 holds with* $K_1 = \mathrm{poly}(c, m, \frac{1}{\alpha(\boldsymbol{\phi})})$*,* $M_1 = \mathrm{poly}(c, m)$*,* $M_2 = \mathrm{poly}(c, m)$*. However, the specific polynomials depend on the choice of congestion function.*

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.** In practice, the platform operator at best 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, Freund, and Lykouris (2016) even though MBP requires *no* knowledge of $\boldsymbol{\phi}$, whereas the policy of Banerjee, Freund, and Lykouris (2016) requires *exact* knowledge of $\boldsymbol{\phi}$: As shown in Banerjee, Kanoria, and Qian (2018, Proposition 4), if the estimate of demand arrival rates is imperfect, any state independent policy (such as that of Banerjee, Freund, and Lykouris 2016) generically suffers a long run (steady state) per customer optimality gap of $\Omega(1)$ (as $K \to \infty$). Note that the greedy policy (which admits a demand whenever a supply unit is available) also typically suffers a steady state per period optimality gap of $\Omega(1)$; see Example 1 in Section 2.

(2) **Guarantee on transient performance.** In contrast with Banerjee, Freund, and Lykouris (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 in physical time, and where a non-trivial fraction of cars are busy at any time, see, e.g., Braverman et al. 2019). Let T^{real} denote the horizon in physical time. As K increases in the large market regime, the primitive $\boldsymbol{\phi}$ remains unchanged, while $T = \Theta(K \cdot T^{\text{real}})$ since there are $\Theta(K)$ arrivals per unit of physical time. Hence, we can rewrite our performance guarantee as

$$W_T^* - W_T^{\text{MBP}} \leq M \left(\frac{1}{T^{\text{real}}} + \frac{1}{K} \right) \xrightarrow{K \to \infty} \frac{M}{T^{\text{real}}}.$$

Our bound on the optimality gap per customer in steady state is M_2/K, matching that of Banerjee, Freund, and Lykouris (2016) in its scaling with K. (However, our constant M_2 is quadratic in the number of nodes m, whereas the constant in the other paper is linear in m.)

(3) **Flexibility in the choice of congestion function.** Because of the richness of the class of congestion functions covered in Appendix C which generalizes Theorem 1, the system controller now has the additional flexibility to choose a suitable congestion function $f(\cdot)$. For example, in our setting the performance guarantee for the congestion function given in (7) (Theorem 1) is more attractive than that for the linear congestion function $f(\bar{q}) = c \cdot \bar{q}$ (Remark 2) 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. (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 (7).) From a practical perspective, this flexibility can allow significant performance gains to be unlocked by making an appropriate choice of $f(\cdot)$, as evidenced by our numerical experiments in Section 7.1 and Appendix D.

4 The MBP Policies and Mirror Descent

In this section, we describe the main intuition behind the success of MBP policies, namely, that they execute (dual) mirror descent 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.

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:

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

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

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

One interprets x_{jk} as the fraction of type (j,k) demand which is accepted, and the objective (10) as the rate at which 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^{SPP} be the optimal value of SPP. The following proposition formalizes that, as is typical in such settings, W^{SPP} is an upper bound on the optimal steady state (per customer) payoff W^*. It further establishes that the optimal finite horizon per customer payoff W_T^* cannot be much larger than W^{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) to accommodate the fact that flow balance need not be exactly satisfied over a finite horizon.

4.2 MBP Executes Dual Stochastic Mirror Descent on SPP

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

$$\text{minimize}_{\mathbf{y}}\ g(\mathbf{y})\,,$$

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

where $(x)^+ \triangleq \max\{0, x\}$. Here $\mathbf{y}$ are the dual variables corresponding to the flow balance constraints (11), and have the interpretation of "congestion costs" (Neely 2010), 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 describe the interpretation of BP as stochastic gradient descent, and the interpretation of MBP as stochastic mirror descent, on problem (14).

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 by Tassiulas and Ephremides (1992), see also, e.g., 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 (8), and some $c > 0$ and greedily maximizes 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), when the current state is in the *interior* of the state space, i.e., when $q_j > 0$ for all $j \in V$ (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 the definition of normalized queue length (8)) and second equality is just the expectation of 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 $\mathbf{y}[t]$.

Eq. (16) shows that the evolution of $\mathbf{y}[t]$ when $\mathbf{q}[t] > 0$ 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, approximate minimization of delay/workload, and approximate 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* (8) *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

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

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 $f(\bar{q}_j)$, 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 via an appropriate choice of mirror map (see, e.g., Nemirovsky and Yudin 1983, Beck and Teboulle 2003). 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 provably 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 2 (BP is far from optimal if c is not large enough)**.** *Consider the network introduced in Example 1. Suppose the platform employs backpressure where the shadow prices are taken to be proportional to (normalized) queue lengths* $\mathbf{y}[t] = c \cdot \bar{\mathbf{q}}[t]$ *with* $c < \frac{3}{2}w$.

Let $\mathbf{y}^*$ *be the optimal dual variables in* (14)*. By complementary slackness we have that the set of dual optima are* $\mathbf{y}^*$ *which satisfy*

$$\frac{w}{2} + y_1^* - y_2^* \geq 0\,, \quad \frac{w}{2} + y_2^* - y_1^* = 0\,, \quad w + y_2^* - y_3^* = 0\,, \quad \frac{w}{2} + y_3^* - y_2^* \geq 0\,.$$

Hence $\mathbf{y}^*$ *takes the form* $\mathbf{y}^* = (y_1^*, y_1^* - \frac{w}{2}, y_1^* + \frac{w}{2})$ *for arbitrary* $y_1^* \in \mathbb{R}$*. Let* $\bar{\mathbf{q}}^* \triangleq \mathbf{y}^*/c$ *be the queue* *lengths corresponding to the optimal dual variables in* (14) *with the additional constraint that the normalized queue lengths sum to* 1*. Simple algebra yields* $\bar{\mathbf{q}}^* = (\frac{1}{3}, \frac{2c-3w}{6c}, \frac{2c+3w}{6c})$*. Because* $c < \frac{3}{2}w$ *we have* $\bar{q}_2^* < 0$*, and so* $\bar{\mathbf{q}}^*$ *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 3 (BP has positive Lyapunov drift at a certain state)**.** *Again consider Example 1 and let* $c \geq \frac{3}{2}w$*. A typical 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 $\|\bar{\mathbf{q}}[t] - \bar{\mathbf{q}}^*\|_2 = \Omega(1)$*. Suppose at time* t *we have*[8] $\bar{\mathbf{q}}[t] = (\frac{2}{3}, 0, \frac{1}{3})$*; in particular, queue 2 is empty. Note that at* $\bar{\mathbf{q}}[t]$*, BP can only fulfill the demand going from* 1 *to* 2 *and from* 3 *to* 2 *because of the no-underflow constraint. Straightforward calculation shows that the "drift" is positive for large enough* K *if* $\epsilon < \frac{w}{2c+3w}$.

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 (7) 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 carrying formal guarantees, the MBP policy also achieves better performance than BP in simulations (Section 7.1 and Appendix D).

5 Proof of Theorem 1

In this section we provide the key lemmas that lead to a proof of Theorem 1. Our analysis generalizes and refines the so-called Lyapunov drift method in the network control literature (see, e.g., Neely 2010). It 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 (Lemmas 2 and 3).

[8] 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.

(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.

We use the antiderivative of $\mathbf{f}(\cdot)$ as our Lyapunov function; for the congestion function f in (7), this is

$$F(\bar{\mathbf{q}}) \triangleq -2\sqrt{m}\sum_{j\in V}\sqrt{\bar{q}_j}\,. \tag{18}$$

Motivation for our choice of Lyapunov function. We utilize our key observation that MBP executes mirror descent on the dual of SPP (see Section 4.2) to find a suitable (uncentered) Lyapunov function. The standard proof of convergence of mirror descent uses the Bregman divergence $B_F(\bar{\mathbf{q}},\bar{\mathbf{q}}^*)$, generated by the antiderivative $F(\cdot)$ of the inverse mirror map, as the Lyapunov function (note that $B_F(\bar{\mathbf{q}},\bar{\mathbf{q}}^*)$ is a "centered" function in that it achieves its minimum at $\bar{\mathbf{q}}^*$; this function generalizes the centered quadratic function used to analyze stochastic gradient descent). We use the "uncentered" version of the Bregman divergence, which is nothing but F itself, as our Lyapunov function; this choice turns out to be natural for studying the time-averaged performance (rather than convergence of the last iterate). Since the congestion function corresponds to the inverse mirror map, our F is simply the antiderivative of the congestion function.[9]

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 analysis combines (a modification of) the standard approach for stochastic mirror descent algorithms (see, e.g., Nemirovsky and Yudin 1983, Beck and Teboulle 2003) with a novel argument that bounds the suboptimality contributed by underflow.

Recall that W^{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 (proved in Appendix B):

Lemma 1 (Suboptimality of MBP in one period)**.** *Consider congestion functions $f(\cdot)$s that are*

[9] An alternate viewpoint is that our setting and policy fit into the "drift-plus-penalty" framework in the network control literature (Neely 2010), with the Lyapunov function which is the antiderivative of the congestion functions. Previous work focuses on the quadratic Lyapunov function.

strictly increasing and continuously differentiable. We have the following decomposition:

$$W^{\text{SPP}} - \mathbb{E}[v^{\text{MBP}}[t]|\bar{\mathbf{q}}[t]] \leq \underbrace{\tilde{K}\left(F(\bar{\mathbf{q}}[t]) - \mathbb{E}[F(\bar{\mathbf{q}}[t+1])|\bar{\mathbf{q}}[t]]\right)}_{\mathcal{V}_1} + \underbrace{\frac{1}{2\tilde{K}} \cdot \max_{j \in V} \left|f'(\bar{q}_j[t])\right|}_{\mathcal{V}_2}$$

$$+ \underbrace{\left(W^{\text{SPP}} - g(\mathbf{f}(\bar{\mathbf{q}}[t]))\right)}_{\mathcal{V}_3} + \underbrace{\mathbb{1}\left\{q_j[t] = 0, \exists j \in V\right\}}_{\mathcal{V}_4}. \tag{19}$$

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

In the next subsection, we outline our novel analysis showing that the sum of the last three terms $\mathcal{V}_2 + \mathcal{V}_3 + \mathcal{V}_4$ is small. As a result, $\mathcal{V}_1$ 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, and this cannot be the case on average since the Lyapunov function value must remain bounded.

5.2 Bounding Single Period Payoff Loss

In this section we proceed to upper bound $\mathcal{V}_2+\mathcal{V}_3+\mathcal{V}_4$ on the RHS of (19). Observe that the terms $\mathcal{V}_2$ and $\mathcal{V}_4$ are non-negative, while $\mathcal{V}_3$ is non-positive, thus the goal is to show that $\mathcal{V}_3$ compensates for $\mathcal{V}_2+\mathcal{V}_4$. First notice that $\mathcal{V}_2$ is large when there exist very short queues (because the congestion function (7) changes rapidly only for short queue lengths), and $\mathcal{V}_4$ is non-zero only when some queues are empty. Helpfully, it turns out that $\mathcal{V}_3$ 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 $\mathcal{V}_3$ that becomes more negative as the shortest queue length decreases.

Lemma 2. *Consider congestion functions $f(\cdot)$s that are strictly increasing and continuously differentiable, and any $\boldsymbol{\phi}$ with connectedness $\alpha(\boldsymbol{\phi}) > 0$. We have*

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

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

The following lemma bounds $\mathcal{V}_2 + \mathcal{V}_3 + \mathcal{V}_4$. The proof is in Appendix C. (In fact we prove a general version of the lemma which applies to all congestion functions that satisfy certain growth conditions formalized in Condition 4 in Appendix C. The growth conditions serve to ensure that $\mathcal{V}_3$ compensates for $\mathcal{V}_2 + \mathcal{V}_4$.)

Lemma 3. *Consider the congestion function* (7)*, 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* $K \geq K_1$*,*

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

for $M_2 = Cm^2$*, where* $C > 0$ *is a universal constant (which does not depend on* K*,* m *or* $\alpha(\boldsymbol{\phi})$*). Here* $\tilde{K}$ *was defined in* (8)*.*

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 $\mathcal{V}_1$ form a telescoping sum.

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

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

Take the sum of both sides of the inequality (55) from $t = 0$ to $t = T - 1$, and divide the sum by T. This yields

$$W^{\mathrm{SPP}} - W_T^{\mathrm{MBP}} \leq \frac{\tilde{K}}{T}\left(\mathbb{E}[F(\bar{\mathbf{q}}[0])] - \mathbb{E}[F(\bar{\mathbf{q}}[T])]\right) + M_2\frac{1}{\tilde{K}} \qquad \text{for } K \geq K_1\,.$$

Using Proposition 1 and the inequality above, we have

$$\begin{aligned}
L_T^{\mathrm{MBP}} = W_T^* - W_T^{\mathrm{MBP}} &\leq W^{\mathrm{SPP}} + m\frac{K}{T} - W_T^{\mathrm{MBP}} \\
&\leq \frac{\tilde{K}}{T}\left(m + \mathbb{E}[F(\bar{\mathbf{q}}[0])] - \mathbb{E}[F(\bar{\mathbf{q}}[T])]\right) + M_2\frac{1}{\tilde{K}} \\
&\leq \frac{\tilde{K}}{T}\left(m + \sup_{\bar{\mathbf{q}}_1, \bar{\mathbf{q}}_2 \in \Omega}\left(F(\bar{\mathbf{q}}_1) - F(\bar{\mathbf{q}}_2)\right)\right) + M_2\frac{1}{\tilde{K}}\,,
\end{aligned}$$

Let $M_1 \triangleq m + \sup_{\bar{\mathbf{q}}_1, \bar{\mathbf{q}}_2 \in \Omega}\left(F(\bar{\mathbf{q}}_1) - F(\bar{\mathbf{q}}_2)\right)$. 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

$$M_1 \leq m - \inf_{\bar{\mathbf{q}} \in \Omega} F(\bar{\mathbf{q}}) \leq m - F\left(\frac{1}{m}\mathbf{1}\right) = 3m\,.$$

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 $M_1 = Cm$, $M_2 = Cm^2$ and C does not depend on m, K, or $\alpha(\boldsymbol{\phi})$. Moreover, $\tilde{K} = K + m\sqrt{K} \in [K, 2K]$ taking $K_1 \geq m^2$. This concludes the proof. $\square$

6 Generalizations and Extensions

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 queues at a subset of nodes $V_{\mathrm{b}} \subset V$ have a finite buffer constraint. For $j \in V_{\mathrm{b}}$, denote the buffer size by $d_j = \bar{d}_j K$ for some scaled buffer size $\bar{d}_j \in (0, 1)$. (If $\bar{d}_j \geq 1$, the buffer size exceeds the number of supply units $d_j \geq K$ and there is no constraint as a result, i.e., $j \notin V_{\mathrm{b}}$.) We will find it convenient to define $\bar{d}_j = 1$ for each $j \in V \backslash V_{\mathrm{b}}$. To avoid the infeasible case where the buffers are too small to accommodate all supply units, we assume that $\sum_{j\in V} \bar{d}_j > 1$. Throughout Section 6, the normalized state space will be

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

Similar to the case of entry control, we need to keep $\bar{\mathbf{q}}$ in the interior of Ω, 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 $\mathbf{q}$. When $\bar{d}_j = 1$ for all $j \in V$, the definition of $\bar{q}_j$ in (21) reduces to the one in (8). The congestion functions $(f_j(\cdot))_{j\in V}$ are monotone increasing functions that map (normalized) queue lengths to congestion costs. Here we will state our main results for the congestion function vector

$$f_j(\bar{q}_j) \triangleq \begin{cases} \sqrt{m}\cdot C_b \cdot \left(\left(1-\frac{\bar{q}_j}{\bar{d}_j}\right)^{-\frac{1}{2}} - \left(\frac{\bar{q}_j}{\bar{d}_j}\right)^{-\frac{1}{2}} - D_b\right), & \forall j \in V_{\mathrm{b}}\,, \\ -\sqrt{m}\cdot \bar{q}_j^{\,-\frac{1}{2}} & \forall j \in V\backslash V_{\mathrm{b}}\,. \end{cases} \tag{22}$$

Here C_b and D_b are normalizing constants[10] chosen to ensure that (i) for all $j, k \in V$, we have that $f_j(\bar{q}_j) = f_k(\bar{q}_k)$ when both queues are empty $q_j = q_k = 0$; (ii) for all $j, k \in V_{\mathrm{b}}$, we have that $f_j(\bar{q}_j) = f_k(\bar{q}_k)$ when both queues are full $q_j = d_j$, $q_k = d_k$. (We state the results for other choices of congestion functions in Appendix C.)

Note that $f_j(\cdot)$ in (22) is identical to $f(\cdot)$ in (7) for $j \notin V_b$, i.e., (22) is a generalization of (7) to the case where some queues have buffer constraints. The intuitive reason (22) is a suitable congestion function is that it enables 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, and furthermore allowing demand arrival rates to vary in time. Formally, instead of an origin node and a destination node, in this setting each demand unit has an abstract *type* $\tau \in \mathcal{T}$, and the type for the demand unit in period t is drawn from distribution $\boldsymbol{\phi}^t = (\phi^t_\tau)_{\tau\in\mathcal{T}}$, independently across t. The demand type at period t is denoted by $\tau[t]$. Each demand type $\tau \in \mathcal{T}$ has a pick-up neighborhood $\mathcal{P}(\tau) \subset V, \mathcal{P}(\tau) \neq \emptyset$ and drop-off neighborhood $\mathcal{D}(\tau) \subset V, \mathcal{D}(\tau) \neq \emptyset$. The sets $(\mathcal{P}(\tau))_{\tau\in V}$ and $(\mathcal{D}(\tau))_{\tau\in V}$ are model primitives. (In shared transportation systems, each demand type τ may correspond to an (origin, destination) pair in V^2, with $\mathcal{P}(\tau)$ being nodes close to the origin and $\mathcal{D}(\tau)$ being nodes close to the destination. In the special case that $\mathcal{P}(\tau)$ and $\mathcal{D}(\tau)$ are singletons for each $\tau \in \mathcal{T}$ we recover the illustrative model in Section 2.)

The platform control and payoff in this setting are as follows. At time t, after observing the demand type $\tau[t] = \tau$, the system makes a decision

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

Here $x_{j\tau k}[t] = 1$ stands for the platform choosing pick-up node $j \in \mathcal{P}(\tau)$ and drop-off node $k \in \mathcal{D}(\tau)$, causing a supply unit to be relocated from j to k. The constraint in (23) captures that each demand unit is either served by one supply unit, or not served. With $x_{j\tau k}[t] = 1$, the system collects payoff

[10]Define $\epsilon \triangleq \frac{\delta_K}{K}$. Let $h_{\mathrm{b}}(\bar{q}) \triangleq (1-\bar{q})^{-\frac{1}{2}} - \bar{q}^{-\frac{1}{2}}$ and $h(\bar{q}) \triangleq -\bar{q}^{-\frac{1}{2}}$. Define $C_b \triangleq \frac{h(\epsilon)-h(1/\sum_{j\in V}\bar{d}_j)}{h_{\mathrm{b}}(\epsilon)-h_{\mathrm{b}}(1/\sum_{j\in V}\bar{d}_j)}$ and $D_b \triangleq h_{\mathrm{b}}(1/\sum_{j\in V}\bar{d}_j) - C_b^{-1}h(1/\sum_{j\in V}\bar{d}_j)$. In addition to the properties listed in the main text, we also have that $f_j(\bar{d}_j/\sum_{j\in V}\bar{d}_j)$ has the same value for all $j \in V$. These properties are useful in the following analysis.

$v[t] = w_{j\tau k}$. Without loss of generality, we assume the scaling

$$\max_{\tau\in\mathcal{T},j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} |w_{j\tau k}| = 1 \,.$$

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

$$x_{j\tau k}[t] = 0 \quad \text{if} \quad q_j[t] = 0 \text{ or } q_k[t] = d_k \,.$$

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

$$\mathbf{q}[t+1] = \mathbf{q}[t] + \sum_{\tau\in\mathcal{T},j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} (-\mathbf{e}_j + \mathbf{e}_k) x_{j\tau k}[t] \,. \tag{24}$$

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^π and L^π as in Section 2 via (3)-(5).

The dual problem to the SPP in period t in the JEA setting (see Appendix A.1.1 for the SPP, which we denote by SPP^t) is

$$\text{minimize}_{\mathbf{y}}\; g^t_{\mathrm{JEA}}(\mathbf{y}), \quad \text{for } g^t_{\mathrm{JEA}}(\mathbf{y}) \triangleq \sum_{\tau\in\mathcal{T}} \phi^t_\tau \max_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} \left(w_{j\tau k} + y_j - y_k\right)^+ \,. \tag{25}$$

As before, MBP is defined to achieve the argmax in the definition of the dual objective g_{JEA}, with the ys replaced by congestion costs: (i) Again, decisions are made based on payoffs adjusted by congestion costs, and demand units which generate (weakly) positive adjusted payoff are admitted. (ii) The pickup and dropoff locations are chosen to maximize the adjusted payoff.

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

At the start of period t, the system observes demand type $\tau[t] = \tau$.

$(j^*, k^*) \leftarrow \mathrm{argmax}_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} w_{j\tau k} + f_j(\bar{q}_j[t]) - f_k(\bar{q}_k[t])$

if $w_{j^*\tau k^*} + f_{j^*}(\bar{q}_{j^*}[t]) - f_{k^*}(\bar{q}_{k^*}[t]) \geq 0$ **and** $q_{j^*}[t] > 0,\ q_{k^*}[t] < d_{k^*}$ **then**

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

else

  $x_{j^*\tau k^*}[t] \leftarrow 0$, i.e., drop the incoming demand;

end

The queue lengths update as $\bar{\mathbf{q}}[t+1] = \bar{\mathbf{q}}[t] - \frac{1}{K} x_{j^*\tau k^*}[t](\mathbf{e}_{j^*} - \mathbf{e}_{k^*})$.

We make the following connectivity assumption on the primitives $(\boldsymbol{\phi}^t, \mathcal{P}, \mathcal{D})$ for all t in the horizon.

Condition 2 (Strong Connectivity of $(\boldsymbol{\phi}^t, \mathcal{P}, \mathcal{D})$)**.** *For any demand arrival rates $\boldsymbol{\phi}$, define the* *connectedness of triple* ($\boldsymbol{\phi}$,$\mathcal{P}$,$\mathcal{D}$) *as*

$$\alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D}) \triangleq \min_{S \subsetneq V, S \neq \emptyset} \sum_{\tau \in \mathcal{P}^{-1}(S) \cap \mathcal{D}^{-1}(V \backslash S)} \phi_\tau \,. \tag{26}$$

Here $\mathcal{P}^{-1}(S) \triangleq \{\tau \in \mathcal{T} : \mathcal{P}(\tau) \cap S \neq \emptyset\}$ *is the set of demand types for which nodes* S *can serve as a pickup node; and* $\mathcal{D}^{-1}(\cdot)$ *is defined similarly. We assume that for some* $\alpha_{\min} > 0$*, for all* t *in the horizon it holds that* ($\boldsymbol{\phi}^t$,$\mathcal{P}$,$\mathcal{D}$) *is* $\alpha_{\min}$*-strongly connected, namely,* $\alpha(\boldsymbol{\phi}^t, \mathcal{P}, \mathcal{D}) \geq \alpha_{\min}$.

If each type $\tau \in \mathcal{T}$ corresponds to an origin-destination pair $\tau = (j,k) \in V^2$ and $\mathcal{P}(\tau) = \{j\}$, $\mathcal{D}(\tau) = \{k\}$ and demand arrival rates are stationary $\boldsymbol{\phi}^t = \boldsymbol{\phi}$, 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 (6).

Definition 1. *We say that demand arrival rates vary* η*-slowly for some* $\eta \geq 0$ *if* $\|\boldsymbol{\phi}^{t+1} - \boldsymbol{\phi}^t\|_1 \leq \eta$ *for all* $t \geq 0$ *in the horizon of interest.*

Note that any sequence of demand arrival rates varies 2-slowly, so $\eta \in [0,2]$, with $\eta = 0$ being the case of stationary demand arrival rates.

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

Theorem 2. *Fix a set* V *of* $m \triangleq |V| > 1$ *nodes, a subset* $V_{\mathrm{b}} \subseteq V$ *of buffer-constrained nodes with scaled buffer sizes* $\bar{d}_j \in (0,1)$ $\forall j \in V_{\mathrm{b}}$ *satisfying*[11] $\sum_{j \in V} \bar{d}_j > 1$*, and a minimum connectivity* $\alpha_{\min} > 0$*. Then there exists* $K_1 = \mathrm{poly}\left(m, \bar{\mathbf{d}}, \frac{1}{\alpha_{\min}}\right)$, $M_1 = Cm$*, and* $M_2 = C\frac{\sqrt{m}}{\min_{j \in V} \bar{d}_j}\left(\frac{\sum_{j \in V} \bar{d}_j}{\min\{\sum_{j \in V} \bar{d}_j - 1, 1\}}\right)^{3/2}$ *where* C *is a universal constant that does not depend on* m, $\bar{\mathbf{d}}$, η *or* $\alpha_{\min}$*, such that for the congestion functions* $(f_j(\cdot))_{j \in V}$ *defined in* (22)*, the following guarantee holds for Algorithm 2. For any horizon* T*, any* $K \geq K_1$*, and any sequence of demand arrival rates* $(\boldsymbol{\phi}^t)_{t=0}^{T-1}$ *which varies* η*-slowly (for some* $\eta \in [0,2]$*) and pickup and dropoff neighborhoods* $\mathcal{P}$ *and* $\mathcal{D}$ *such that* $(\boldsymbol{\phi}^t, \mathcal{P}, \mathcal{D})$ *is* $\alpha_{\min}$*-strongly connected for all* $t \leq T-1$*, we have*

$$L_T^{\mathrm{MBP}} \leq M_1 \cdot \left(\frac{K}{T} + \sqrt{\eta K}\right) + M_2 \cdot \frac{1}{K} \,.$$

In Appendix C we prove a general version of Theorem 2 which provides a performance guarantee 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

[11] Recall that we define $\bar{d}_j \triangleq 1$ for all $j \in V \backslash V_{\mathrm{b}}$.

pair, and decide an assignment at each period to maximize payoff. Our model here will be similar to that of Banerjee, Freund, and Lykouris (2016), except that the platform does *not* know demand arrival rates, and we allow a finite horizon. The proposed algorithm will be a generalization of backpressure based joint-rate-scheduling control policies (see, e.g., Lin and Shroff 2004, Eryilmaz and Srikant 2007). The demand types τ, pick-up neighborhood $\mathcal{P}(\tau)$ and drop-off neighborhood $\mathcal{D}(\tau)$ are defined in the same way as in section 6.2. For simplicity, we assume that the demand type distribution $\boldsymbol{\phi} = (\phi_\tau)_{\tau\in\mathcal{T}}$ is time invariant in this subsection.

The platform control and payoff in this setting are as follows. At time t, after observing the demand type $\tau[t] = \tau$, the system chooses a *price* $p_\tau[t] \in [p_\tau^{\min}, p_\tau^{\max}]$ and a decision

$$(x_{j\tau k}[t])_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} \in \{0,1\}^{|\mathcal{P}(\tau)|\cdot|\mathcal{D}(\tau)|} \quad \text{such that} \quad \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} x_{j\tau k}[t] \le 1\,. \tag{27}$$

As before we require

$$x_{j\tau k}[t] = 0 \quad \text{if} \quad q_j[t] = 0 \text{ or } q_k[t] = d_k\,.$$

The result of the platform control is as follows:

(1) Upon seeing the price, the arriving demand unit will decline (to buy) with probability $F_\tau(p_\tau[t])$, where $F_\tau(\cdot)$ is the cumulative distribution function of type τ demand's willingness-to-pay.

(2) If the demand accepts (i.e., buys), the system state updates as per

$$\mathbf{q}[t+1] = \mathbf{q}[t] + \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} (-\mathbf{e}_j + \mathbf{e}_k)x_{j\tau k}[t]\,. \tag{28}$$

Meanwhile, the platform collects payoff $v[t] = p_\tau[t] - c_{j\tau k}$ where $c_{j\tau k}$ is the "cost" of serving a demand unit of type τ using pick-up node j and drop-off node k.

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

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

Condition 3. *(1) Assume*[12] $F_\tau(p_\tau^{\min}) = 0$ *and that* $F_\tau(p_\tau^{\max}) = 1$.

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

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

(3) The revenue functions $r_\tau(\mu_\tau) \triangleq \mu_\tau \cdot p_\tau(\mu_\tau)$ are concave and twice continuously differentiable, where μ_τ denotes the fraction of demand of type τ which is realized (i.e., willing to pay the price offered).

As a consequence of Condition 3 parts 1 and 2, the willingness to pay distribution $F_\tau(\cdot)$ has an inverse denoted as $p_\tau(\mu_\tau) : [0,1] \to [p_\tau^{\min}, p_\tau^{\max}]$ which gives the price which will cause any desired fraction $\mu_\tau \in [0,1]$ of demand to be realized. (The concavity assumption in part 3 of the condition is stated in terms of this function $p_\tau(\cdot)$.) Without loss of generality, let $\max_{\tau\in\mathcal{T}} p_\tau^{\max} + \max_{j,k\in V,\tau\in\mathcal{T}}|c_{j\tau k}|= 1$.

In the JPA setting, the net demand $\phi_\tau\mu_\tau$ 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 should modulate the quantity of demand through changing the prices (and serving all the demand which is then realized) rather than apply entry control (i.e., dropping some demand proactively). Our MBP policy for this setting will have this feature.

The dual problem to the SPP in the JPA setting (the SPP is stated in Appendix A.2) is[13]

$$\text{minimize}_{\mathbf{y}}\, g_{\text{JPA}}(\mathbf{y}) \quad \text{for } g_{\text{JPA}}(\mathbf{y}) \triangleq \sum_{\tau\in\mathcal{T}} \phi_\tau \max_{\{0\le\mu_\tau\le 1\}} \left(r_\tau(\mu_\tau) + \mu_\tau \max_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} \left(-c_{j\tau k} + y_j - y_k\right)\right). \tag{29}$$

Once again, the MBP policy (Algorithm 3 below) is defined to achieve the argmaxes in the definition of the dual objective $g_{\text{JPA}}(\cdot)$ with the ys replaced by congestion costs: MBP dynamically sets prices p_τ such that mean fraction of demand realized under the policy is the outer argmax in the definition (29) of $g_{\text{JPA}}(\cdot)$, and the assignment decision of MBP achieves the inner argmax in the definition (29) of $g_{\text{JPA}}(\cdot)$. The policy again has the property that it executes stochastic mirror descent on the dual objective $g_{\text{JPA}}(\cdot)$.

The MBP policy retains the advantage that it does not require any prior knowledge of gross demand $\boldsymbol{\phi}$. We assume that the willingness-to-pay distributions $F_\tau(\cdot)$s are exactly known to the platform; it may be possible to relax this assumption via a modified policy which "learns" the $F_\tau(\cdot)$s, however, pursuing this direction is beyond the scope of the present paper.

[13]The derivation of the dual objective is in Appendix B.

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

At the start of period t, the system observes $\tau[t] = \tau$.

$(j^*, k^*) \leftarrow \arg\max_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} \left\{-c_{j\tau k} + f_j(\bar{q}_j[t]) - f_k(\bar{q}_k[t])\right\}$;

if $q_{j^*}[t] > 0$, $q_{k^*}[t] < d_{k^*}$ **then**

  $\mu_\tau[t] \leftarrow \mathrm{argmax}_{\mu_\tau\in[0,1]} \left\{r_\tau(\mu_\tau) + \mu_\tau \cdot (-c_{j^*\tau k^*} + f_{j^*}(\bar{q}_{j^*}[t]) - f_{k^*}(\bar{q}_{k^*}[t]))\right\}$;

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

  $x_{j^*\tau k^*}[t] \leftarrow 1$, i.e., serve the incoming demand (if it stays) by pick up from j^* and drop off at k^*;

else

  $x_{j^*\tau k^*}[t] \leftarrow 0$, i.e., drop the incoming demand;

end

The queue lengths update as $\bar{\mathbf{q}}[t+1] = \bar{\mathbf{q}}[t] - \frac{1}{K} x_{j^*\tau k^*}[t](\mathbf{e}_{j^*} - \mathbf{e}_{k^*})$.

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

(1) The computed prices satisfy $p_\tau[t] \in [p_\tau^{\min}, p_\tau^{\max}]$ (by the observation following Condition 3).

(2) The optimization problem for computing $\mu_\tau[t]$ is a one-dimensional concave maximization problem (Condition 3 part 3), hence $\mu_\tau[t]$ can be efficiently computed.

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

Theorem 3. *Fix a set V of $m = |V| > 1$ nodes, scaled buffer sizes $\bar{\mathbf{d}} = (\bar{d}_j)_{j\in V}$ with*[14] *$\bar{d}_j \in (0, 1]$ and $\sum_{j\in V} \bar{d}_j > 1$, minimum and maximum allowed prices $(p_\tau^{\min}, p_\tau^{\max})_{\tau\in\mathcal{T}}$, any $(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})$ that satisfy Condition 2 (strong connectivity), and willingness-to-pay distributions $(F_\tau)_{\tau\in\mathcal{T}}$ that satisfy Condition 3. Then there exist $K_1 < \infty$, $M_1 = \mathrm{poly}\left(m, \bar{\mathbf{d}}\right)$, and $M_2 = \mathrm{poly}\left(m, \bar{\mathbf{d}}\right)$ such that for the congestion functions $(f_j(\cdot))$s defined in (22), the following guarantee holds for Algorithm 3. For any horizon T and for any $K \geq K_1$, we have*

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

7 Application to Shared Transportation Systems

Our setting can be mapped to shared transportation systems such as bike sharing and ride-hailing systems. In this context, the nodes in our model correspond to geographical locations, while supply units and demand units correspond to vehicles and customers, respectively.

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 dynamically incentivize certain trips by

[14] Recall that we use $\bar{d}_j \triangleq 1$ for nodes $j \in V \backslash V_b$ which do not have a buffer size constraint.

awarding points (with cash value) depending on a trip's pickup and dropoff locations (Chung et al. 2018). Thus the problem of designing a dynamic incentive program is addressed (in a stylized way) by the pricing setting we study (the joint pricing-assignment setting studied Section 6.3, but with no assignment flexibility). MBP tells the system operator, quantitatively, how to reward rides that relocate bikes to locations which have a scarcity of bikes. In docked bike sharing systems, there is a constraint on the number of docks available at each location. Such constraints are seamlessly handled in our framework as detailed earlier in Section 6.1. One concern may be that our model ignores travel delays. However, in most bike sharing systems, the fraction of bikes in transit at any time is typically quite small (under 10-20%).[15] As a result, we expect our control insights to retain their power despite the presence of delays. (Indeed, we will numerically demonstrate in Section 7.1 that this is the case in a realistic ridehailing setting; see the excess supply case where MBP performs well even when the vast majority of supply is in transit at any time.) We leave a detailed study of bike sharing platforms to future work.

Online control of ride-hailing platforms. Ride-hailing platforms make dynamic decisions to optimize their objectives (e.g., revenue, welfare, etc.). For most ride-hailing platforms in North America, pricing is used to modulate demand. 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. In both cases, the platform further decides 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 captured, respectively, by the joint entry-assignment (JEA)[16] and joint pricing-assignment (JPA) models studied in Section 6. A concern may be that travel delays play a significant role in ride-hailing, whereas delays are ignored in our theory. In the following subsection, we summarize a numerical investigation of ride-hailing focusing on entry and assignment controls only (a full description is provided in Appendix D). We find that MBP performs well despite the presence of travel delays. In order to address the case where the available supply is scarce, we heuristically adapt MBP to incorporate the Little's law constraint (Section 7.1.1).

[15]The report https://nacto.org/bike-share-statistics-2017/ tells us that U.S. dock-based systems produced an average of 1.7 rides/bike/day, while dockless bike share systems nationally had an average of about 0.3 rides/bike/day. Average trip duration was 12 minutes for pass holders (subscribers) and 28 mins for casual users. In other words, for most systems, each bike was used less than 1 hour per day, which implies that less than 10% of bikes are in use at any given time during day hours (in fact the utilization is below 20% even during rush hours).

[16]The JEA setting can be mapped to ride-hailing as follows: there is a demand type τ corresponding to each (origin, destination) pair $(j,k) = V^2$, with $\mathcal{P}(\tau)$ being nodes close to the origin j and $\mathcal{D}(\tau)$ being nodes close to the destination k.

7.1 Numerical simulations in a realistic ride-hailing setting

We simulate the MBP policy in a realistic ride-hailing environment using yellow cab data from NYC Taxi & Limousine Commission and travel times from Google Maps. In the interest of space, we provide only a summary of our simulations here and defer a full description to Appendix D.

We allow the platform two control levers: entry control and assignment/dispatch control, similar to the JEA setting[17,18] 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 Appendix D.1 for details on the "fluid requirement") to achieve the value of the static planning problem.

(2) *Scarce supply.* The number of cars fall short (by 25%) of the "fluid requirement", i.e., there are not enough cars to realize the optimal solution of static planning problem (ignoring stochasticity).

Summary of findings. We make a natural modification of the MBP policy (with congestion function (7)) 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. This policy is described below in Section 7.1.1. 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 the 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.

7.1.1 The Supply-Aware MBP Policy

In order to heuristically modify MBP to account for travel times, we begin by observing that the SPP must now include a Little's law constraint. (The same observation was previously leveraged

[17]The correspondence between our (ride-hailing) simulation setting and the JEA setting is as follows: In the ride-hailing setting, the type of a demand is its origin-destination pair, i.e. $\mathcal{T} = V \times V$. For type (j,k) demand, its supply neighborhood is the neighboring locations of j, which we denote by (with a slight abuse of notation) $\mathcal{P}(j)$. We do not consider flexible drop-off, therefore $\mathcal{D}(j,k) = \{k\}$.

[18]In our simulations, we focus on the special case where demand is stationary instead of time-varying, even though MBP policies are expected to work well if demand varies slowly over time. We make this choice because it allows us to compare performance against that of the policy proposed in Banerjee et al. (2016) for the stationary demand setting.

by Braverman et al. 2019 and Banerjee, Freund, and Lykouris 2016 to formally handle travel times, albeit under the assumption that travel times are i.i.d. exponentially distributed.) Our heuristic modification of MBP will maintain an estimate of the shadow price corresponding to the Little's law constraint, and penalize rides appropriately.

Applying Little's Law, if the optimal solution $\mathbf{z}^*$ of the SPP (32)-(34) (see Appendix A.1.1; here we work with the special case where $\boldsymbol{\phi}$ does not depend on t) is realized as the average long run assignment, the mean number of cars which are occupied in picking up or transporting customers is $\sum_{j,k\in V}\sum_{i\in\mathcal{P}(j)} D_{ijk}\cdot z^*_{ijk}$, for $D_{ijk} \triangleq \tilde{D}_{ij} + \hat{D}_{jk}$, where $\tilde{D}_{ij}$ is the pickup time from i to j and $\hat{D}_{jk}$ is the travel time from j to k. We augment the SPP with the additional supply constraint

$$\sum_{j,k\in V}\sum_{i\in\mathcal{P}(j)} D_{ijk}\cdot z_{ijk} \le K\,. \tag{30}$$

We propose and test in the simulation 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:

$$\begin{aligned}
&i^* \leftarrow \arg\max_{i\in\mathcal{P}(j)} \left\{w_{ijk} + f(\bar{q}_i[t]) - f(\bar{q}_k[t]) - v[t]D_{ijk}\right\}\\
&\textbf{If } w_{i^*jk} + f(\bar{q}_{i^*}[t]) - f(\bar{q}_k[t]) - v[t]D_{i^*jk} \ge 0 \textbf{ and } q_{i^*}[t] > 0\,, \text{dispatch from } i^*,\ \textbf{else} \text{ Drop},
\end{aligned}$$

We define the tightened supply constraint as

$$\sum_{j,k\in V}\sum_{i\in\mathcal{P}(j)} D_{ijk}\cdot z_{ijk} \le 0.95K\,, \tag{31}$$

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.[19] Here $v[t]$ is the current estimate of the shadow price for a "tightened" version of supply constraint (30). We use the congestion function $f_j(\bar{q}_j) = \sqrt{m}\cdot\bar{q}_j^{-1/2}$, i.e. the one given in (7), in our numerical simulations. An adaptation here is that the queue lengths are normalized by the estimated number of *free* cars instead of K, which we set as $0.05K$ to be consistent with the "utilization" parameter we choose. We update $v[t]$ as

$$v[t+1] = \left[v[t] + \frac{1}{K}\left(\sum_{j,k\in V}\sum_{i\in\mathcal{P}(j)} D_{ijk}\cdot \mathbb{1}\{(o[t],d[t]) = (j,k), \text{MBP dispatches from } i\} - 0.95K\right)\right]^+ .$$

An iteration of supply-aware MBP is equivalent to executing a (dual) stochastic mirror descent

[19] 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.

step on the supply-aware SPP with objective (32) and constraints (33), (34) and (31).

8 Application to Scrip Systems

In this section, we illustrate the application of our model to scrip systems. A scrip system is a nonmonetary trade economy where agents use scrips (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 admission-and-provider-selection rule: If an agent is running low on scrip balance, should they be allowed to request services? If yes, and if there are several possible providers for a trade, who should be selected for service?

We introduce a natural model of a scrip system 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 the service provider should be. The setting is seen to be a special case of the joint entry-assignment (JEA) setting studied in Section 6; yielding a simple MBP control rule that comes with the guarantee that it asymptotically maximizes social welfare.

8.1 Model of Scrip Systems

We now describe a model of a service exchange (i.e., a scrip system). Consider an economy with a finite number of agents indexed by $j \in V$. There are finitely many types of service types Σ indexed by $\sigma \in \Sigma$. A demand type $\tau = (j, \sigma)$ is specified by the requestor $j \in V$ along with the requested service type $\sigma \in \Sigma$, i.e., the set of demand types $\mathcal{T} \subseteq V \times \Sigma$. If the demand is served, the requestor pays a scrip to the service provider. Accordingly, for each demand type $\tau = (j, \sigma)$, we define the *compatible* set of agents who can serve it as $\mathcal{D}(\tau) \subseteq V \backslash \{j\}$. We again consider a slotted time model, where in each period exactly one service request arises, with demand type drawn i.i.d. from the distribution[20] $\boldsymbol{\phi} = (\phi_\tau)_{\tau \in \mathcal{T}}$. There are a fixed number K of scrips in circulation, distributed among the agents. For each $\tau = (j, \sigma) \in \mathcal{T}$, serving a demand type $\tau = (j, \sigma)$ generates payoff $w_{j\sigma}$.

Observe that our model here is a special case of the JEA setting.[21]

Comparison with the model in Johnson et al. (Johnson et al. 2014). The work Johnson et al. (2014) consider the case where there is only *one* type of service which *all* agents can provide, and requests arrive at the same rate from all agents. One one hand, we significantly generalize their model by considering heterogeneous service types, general compatibility structures, and asymmetric service request arrivals. They obtain an optimal rule for the symmetric fully connected 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 8.2).

8.2 MBP Control Rule and Asymptotic Optimality

Since the model above is a special case of the JEA setting, we immediately obtain an MBP control rule for scrip systems that achieves asymptotic optimality as a special case of Algorithm 2 and Theorem 2. This control rule is specified in Algorithm 4 below. The congestion function $f(\cdot)$ can again be chosen flexibly; we state our formal guarantee for the congestion function in (7). Denote the normalized number of scrips (defined in (8)) in the possession of agent i by $\bar{q}_i$.

ALGORITHM 4: MBP Admission-and-provider-selection rule for scrip systems

At the start of period t, the central planner receives a request from agent j for service type σ, i.e., demand type $\tau = (j, \sigma)$ arises.

if $w_{j\sigma} + f(\bar{q}_j[t]) - \min_{k \in \mathcal{D}(\tau)} f(\bar{q}_k[t]) \geq 0$ **and** $\bar{q}_i[t] > 0$ **then**

  $k^* \leftarrow \mathrm{argmin}_{k \in \mathcal{D}(\tau)} f(\bar{q}_k[t])$,

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

else

  Reject the service request from agent j;

end

Theorem 2 immediately implies the following performance guarantee for Algorithm 4.

Corollary 1. *Consider a set of m agents and any demand type distribution and compatibilities $(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})$ (where $\mathcal{P}$ is identity) that satisfy Condition 2. Then there exists* $K_1 = \mathrm{poly}\left(m, \frac{1}{\alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})}\right)$

[20] Time-varying demand arrival rates can be seamlessly handled since they are permitted in the JEA setting; we work with stationary arrival rates only for the sake of brevity.

[21] This can be seen as follows: For each demand type $\tau \in \mathcal{T}$, the compatible set of service providers $\mathcal{D}(\tau)$ is identified with the "dropoff neighborhood" for τ. The "pickup neighborhood" is a singleton set consisting of the requestor $\mathcal{P}(\tau) = \{j\}$. Finally, for each $k \in \mathcal{D}(\tau)$ we define the payoff $w_{j\tau k} \triangleq w_{j\sigma}$. The primitives $V, \mathcal{P}, \mathcal{D}, \boldsymbol{\phi}$ and $(w_{j\tau k})_{\tau=(j,\sigma)\in\mathcal{T}, k \in \mathcal{D}(\tau)}$ fully specify the JEA setting.

and a universal $C > 0$ that does not depend on m, K or $\alpha(\boldsymbol{\phi}, \mathcal{P}, \mathcal{D})$, such that for the congestion function $f(\cdot)$ defined in (7)*, for any $K \geq K_1$, the following guarantee holds for Algorithm 4*

$$L_T^{\text{MBP}} \leq M_1 \cdot \frac{K}{T} + M_2 \cdot \frac{1}{K}, \qquad \text{and} \qquad L^{\text{MBP}} \leq M_2 \cdot \frac{1}{K}, \qquad \text{for } M_1 \triangleq Cm \text{ and } M_2 \triangleq Cm^2 .$$

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)).[22] When a trade is approved, our policy also chooses the compatible trade partner with the fewest scrips as service provider.

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 sufficiently patient, and (ii) agents benefit from trading, i.e., agents derive more value from receiving service than the cost they incur from providing service.)

9 Discussion

In this paper we considered the payoff maximizing dynamic control of a closed network of resources. We proposed 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, while retaining the simplicity of backpressure. The MBP policy overcomes the challenge stemming from the no-underflow constraint and it does not require any knowledge of demand arrival rates. We proved that MBP achieves good transient performance for demand arrival rates which are stationary or vary slowly over time, losing at most $O\left(\frac{K}{T} + \frac{1}{K} + \sqrt{\eta K}\right)$ payoff per customer, where K is the number of supply units, T is the number of customers over the horizon of interest, and η is the maximum change in demand arrival rates per customer arrival. We considered a variety of

[22]For example, consider a setting with two agents j_1 and j_2. Denote the demand type requested by j_1 as τ_1 (this demand type can be served by j_2) and similarly define τ_2. Under the mild condition $\phi_{\tau_1} \neq \phi_{\tau_2}$, the planner will be forced to decline a positive fraction of requests.

control levers: entry control, assignment control and pricing, and allowed for finite buffer sizes. We discussed the application of our results to the control of shared transporation systems and scrip systems.

One natural question is whether our bounds capture the right scaling of the per customer optimality gap of MBP with K, T and η, relative to the best policy which is given exact demand arrival rates and horizon length T in advance. Consider the joint entry-assignment setting (Section 6.2). It is not hard to construct examples showing that each of the terms in our bound is unavoidable: a $1/K$ optimality gap arises in steady state (under stationary demand arrival rates) for instance in a two-node entry-control-only example where the two demand arrival rates are exactly equal to each other, the K/T term arises because over a finite horizon the flow balance constraints need not be satisfied exactly and MBP does not exploit this flexibility fully, and the $\sqrt{\eta K}$ term arises in examples where demand arrival rates oscillate (with a period of order $\sqrt{K/\eta}$) but MBP does not take full advantage of the flexibility to allow queue lengths to oscillate alongside. We omit these examples in the interest of space.

We point out some interesting directions that emerge from our work:

1. *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 state regret will decay exponentially in K.

Another promising direction is to pursue the viewpoint that there is an MBP policy which (very nearly) maximizes the steady state rate of payoff generation, specifically for the choice of congestion functions $f_j(\cdot)$ that are the discrete derivatives of the relative value function $F(\bar{\mathbf{q}})$ (for the average payoff maximization dynamic programming problem) with respect to $\bar{q}_j$; see Chapter 7.4 of Bertsekas (1995) for background on dynamic programming. Thus, estimates of the relative value function $F(\bar{\mathbf{q}})$ can guide the choice of congestion function.

2. *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, as well as performance under demand arrival rates which vary slowly in time. We conclude with a concrete problem which one may try to address using MBP: 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 near optimal performance for more general matching queue systems.

References

I. Adan and G. Weiss. 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, 2012.

N. Agarwal, I. Ashlagi, E. Azevedo, C. R. Featherstone, and Ö. Karaduman. Market failure in kidney exchange. *American Economic Review*, 109(11):4026–70, 2019.

S. Balseiro, D. B. Brown, and C. Chen. Dynamic pricing of relocating resources in large networks. *Available at SSRN 3313737*, 2019.

S. Banerjee, D. Freund, and T. Lykouris. Pricing and optimization in shared vehicle systems: An approximation framework. *arXiv preprint arXiv:1608.06819*, 2016.

S. Banerjee, Y. Kanoria, and P. Qian. Dynamic assignment control of a closed queueing network under complete resource pooling. *arXiv preprint arXiv:1803.04959*, 2018. Also in ACM SIGMETRICS 2018.

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

D. P. Bertsekas. *Dynamic programming and optimal control*, volume 1. Athena scientific Belmont, MA, 1995.

A. Braverman, J. G. Dai, X. Liu, and L. Ying. Empty-car routing in ridesharing systems. *Operations Research*, 67(5):1437–1452, 2019.

S. Bubeck, M. B. Cohen, Y. T. Lee, J. R. Lee, and A. Mądry. K-server via multiscale entropic regularization. In *Proceedings of the 50th annual ACM SIGACT symposium on theory of computing*, pages 3–16, 2018.

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

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

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

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

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

J. G. Dai and W. Lin. Maximum pressure policies in stochastic processing networks. *Operations Research*, 53(2):197–218, 2005.

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

A. Désir, V. Goyal, Y. Wei, and J. Zhang. Sparse process flexibility designs: is the long chain really optimal? *Operations Research*, 64(2):416–431, 2016.

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

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

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

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

V. Gupta and A. Radovanović. Interior-point-based online stochastic bin packing. *Operations Research*, 2020.

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

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

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

L. Huang and M. J. Neely. Utility optimal scheduling in processing networks. *Performance Evaluation*, 68 (11):1002–1021, 2011.

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

K. Johnson, D. Simchi-Levi, and P. Sun. Analyzing scrip systems. *Operations Research*, 62(3):524–534, 2014.

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

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

H. Ma, F. Fang, and D. C. Parkes. Spatio-temporal pricing for ridesharing platforms. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 583–583, 2019.

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

J. Mairesse and P. Moyal. Stability of the stochastic matching model. *Journal of Applied Probability*, 53(4): 1064–1077, 2016.

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

A. S. Nemirovsky and D. B. Yudin. Problem complexity and method efficiency in optimization. 1983.

E. Ozkan and A. R. Ward. Dynamic matching for real-time ridesharing. 2016.

C. Shi, Y. Wei, and Y. Zhong. Process flexibility for multiperiod production systems. *Operations Research*, 67(5):1300–1320, 2019.

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

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

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

K. T. Talluri and G. J. Van Ryzin. *The theory and practice of revenue management*, volume 68. Springer Science & Business Media, 2006.

L. Tassiulas and A. Ephremides. 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, 1992.

D. P. Williamson. *Network Flow Algorithms*. Cambridge University Press, 2019.

Appendix to "Blind Dynamic Resource Allocation in Closed Networks via Mirror Backpressure"

Organization of the appendix. 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). In the appendix, we will only prove the results for JEA and JPA since entry control is a special case of JEA. For most parts of the proof, the proof of JEA can be easily extended to JPA. For particular lemmas/propositions, the proofs of the JPA setting are more involved. For easier reading, we put analogous results together.

The appendix is organized as follows.

1. In Appendix A we prove Proposition 1, i.e., that the value of SPP is an upper bound of the best achievable per customer payoff. We will prove the counterpart of Proposition 1 for JEA (Proposition 2) and JPA (Proposition 3) settings.

2. In Appendix B, we perform the Lyapunov analysis and analyze the geometry of the dual problem (14), and prove Lemma 1 and Lemma 2. We will prove the counterpart of these lemmas for JEA and JPA settings.

3. In Appendix C we prove Lemma 3. We also prove a general result (Theorem 4), and show that it implies Theorems 1, 2, and 3.

4. In Appendix D, we provide further details of the simulation setting.

A Finite Horizon Payoff Upper Bound: Proof of Proposition 1

In this section, we prove the finite horizon payoff upper bounds for JEA (Proposition 2) and JPA setting (Proposition 3). Proposition 1 is implied by Proposition 2.

A.1 Joint Entry-Assignment Setting

Consider the JEA setting defined in Section 6.2, which allows for flexible assignment and time-varying demand.

We will state and prove a generalization of Proposition 1 (fluid-based upper bound on the payoff) to the JEA setting. Before that we introduce some linear programs which generalize the static planning problem (10)-(12), and establish a lemma relating their values to each other.

A.1.1 Relevant linear programs

Fix a horizon T. We will consider the following linear program at time t, based on the current demand arrival rates $\boldsymbol{\phi}^t$:

$$\mathrm{SPP}^t: \quad \text{maximize}_{\mathbf{z}} \quad \sum_{\tau\in\mathcal{T}, j\in\mathcal{P}(\tau), k\in\mathcal{D}(\tau)} w_{j\tau k}\, z_{j\tau k} \tag{32}$$

$$\text{s.t.} \quad \sum_{\tau\in\mathcal{T}, j\in\mathcal{P}(\tau), k\in\mathcal{D}(\tau)} z_{j\tau k}(\mathbf{e}_j - \mathbf{e}_k) = \mathbf{0} \qquad \text{(flow balance)}\,, \tag{33}$$

$$\sum_{j\in\mathcal{P}(\tau), k\in\mathcal{D}(\tau)} z_{j\tau k} \le \phi^t_\tau\,,\ z_{j\tau k} \ge 0 \quad \forall j,k\in V,\ \tau\in\mathcal{T}\,. \text{ (demand constraint)}\,. \tag{34}$$

The variable $z_{j\tau k}$ can be interpreted as the flow of demand type τ being served by pickup location j and dropoff location k. (Note that our LP formulation here has a cosmetic difference from that in (10)-(12): here we find that it simplifies our analysis to use the *flows* $z_{j\tau k}$ as the LP variables instead of using the *fractions* $x_{j\tau k}$ of demand of type τ served by pickup location j and dropoff location k as the variables. The correspondence is simply $z_{j\tau k} \leftrightarrow \phi^t_\tau x_{j\tau k}$.) We denote the value of SPP^t by W^{SPP^t}.

Define the average demand arrival rates

$$\bar{\boldsymbol{\phi}} \triangleq \frac{1}{T}\sum_{t=0}^{T-1} \boldsymbol{\phi}^t \,. \tag{35}$$

We define an "average" linear program $\overline{\mathrm{SPP}}$ as the linear program given by (32), (33), and the averaged demand constraint

$$\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} z_{j\tau k} \leq \bar{\phi}_\tau \,, \; z_{j\tau k} \geq 0 \quad \forall j,k \in V, \; \tau \in \mathcal{T} \,. \text{ (demand constraint)} \,. \tag{36}$$

We denote the value of $\overline{\mathrm{SPP}}$ by $W^{\overline{\mathrm{SPP}}}$.

Although we will not use this property, note that $W^{\overline{\mathrm{SPP}}} \geq \frac{1}{T}\sum_{t=0}^{T-1} W^{\mathrm{SPP}^t}$ since if $\mathbf{z}^t$ is feasible for SPP^t for each t then $\bar{\mathbf{z}} = \frac{1}{T}\sum_{t=0}^{T-1}\mathbf{z}^t$ is feasible for $\overline{\mathrm{SPP}}$. Rather, we will prove and then leverage the property that $W^{\overline{\mathrm{SPP}}}$ is not much larger than W^{SPP^t} for any $t \leq T-1$ if the demand arrival rates vary slowly with t.

Lemma 4. *Suppose the demand arrival rates vary η-slowly (Definition 1) for some $\eta > 0$. Fix a horizon T. For any $0 \leq t \leq T-1$ we have*

$$W^{\mathrm{SPP}^t} \geq W^{\overline{\mathrm{SPP}}} - \eta T m/2 \,. \tag{37}$$

Proof. Since $\|\boldsymbol{\phi}^{t'+1} - \boldsymbol{\phi}^{t'}\|_1 \leq \eta$ for all t', we know that

$$\|\boldsymbol{\phi}^t - \bar{\boldsymbol{\phi}}\|_1 \leq \eta T/2 \,. \tag{38}$$

Let $\bar{\mathbf{z}}$ be an optimal solution to $\overline{\mathrm{SPP}}$. If $\bar{\mathbf{z}}$ is feasible for SPP^t we are done. Suppose not. Using the standard flow decomposition approach (see, e.g., Williamson 2019, the interested reader can also find the flow decomposition argument in the proof of Lemma 6 below), the flow $\bar{\mathbf{z}}$ can be decomposed into flows along directed cycles, since it satisfies the flow balance constraints (33): directed cycles $\mathcal{C}$ carrying flow $f_\mathcal{C} > 0$ in the decomposition take the form $\mathcal{C} = ((j_1,\tau_1,j_2),(j_2,\tau_2,j_3),\cdots,(j_s,\tau_s,j_{s+1}=j_1))$ where the nodes $j_1, j_2, \ldots, j_s$ are distinct from each other, and for each $r = 1,2,\ldots,s$, there is a flow from j_r to j_{r+1} due to demand type τ_r. We have

$$\bar{z}_{j\tau k} = \sum_{\mathcal{C}\ni(j,\tau,k)} f_\mathcal{C} \qquad \text{for all } \tau\in\mathcal{T}, j\in\mathcal{P}(\tau), k\in\mathcal{D}(\tau) \,. \tag{39}$$

(The number of cycles in the decomposition is bounded above by $\sum_{\tau\in\mathcal{T}}|\mathcal{P}(\tau)||\mathcal{D}(\tau)|$, but our argument will not be affected by the number of cycles. In fact our argument can handle an infinity of demand types by replacing sums with integrals.)

Starting from the flow $\bar{\mathbf{z}}$ and the associated cycle decomposition (39), we reduce the flows $(f_\mathcal{C})$ along the cycles via the following iterative process, in order to obtain $\mathbf{z}^t$ which is feasible for the problem SPP^t:

Consider each demand type $\tau \in \mathcal{T}$ in turn and do the following. Define the (current) arrival rate violation as

$$\delta_\tau \triangleq \left(\sum_{\mathcal{C}} f_{\mathcal{C}} \cdot \mathrm{count}(\mathcal{C}, \tau) \; - \; \phi_\tau^t \right)_+ .$$

where $\mathrm{count}(\mathcal{C}, \tau)$ is the number of times demand type τ appears in cycle $\mathcal{C}$. If $\delta_\tau = 0$ do nothing. If $\delta_\tau > 0$, reduce the flows in cycles containing τ sufficiently that after the reduction $\sum_{\mathcal{C}} f_{\mathcal{C}} \cdot \mathrm{count}(\mathcal{C}, \tau) = \phi_\tau^t$ holds (the reduction can be divided arbitrarily between the different cycles containing τ; subject to the constraints that no cycle-flow should increase and no cycle-flow should go below zero). Note that the payoff loss resulting from this reduction is bounded above by $\delta_\tau m$ since each cycle length is at most m (since no node is repeated in a cycle), the ws are assumed to be bounded by 1, and the total reduction in cycle flows is at most δ_τ.

This simple process maintains the following properties:

- The flow balance constraint (33) is satisfied throughout.
- Cycle-flows are non-increasing during the process. Cycle-flows never drop below zero.
- For all demand types which have already been processed so far, the arrival rate constraint is satisfied. Formally: During the process, denote the current value of the right-hand side of (39) by $z_{j\tau k}$. Then $\sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} z_{j\tau k} = \sum_{\mathcal{C}} f_{\mathcal{C}} \cdot \mathrm{count}(\mathcal{C}, \tau) \leq \phi_\tau^t$ for all demand types τ which have already been processed.

In particular, at the end of the process, we arrive at flows $\mathbf{z}^t$ which are feasible for SPP^t. It remains to show that the payoff lost due to the reduction in flows is bounded by $\eta T m/2$.

Since flows are non-increasing and the initial flows are feasible for $\overline{\mathrm{SPP}}$, we have that $\delta_\tau \leq \left(\bar{\phi}_\tau - \phi_\tau^t\right)_+$ for all $\tau \in \mathcal{T}$. Since the payoff lost while processing demand type τ is bounded above by $\delta_\tau m$ (as we argued above), the total loss in payoff lost is then bounded above by

$$m \sum_{\tau \in \mathcal{T}} \delta_\tau \leq m \sum_{\tau \in \mathcal{T}} \left(\bar{\phi}_\tau - \phi_\tau^t\right)_+ \leq m \|\boldsymbol{\phi}^t - \bar{\boldsymbol{\phi}}\|_1 \leq m\eta T/2 ,$$

where we used (38) in the last inequality. Thus, we have constructed a feasible solution $\mathbf{z}^t$ to SPP^t which achieves payoff at least $W^{\overline{\mathrm{SPP}}} - \eta T m/2$. The lemma follows. □

A.1.2 Upper bound on the payoff

We state below the generalization of Proposition 1 to the JEA setting with time-varying demand arrival rates.

Proposition 2. *For any horizon $T < \infty$, any K and any starting state $\mathbf{q}[0]$, the best achievable finite horizon average payoff W_T^* in the JEA setting is upper bounded as*

$$W_T^* \leq W^{\overline{\mathrm{SPP}}} + m \cdot \frac{K}{T} .$$

Here $W^{\overline{\mathrm{SPP}}}$ is the optimal value of $\overline{\mathrm{SPP}}$ given by (32), (33) *and* (36).

The idea behind Proposition 2 is as follows. As is typical in such settings, $W^{\overline{\mathrm{SPP}}}$ is an upper bound on the payoff if the flow constraints are satisfied in expectation. However, since the flow constraints can be slightly violated in the finite horizon setting under consideration, we obtain an upper bound by slightly relaxing the flow constraint (33) in the SPP to

$$\left| \mathbf{1}_S^\top \left(\sum_{\tau\in\mathcal{T}, j\in\mathcal{P}(\tau), k\in\mathcal{D}(\tau)} z_{j\tau k}(\mathbf{e}_j - \mathbf{e}_k) \right) \right| \le \frac{K}{T} \qquad \forall\, S \subseteq V\,, \tag{40}$$

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 2. The first lemma (Lemma 5) shows that the expected payoff cannot exceed the value of the finite horizon demand-averaged SPP, i.e., the linear program defined by (32), (40) and (36).

Lemma 5. *For any horizon $T < \infty$, any K and any starting state $\mathbf{q}[0]$, the expected payoff generated by any feasible joint entry-assignment control policy π is upper bounded by the value of the linear program defined by* (32)*, the approximate flow balance constraints* (40) *and time-averaged demand constraints* (36)*.*

Proof. Let π be any feasible policy. For each $\tau \in \mathcal{T}$ and $j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)$, define

$$\bar{z}_{j\tau k} \triangleq \frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}[x_{j\tau k}[t]\mathbb{I}\{\tau[t] = \tau\}]\,.$$

In words, $\bar{z}_{j\tau k}$ is the average flow over $1 \le t \le T$ of the demand type τ being served by pickup location j and dropoff location k. Since for each t, $z_{j\tau k}[t] \triangleq \mathbb{E}[\mathbb{I}\{\tau[t] = \tau\}x_{j\tau k}]$ satisfies the period-specific demand constraint (34) for all $\tau \in \mathcal{T}, j \in \mathcal{P}[\tau], k \in \mathcal{D}(\tau)$, the averaged constraints (36) must hold for $\bar{\mathbf{z}}$.

We can write the expected per-period payoff collected in the first T periods as:

$$\begin{aligned} W_T^\pi &= \frac{1}{T}\sum_{t=0}^{T-1} \mathbb{E}\left[\sum_{\tau\in\mathcal{T}, j\in\mathcal{P}(\tau), k\in\mathcal{D}(\tau)} w_{j\tau k}\cdot x_{j\tau k}[t]\mathbb{I}\{\tau[t] = \tau\} \right] \\ &= \sum_{\tau\in\mathcal{T}, j\in\mathcal{P}(\tau), k\in\mathcal{D}(\tau)} w_{j\tau k}\cdot \bar{z}_{j\tau k}\,, \end{aligned}$$

where we only used linearity of expectation. In words, the expected per-period payoff is the objective (32) evaluated at $\bar{\mathbf{z}}$. Similarly, for the time-average of the change of queue length we have:

$$\frac{1}{T}\cdot \mathbb{E}[\mathbf{q}[T] - \mathbf{q}[0]] = \sum_{\tau\in\mathcal{T}, j\in\mathcal{P}(\tau), k\in\mathcal{D}(\tau)} \bar{z}_{j\tau k}\cdot(\mathbf{e}_j - \mathbf{e}_k)\,,$$

which implies that $\bar{\mathbf{z}}$ satisfies the approximate flow constraints (40) since $|\sum_{j\in S} q_j[T] - q_j[0]| \le K$ for all $S \subset V$. (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 in magnitude.)

We have shown that $\bar{\mathbf{z}}$ is feasible for the given linear program with constraints (40) and (36), and the expected payoff earned W_T^π is identical to objective (32) evaluated at $\bar{\mathbf{z}}$. It follows that W_T^π is upper bounded by the value of the optimization problem defined by (32), (40) and (36) 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 6). We call $\mathbf{z}$ a (directed) *acyclic flow* if there is no (directed) cycle

$$\mathcal{C} = (\,(j_1,\tau_1,j_2),(j_2,\tau_2,j_3),\cdots,(j_s,\tau_s,j_{s+1} = j_1)\,)\,, \qquad \text{where } j_r \in V \text{ and } \tau_r \in \mathcal{T} \text{ for } r = 1,2,\cdots,s\,,$$

such that

$$z_{j_r,\tau_r,j_{r+1}} > 0 \qquad \text{for all } r = 1, \cdots, s\,.$$

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

Lemma 6. *Any feasible solution $\mathbf{z}^F$ of the finite horizon averaged SPP satisfying approximate flow balance* (40) *and the average demand constraint* (36) *can be decomposed as*

$$\mathbf{z}^{\mathrm{F}} = \mathbf{z}^{\mathrm{S}} + \mathbf{z}^{\mathrm{DAG}}\,, \tag{41}$$

where $\mathbf{z}^{\mathrm{S}}$ is a feasible solution for the SPP satisfying exact flow balance (33) *and* (36)*, and $\mathbf{z}^{\mathrm{DAG}}$ is an acyclic flow satisfying* (40) *and* (36)*.*

Proof. The existence of such a decomposition can be established using a standard flow decomposition argument (see, e.g., Williamson 2019): Start with $\mathbf{z}^{\mathrm{S}} = \mathbf{0}$ and $\mathbf{z}^{\mathrm{DAG}} = \mathbf{z}^{\mathrm{F}}$. Then, iteratively, if $\mathbf{z}^{\mathrm{DAG}}$ includes a cycle $\mathcal{C}$ with a positive flow along $\mathcal{C}$ as above, move a flow of $u(\mathcal{C}) \triangleq \min_{1\le r\le s} z_{j_r,\tau_r,j_{r+1}}$ along $\mathcal{C}$ from $\mathbf{z}^{\mathrm{DAG}}$ to $\mathbf{z}^{\mathrm{S}}$, via the updates

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

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

- $\mathbf{z}^{\mathrm{S}}$ remains feasible for the SPP, in particular, it satisfies flow balance (33).
- $\mathbf{z}^{\mathrm{F}} = \mathbf{z}^{\mathrm{S}} + \mathbf{z}^{\mathrm{DAG}}$ remains true.
- It remains true that

$$\sum_{\tau\in\mathcal{T}, j\in\mathcal{P}(\tau), k\in\mathcal{D}(\tau)} z^{\mathrm{DAG}}_{j\tau k}(\mathbf{e}_j - \mathbf{e}_k) = \sum_{\tau\in\mathcal{T}, j\in\mathcal{P}(\tau), k\in\mathcal{D}(\tau)} z^{\mathrm{F}}_{j\tau k}(\mathbf{e}_j - \mathbf{e}_k)\,.$$

  i.e., $\mathbf{z}^{\mathrm{DAG}}$ has the same net inflow/outflow from each supply node as $\mathbf{z}^{\mathrm{F}}$. In particular, $\mathbf{z}^{\mathrm{DAG}}$ satisfies approximate flow balance (40).

Moreover, the iterative process progresses monotonically: Observe that $\mathbf{z}^{\mathrm{S}}$ coordinate-wise (weakly) increases monotonically, whereas $\mathbf{z}^{\mathrm{DAG}}$ coordinate-wise (weakly) decreases monotonically (but preserves $\mathbf{z}^{\mathrm{DAG}} \ge \mathbf{0}$). Since we also know that $\mathbf{z}^{\mathrm{S}}$ is bounded, it follows that this iterative process converges. Moreover, in the limit it must be that there is no remaining cycle with positive flow in $\mathbf{z}^{\mathrm{DAG}}$ (else we observe a contradiction with the fact that the process has converged). Hence, $\mathbf{z}^{\mathrm{S}}$ and $\mathbf{z}^{\mathrm{DAG}}$ 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 averaged SPP with approximate flow balance constraints (40) cannot be much larger than the value of the averaged program $\overline{\mathrm{SPP}}$ which imposes exact flow balance constraints (33).

Lemma 7. *The value of the linear program defined by* (32)*, the approximate flow balance constraints* (40) *and time-averaged demand constraints* (36) *is bounded above by*

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

where $W^{\overline{\mathrm{SPP}}}$ is the value of the linear program $\overline{\mathrm{SPP}}$ which imposes exact flow balance constraints (33)*.*

Proof. We appeal to the decomposition from Lemma 6 to decompose any feasible solution $\mathbf{z}^{\mathrm{F}}$ to the finite horizon fluid problem as

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

where $\mathbf{z}^{\mathrm{S}}$ is feasible for $\overline{\mathrm{SPP}}$ and $\mathbf{z}^{\mathrm{DAG}}$ is a directed acyclic flow that satisfies approximate flow balance (40) and the averaged demand constraints (34). Hence, the objective (32) can be written as the sum of two terms

$$\sum_{\tau\in\mathcal{T},j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} w_{j\tau k}\cdot z^{\mathrm{F}}_{j\tau k} = \sum_{\tau\in\mathcal{T},j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} w_{j\tau k}\cdot (z^{\mathrm{S}}_{j\tau k} + z^{\mathrm{DAG}}_{j\tau k}) , \tag{42}$$

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

$$\sum_{\tau\in\mathcal{T},j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} w_{j\tau k}\cdot z^{\mathrm{S}}_{j\tau k} \le W^{\overline{\mathrm{SPP}}} .$$

We will now show that

$$\sum_{\tau\in\mathcal{T},j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} w_{j\tau k}\cdot z^{\mathrm{DAG}}_{j\tau k} \le (m-1)\cdot\frac{K}{T} < m\cdot\frac{K}{T} .$$

The lemma will follow, since this will imply an upper bound of $W^{\overline{\mathrm{SPP}}} + m\cdot\frac{K}{T}$ on the objective for any $\mathbf{z}^{\mathrm{F}}$ satisfying (40) and (36).

Consider $\mathbf{z}^{\mathrm{DAG}}$. Since it is an acyclic flow, there is an ordering $(j_1, j_2, \ldots, j_m)$ of the nodes in V such that all positive flows move supply from an earlier node to a later node in this ordering. More precisely, it holds that for any $\tau\in\mathcal{T}$,

$$z^{\mathrm{DAG}}_{j_l,\tau,j_r} = 0 \qquad \forall\, l > r \;\text{ s.t. }\; j_l\in\mathcal{P}(\tau), j_r\in\mathcal{D}(\tau) . \tag{43}$$

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 (43), $\mathbf{z}^{\mathrm{DAG}}$ does not move any supply from $V\backslash A_\ell$ to A_ℓ. Hence we have

$$\begin{aligned}\mathbf{1}^\top_{A_\ell}\left(\sum_{\tau\in\mathcal{T},j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} z^{\mathrm{DAG}}_{j\tau k}(\mathbf{e}_j - \mathbf{e}_k)\right) &= \sum_{\tau\in\mathcal{T},j\in\mathcal{P}(\tau)\cap A_\ell,k\in\mathcal{D}(\tau)\cap(V\backslash A_\ell)} z^{\mathrm{DAG}}_{j\tau k}\\ &\le \frac{K}{T} \qquad \forall\, l = 1, 2, \ldots, m-1 ,\end{aligned} \tag{44}$$

We made use of (40) to obtain the upper bound. Further, note that for each $z^{\mathrm{DAG}}_{j_l,\tau,j_r}$ with $l < r$, the term $z^{\mathrm{DAG}}_{j_l,\tau,j_r}$ is part of the above sum for $\ell = l$. Motivated by this observation, we bound the expected payoff of $\mathbf{z}^{\mathrm{DAG}}$ by first using our assumption $\max_{j,k\in V,\tau\in\mathcal{T}}|w_{j\tau k}|\le 1$ to bound the payoff by the sum of z^{DAG}s (the first inequality below), and then bounding the sum of z^{DAG}s by "allocating" $z^{\mathrm{DAG}}_{j_l,\tau,j_r}$ to the left-hand side of (44) with $\ell = l$ and summing over ℓ (the second inequality below):

$$\begin{aligned}&\sum_{\tau\in\mathcal{T},j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} w_{j\tau k}\cdot z^{\mathrm{DAG}}_{j\tau k}\\ \le &\sum_{\tau\in\mathcal{T},j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} z^{\mathrm{DAG}}_{j\tau k}\end{aligned}$$

$$
\begin{aligned}
&\le \sum_{1\le \ell<m} \sum_{\tau\in\mathcal{T}, j\in\mathcal{P}(\tau)\cap A_\ell, k\in\mathcal{D}(\tau)\cap(V\backslash A_\ell)} z^{\mathrm{DAG}}_{j\tau k} \\
&\le (m-1)\cdot \frac{K}{T}\,.
\end{aligned}
$$

The last inequality uses (44) summed over ℓ. This completes the proof. ☐

Proof of Proposition 2. The proposition follows immediately from Lemmas 5 and 7. ☐

A.2 Joint Pricing-Assignment Setting

Consider the JPA setting defined in Section 6.3. Recall that we assumed stationary demand arrival rates (in contrast to the JEA setting). The static planning problem (SPP) in the JPA setting is

$$
\begin{aligned}
\text{maximize}_{\mathbf{x}} \ & \sum_{\tau\in\mathcal{T}} \phi_\tau \left(r_\tau \left(\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} x_{j\tau k} \right) - \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} c_{j\tau k}\cdot x_{j\tau k} \right) && (45)\\
\text{s.t.} \ & \sum_{\tau\in\mathcal{T}} \phi_\tau \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} x_{j\tau k}(\mathbf{e}_j-\mathbf{e}_k) = \mathbf{0} && \text{(flow balance)} \quad (46)\\
& \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} x_{j\tau k} \le 1\,,\ x_{j\tau k}\ge 0 \quad \forall j,k\in V\,,\ \tau\in\mathcal{T} && \text{(demand constraint)}\,. \quad (47)
\end{aligned}
$$

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 JPA setting are upper bounded as*

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

Here W^{SPP} is the optimal value of SPP (45)-(47).

The main twist of the proof comparing to Proposition 2 is that the objective function in (45) is no longer linear. We first prove a JPA version of Lemma 5.

Lemma 8. *For any horizon $T<\infty$, any K and any starting state $\mathbf{q}[0]$, the expected payoff generated by any JPA policy π is upper bounded by the value of the finite horizon SPP:*

$$
\begin{aligned}
\text{maximize}_{\mathbf{x}} \ & \sum_{\tau\in\mathcal{T}} \phi_\tau\cdot \left(r_\tau \left(\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} x_{j\tau k} \right) - \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} c_{j\tau k}\cdot x_{j\tau k} \right)\\
\text{s.t.} \ & \mathbb{1}_S^\top \left(\sum_{\tau\in\mathcal{T}} \phi_\tau \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} x_{j\tau k}(\mathbf{e}_j-\mathbf{e}_k) \right) \le \frac{K}{T} \qquad \forall S\subseteq V\\
& \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} x_{j\tau k} \le 1\,,\ x_{j\tau k}\ge 0 \qquad \forall j,k\in V\,,\ \tau\in\mathcal{T}\,.
\end{aligned}
$$

Proof. Let π be any feasible JPA policy. For each demand type $\tau\in\mathcal{T}$ and $j\in\mathcal{P}(\tau)$, $k\in\mathcal{D}(\tau)$, define

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

In words, $\bar{x}_{j\tau k}$ is the average rate over the first T periods of picking up type τ demands from node j and dropping them off at node k.

Let $U_\tau[t]$ be the willingness-to-pay of a type τ demand arriving at time t. We decompose the time-average of payoff collected in the first T periods as:

$$\begin{aligned} W_T^\pi &= \frac{1}{T}\sum_{t=0}^{T-1}\mathbb{E}\left[\sum_{\tau\in\mathcal{T}}\mathbb{1}\{\tau[t]=\tau, U_\tau[t]\geq p_\tau[t]\}\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}(p_\tau[t]-c_{j\tau k})\cdot x_{j\tau k}[t]\right] \\ &= \frac{1}{T}\sum_{t=0}^{T-1}\sum_{\tau\in\mathcal{T}}\phi_\tau\cdot\mathbb{E}\left[\mathbb{1}\{U_\tau[t]\geq p_\tau[t]\}\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}(p_\tau[t]-c_{j\tau k})\cdot x_{j\tau k}[t]\,\middle|\,\tau[t]=\tau\right]. \end{aligned}$$

Because $U_\tau[t]$ is independent of $p_\tau[t]$ and $x_{j\tau k}[t]$, we have

$$W_T^\pi = \frac{1}{T}\sum_{t=0}^{T-1}\sum_{\tau\in\mathcal{T}}\phi_\tau\cdot\mathbb{E}\left[\bar{F}_\tau(p_\tau[t])\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}(p_\tau[t]-c_{j\tau k})\cdot x_{j\tau k}[t]\,\middle|\,\tau[t]=\tau\right].$$

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

$$\begin{aligned} W_T^\pi &= \frac{1}{T}\sum_{t=0}^{T-1}\sum_{\tau\in\mathcal{T}}\phi_\tau\cdot\mathbb{E}\left[\left(\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}\hat{x}_{j\tau k}[t]\right)\cdot\bar{F}_\tau^{-1}(\mu_\tau[t])-\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}c_{j\tau k}\cdot\hat{x}_{j\tau k}[t]\,\middle|\,\tau[t]=\tau\right] \\ &\leq \frac{1}{T}\sum_{t=0}^{T-1}\sum_{\tau\in\mathcal{T}}\phi_\tau\cdot\mathbb{E}\left[\left(\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}\hat{x}_{j\tau k}[t]\right)\cdot\bar{F}_\tau^{-1}\left(\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}\hat{x}_{j\tau k}[t]\right)\right. \\ &\quad\left.-\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}c_{j\tau k}\cdot\hat{x}_{j\tau k}[t]\,\middle|\,\tau[t]=\tau\right] \\ &= \frac{1}{T}\sum_{t=0}^{T-1}\sum_{\tau\in\mathcal{T}}\phi_\tau\cdot\mathbb{E}\left[r_\tau\left(\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}\hat{x}_{j\tau k}[t]\right)-\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}c_{j\tau k}\cdot\hat{x}_{j\tau k}[t]\,\middle|\,\tau[t]=\tau\right]. \end{aligned}$$

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

$$W_T^\pi \leq \sum_{\tau\in\mathcal{T}}\phi_\tau\cdot\left(-\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}c_{j\tau k}\cdot\bar{x}_{j\tau k}+\frac{1}{T}\sum_{t=0}^{T-1}r_\tau\left(\mathbb{E}\left[\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}\hat{x}_{j\tau k}[t]\,\middle|\,\tau[t]=\tau\right]\right)\right).$$

Use Jensen's inequality again, we have

$$W_T^\pi \leq \sum_{\tau\in\mathcal{T}}\phi_\tau\cdot\left(-\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}c_{j\tau k}\cdot\bar{x}_{j\tau k}+r_\tau\left(\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}\bar{x}_{j\tau k}\right)\right).$$

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

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

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 (47). Optimizing over $\bar{\mathbf{x}}$ yields the desired result. □

Proof Sketch of Proposition 3. The rest of the proof proceeds almost exactly the same as in Proposition 2. The only caveat is that the equation (42) should be replaced by inequality

$$\begin{aligned}
&\sum_{\tau \in \mathcal{T}} \phi_\tau \left(- \sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} c_{j\tau k} \cdot x^{\mathrm{F}}_{j\tau k} + r_\tau \left(\sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} x^{\mathrm{F}}_{j\tau k} \right) \right) \\
\leq &\sum_{\tau \in \mathcal{T}} \phi_\tau \left(- \sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} c_{j\tau k} \cdot x^{\mathrm{S}}_{j\tau k} + r_\tau \left(\sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} x^{\mathrm{S}}_{j\tau k} \right) \right) \\
&+ \sum_{\tau \in \mathcal{T}} \phi_\tau \left(- \sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} c_{j\tau k} \cdot x^{\mathrm{DAG}}_{j\tau k} + r_\tau \left(\sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} x^{\mathrm{DAG}}_{j\tau k} \right) \right) .
\end{aligned}$$

Here the inequality follows from $\mathbf{x}^{\mathrm{F}} = \mathbf{x}^{\mathrm{S}} + \mathbf{x}^{\mathrm{DAG}}$ and the fact that $r_\tau(\cdot)$ is subadditive by virtue of being a non-negative concave function. □

B Lyapunov Analysis: Proof of Lemma 1 and Lemma 2

In this section, we prove the counterparts of Lemma 1 and Lemma 2 for JEA (Lemma 9 and Lemma 11, resp.) and JPA setting (Lemma 10 and Lemma 12, resp.). Lemma 1 is implied by Lemma 9, and Lemma 2 is implied by Lemma 11.

B.1 Decomposition of Optimality Gap

B.1.1 Generalization of Lemma 1 for the JEA Setting

The following lemma generalizes Lemma 1 for the JEA setting.

Lemma 9. *Consider congestion functions $f_j(\cdot)$s that are strictly increasing and continuously differentiable, and that $f_j(\bar{q}_j) \leq f_k(\bar{q}_k)$ (i) for any $k \in V$ if $q_j = 0$, and (ii) for any $j \in V$ if $q_k = d_k$, $k \in V_{\mathrm{b}}$. We have the following decomposition (W^{SPP^t} is defined in Appendix A.1 and g^t_{JEA} is defined in* (25)*):*

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

Proof. For congestion functions $f_j(\bar{q}_j)$ that are 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:

$$D_F(\bar{\mathbf{q}}_1, \bar{\mathbf{q}}_2) = F(\bar{\mathbf{q}}_1) - F(\bar{\mathbf{q}}_2) - \langle \mathbf{f}(\bar{\mathbf{q}}_1), \bar{\mathbf{q}}_1 - \bar{\mathbf{q}}_2 \rangle . \tag{48}$$

Plugging $\bar{\mathbf{q}}_1 = \bar{\mathbf{q}}[t+1]$, $\bar{\mathbf{q}}_2 = \bar{\mathbf{q}}[t]$ into (48) and rearranging 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 $\frac{1}{\tilde{K}} \sum_{\tau \in \mathcal{T}} \phi_\tau^t \sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} w_{j\tau k} \cdot x_{j\tau k}[t]$ on both sides and taking conditional expectation given $\bar{\mathbf{q}}[t]$, we have:

$$\begin{aligned}
&\mathbb{E}[F(\bar{\mathbf{q}}[t+1])|\bar{\mathbf{q}}[t]] - F(\bar{\mathbf{q}}[t]) - \frac{1}{\tilde{K}} \sum_{\tau \in \mathcal{T}} \phi_\tau^t \sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} w_{j\tau k} \mathbb{E}[x_{j\tau k}[t]|\bar{\mathbf{q}}[t]] \\
= &\underbrace{-\frac{1}{\tilde{K}} \sum_{\tau \in \mathcal{T}} \phi_\tau^t \sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} w_{j\tau k} \mathbb{E}[x_{j\tau k}[t]|\bar{\mathbf{q}}[t]] + \langle \mathbf{f}(\bar{\mathbf{q}}[t]), \mathbb{E}[\bar{\mathbf{q}}[t+1]|\bar{\mathbf{q}}[t]] - \bar{\mathbf{q}}[t] \rangle}_{\text{(I)}} + \underbrace{\mathbb{E}[D_F(\bar{\mathbf{q}}[t+1], \bar{\mathbf{q}}[t])|\bar{\mathbf{q}}[t]]}_{\text{(II)}} .
\end{aligned} \tag{49}$$

Let $x_{j\tau k}^{\mathrm{NOM}}[t]$ be the "nominal" control that ignores the no-underflow constraint, i.e.

$$(x_{j\tau k}^{\mathrm{NOM}})[t] = \begin{cases} 1 & \text{if } w_{j\tau k} + f_j(\bar{q}_j[t]) - f_k(\bar{q}_k[t]) \geq 0 \\ 0 & \text{otherwise.} \end{cases} \tag{50}$$

It immediately follows that

$$(x_{j\tau k}^{\mathrm{MBP}})[t] = (x_{j\tau k}^{\mathrm{NOM}})[t] \cdot \mathbb{1}\{q_j[t] > 0,\ q_k[t] < d_k\} . \tag{51}$$

With a slight abuse of notation, denote $\mathbf{x}^{\mathrm{NOM}}$ as $\tilde{\mathbf{x}}$, $\mathbf{x}^{\mathrm{MBP}}$ as $\mathbf{x}$. Rearranging the terms in (I) and plugging in (51), we have

$$\begin{aligned}
\text{(I)} = &-\frac{1}{\tilde{K}} \sum_{\tau \in \mathcal{T}} \phi_\tau^t \sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} \big(w_{j\tau k} + f_j(\bar{q}_j[t]) - f_k(\bar{q}_k[t])\big) \cdot \mathbb{E}[x_{j\tau k}[t]|\bar{\mathbf{q}}[t]] \\
= &-\frac{1}{\tilde{K}} \sum_{\tau \in \mathcal{T}} \phi_\tau^t \sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} \big(w_{j\tau k} + f_j(\bar{q}_j[t]) - f_k(\bar{q}_k[t])\big) \cdot \mathbb{E}[\tilde{x}_{j\tau k}[t]|\bar{\mathbf{q}}[t]] \\
&+ \frac{1}{\tilde{K}} \sum_{\tau \in \mathcal{T}} \phi_\tau^t \sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} \big(w_{j\tau k} + f_j(\bar{q}_j[t]) - f_k(\bar{q}_k[t])\big) \cdot \mathbb{E}[\tilde{x}_{j\tau k}[t]|\bar{\mathbf{q}}[t]] \cdot \mathbb{1}\big\{q_j[t] = 0 \text{ or } q_k[t] = d_k\big\} .
\end{aligned}$$

By definition of the nominal control $\tilde{\mathbf{x}}$, we have:

$$\begin{aligned}
&-\frac{1}{\tilde{K}} \sum_{\tau \in \mathcal{T}} \phi_\tau^t \sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} \big(w_{j\tau k} + f_j(\bar{q}_j[t]) - f_k(\bar{q}_k[t])\big) \cdot \mathbb{E}[\tilde{x}_{j\tau k}[t]|\bar{\mathbf{q}}[t]] \\
= &-\frac{1}{\tilde{K}} \sum_{\tau \in \mathcal{T}} \phi_\tau^t \sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} \big(w_{j\tau k} + f_j(\bar{q}_j[t]) - f_k(\bar{q}_k[t])\big)^+ \\
= &-\frac{1}{\tilde{K}} \cdot g_{\mathrm{JEA}}^t(\mathbf{f}(\bar{\mathbf{q}}[t])) .
\end{aligned}$$

Using the fact that $\max_{j,k\in V,\tau\in\mathcal{T}}|w_{j\tau k}|= 1$, we have

$$\begin{aligned}
&\frac{1}{\tilde{K}}\sum_{\tau\in\mathcal{T}}\phi_\tau^t \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} \big(w_{j\tau k}+f_j(\bar{q}_j[t])-f_k(\bar{q}_k[t])\big)\cdot\mathbb{E}[\tilde{x}_{j\tau k}[t]|\bar{\mathbf{q}}[t]]\cdot\mathbb{1}\left\{q_j[t]=0 \text{ or } q_k[t]=d_k\right\}\\
\le\ &\frac{1}{\tilde{K}}\sum_{\tau\in\mathcal{T}}\phi_\tau^t\cdot\mathbb{1}\left\{q_j[t]=0 \text{ or } d_j,\ \exists j\right\}\\
&+\frac{1}{\tilde{K}}\sum_{\tau\in\mathcal{T}}\phi_\tau^t \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} \big(f_j(\bar{q}_j[t])-f_k(\bar{q}_k[t])\big)^+\cdot\mathbb{1}\left\{q_j[t]=0 \text{ or } q_k[t]=d_k\right\}\\
\le\ &\frac{1}{\tilde{K}}\cdot\mathbb{1}\left\{q_j[t]=0 \text{ or } d_j,\ \exists j\right\}\,.
\end{aligned}$$

Here the last inequality follows from the assumption that $f_j(\bar{q}_j[t])\le f_k(\bar{q}_k[t])$ for any $j,k\in V$ when $q_j[t]=0$ or $q_k[t]=d_k$. (Condition (ii) in the lemma as stated only covers $k\in V_{\mathrm{b}}$. However, in case where $k\notin V_{\mathrm{b}}$, i.e., $d_k=K$, and $q_k[t]=d_k$ holds, then we automatically have $q_k[t]=K\ \Rightarrow\ q_j[t]=0$ and condition (i) kicks in, i.e., condition (ii) in fact holds for all $k\in V$.) Note that when no queue has finite buffer constraints as in the illustrative model in Section 2, such assumption is satisfied by any congestion function such that $f_j(\bar{q}_j)=f(\bar{q}_j)$ for all $j\in V$ where $f(\cdot)$ is a monotonically increasing function.

Combining the above inequality and equality yields

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

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[23]

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

Plugging the above bounds on (I) and (II) into (49), we have

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

Rearranging the terms yields:

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

Adding W^{SPP^t} to both sides concludes the proof. □

B.1.2 Joint Pricing-Assignment Setting

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

Lemma 10. *Consider congestion functions $f_j(\cdot)$s that are strictly increasing and continuously differentiable, and that $f_j(\bar{q}_j)\le f_k(\bar{q}_k)$ (i) for any $k\in V$ if $q_j=0$, and (ii) for any $j\in V$ if*

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

$q_k = d_k$, $k \in V_{\text{b}}$. *We have the following decomposition:*

$$W^* - \mathbb{E}[v^{\text{MBP}}[t]|\bar{\mathbf{q}}[t]] \le \tilde{K}\left(F(\bar{\mathbf{q}}[t]) - \mathbb{E}[F(\bar{\mathbf{q}}[t+1])|\bar{\mathbf{q}}[t]]\right) + \frac{1}{2\tilde{K}} \cdot \max_{j\in V} f_j'(\bar{q}_j[t]) + \left(W^{\text{SPP}} - g_{\text{JPA}}(\mathbf{f}(\bar{\mathbf{q}}[t]))\right) + \mathbb{1}\left\{q_j[t] = 0 \text{ or } d_j, \exists j\right\}, \tag{52}$$

where $g_{\text{JPA}}(\mathbf{y})$ *is defined in* (29).

Proof Sketch. The proof is analogous to Lemma 9. To use the strong duality argument, we prove below that $g_{\text{JPA}}(\cdot)$ defined in (29) is indeed the partial dual function of the SPP (45)-(47). 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 (46). We have

$$\begin{aligned} g_{\text{JPA}}(\mathbf{y}) = & \max_{\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} x_{j\tau k}\le 1, x_{j\tau k}\ge 0} \sum_{\tau\in\mathcal{T}} \phi_\tau \left(r_\tau\left(\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} x_{j\tau k}\right) \right. \\ & \left. + \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} \left(-c_{j\tau k} + y_j - y_k\right) x_{j\tau k} \right) \\ = & \sum_{\tau\in\mathcal{T}} \phi_\tau \max_{\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} x_{j\tau k}\le 1, x_{j\tau k}\ge 0} \left(r_\tau\left(\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} x_{j\tau k}\right) \right. \\ & \left. + \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} \left(-c_{j\tau k} + y_j - y_k\right) x_{j\tau k} \right). \end{aligned}$$

Let $\mu_\tau = \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} x_{j\tau k}$, we have

$$\begin{aligned} g_{\text{JPA}}(\mathbf{y}) = & \sum_{\tau\in\mathcal{T}} \phi_\tau \max_{0\le\mu_\tau\le 1} \max_{\sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} x_{j\tau k}=\mu_\tau, x_{j\tau k}\ge 0} \left(r_\tau\left(\mu_\tau\right) + \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} \left(-c_{j\tau k} + y_j - y_k\right) x_{j\tau k} \right) \\ = & \sum_{\tau\in\mathcal{T}} \max_{0\le\mu_\tau\le 1} \left(r_\tau\left(\mu_\tau\right) + \mu_\tau \max_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} \left(-c_{j\tau k} + y_j - y_k\right) \right). \end{aligned}$$

□

B.2 Geometry of the Dual Function

B.2.1 Generalization of Lemma 2 for the JEA Setting

The following lemma generalizes Lemma 2 for the JEA setting.

Lemma 11. *Consider congestion functions* $(f_j(\cdot))_{j\in V}$ *that are strictly increasing and continuously differentiable, and any triple* $(\boldsymbol{\phi}^t, \mathcal{P}, \mathcal{D})$ *with connectedness at least* $\alpha_{\min} > 0$. *Then the term* $\mathcal{V}_3 = W^{\text{SPP}^t} - g^t_{\text{JEA}}(\bar{\mathbf{q}}[t])$ *(see Lemma 9) is bounded above as*

$$\mathcal{V}_3 \le -\alpha_{\min} \cdot \left[\max_{j\in V} f_j(\bar{q}_j[t]) - \min_{j\in V} f_j(\bar{q}_j[t]) - 2m\right]^+ .$$

Proof. Consider $\mathbf{y} \triangleq (f_j(\bar{q}_j[t]))_{j\in V}$ and order the nodes in V in decreasing order of y_j as $y_{i_1} \geq y_{i_2} \geq \cdots y_{i_m}$. For $r = 1$ to $r = m-1$, we repeat the following procedure: if $y_{i_r} - y_{i_{r+1}} \leq 2$, then do nothing and move on to $r+1$; if otherwise, perform the following update:

$$y_{i_k} \leftarrow \; y_{i_k} - \big(y_{i_r} - y_{i_{r+1}} - 2\big) \qquad \forall 1 \leq k \leq r\,.$$

Recall that $g(\mathbf{y}) = \sum_{\tau\in\mathcal{T}} \phi_\tau \sum_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}[w_{j\tau k} + y_j - y_k]^+$. 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, for $\tau \in \mathcal{P}^{-1}(j)\cap\mathcal{D}^{-1}(k)$,

$$w_{j\tau k} + y_j - y_k \geq \; w_{j\tau k} + y_{i_r} - \big(y_{i_r} - y_{i_{r+1}} - 2\big) - y_{i_{r+1}} \geq w_{j\tau k} + 2 > 0\,,$$

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

$$\left(\sum_{\tau\in\mathcal{P}^{-1}(\{i_1,\cdots,i_r\})\cap\mathcal{D}^{-1}(\{i_{r+1},\cdots,i_m\})} \phi_\tau\right) \cdot \big[y_{i_r} - y_{i_{r+1}} - 2\big]^+ \,.$$

Note that the first term is lower bounded by $\alpha_{\min}$ defined in (26). As a result, after the finishing the procedure, $g(\mathbf{y})$ decreased by at least:

$$\alpha_{\min} \cdot \sum_{r=1}^{m-1} \big[y_{i_r} - y_{i_{r+1}} - 2\big]^+ \geq \; \alpha_{\min} \cdot \big[y_{i_1} - y_{i_m} - 2m\big]^+ \,.$$

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

$$g^t_{\text{JEA}}(\mathbf{y}) - W^{\text{SPP}^t} \geq \; \alpha_{\min} \cdot \left[\max_{j\in V} y_j - \min_{k\in V} y_k - 2m\right]^+ \,.$$

This concludes the proof. $\square$

B.2.2 Joint Pricing-Assignment Setting

The following lemma is the counterpart of Lemma 2 for the JPA setting.

Lemma 12. *Consider congestion functions $(f_j(\cdot))_{j\in V}$ that are strictly increasing and continuously differentiable, and any $\boldsymbol{\phi}$ with connectedness $\alpha(\boldsymbol{\phi},\mathcal{P},\mathcal{D}) > 0$. We have*

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

where W^{SPP} is the value of SPP (45)-(47)*, and $\alpha(\boldsymbol{\phi},\mathcal{P},\mathcal{D})$ is defined in* (26)*.*

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

$$\text{argmax}_{\{0\leq\mu_\tau\leq 1\}} \left(r_\tau(\mu_\tau) + \mu_\tau \cdot \max_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} \big(-c_{j\tau k} + y_j - y_k\big)\right) = 1\,,$$

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

$$\operatorname{argmax}_{\{0 \le \mu_\tau \le 1\}} \left(r_\tau(\mu_\tau) + \mu_\tau \cdot \max_{k \in \mathcal{P}(\tau), j \in \mathcal{D}(\tau)} \left(-c_{k\tau j} + y_k - y_j \right) \right) = 0 \,.$$

$\square$

C Proofs of Lemma 3 and Theorems 2 and 3

In this section, we first show that Lemma 3 and its counterparts for JEA and JPA settings hold if the congestion function satisfy certain growth conditions. Together with the lemmas derived in Appendix B, we conclude that Theorems 2 and 3 hold for any congestion function that satisfies the growth conditions. Finally, we verify the growth condition for several congestion functions including (22).

C.1 A Sufficient Condition of Lemma 3 and its Counterparts for JEA and JPA

Since the statements and proofs for the JEA and JPA settings are almost identical, we only provide them for the JEA setting to avoid redundancy. The generalization of Lemma 3 to the JEA setting is as follows:

Lemma 13. *Consider the congestion function* (22)*. Consider a set V of $m = |V| > 1$ nodes, a subset $V_{\mathrm{b}} \subset V$ of buffer-constrained nodes with scaled buffer sizes $\bar{d}_j \in (0,1)$ (recall that we define $\bar{d}_j \triangleq 1$ for all $j \in V \backslash V_{\mathrm{b}}$) satisfying $\sum_{j \in V} \bar{d}_j > 1$, and any $(\boldsymbol{\phi}^t, \mathcal{P}, \mathcal{D})$ that satisfies Condition 2 with $\alpha_{\min} > 0$. Then there exists $K_1 = \operatorname{poly}\left(m, \bar{\mathbf{d}}, \frac{1}{\alpha_{\min}}\right)$ such that for $K \ge K_1$,*

$$\mathcal{V}_2 + \mathcal{V}_3 + \mathcal{V}_4 \le M_2 \cdot \frac{1}{\tilde{K}} \,, \qquad \text{for } M_2 = C \frac{\sqrt{m}}{\min_{j \in V} \bar{d}_j} \left(\frac{\sum_{j \in V} \bar{d}_j}{\min\{\sum_{j \in V} \bar{d}_j - 1, 1\}} \right)^{3/2} ,$$

where $\mathcal{V}_2, \mathcal{V}_3, \mathcal{V}_4$ were defined in Lemma 9, $\tilde{K}$ was defined in (21)*, and $C > 0$ is a universal constant that is independent of m, $\bar{\mathbf{d}}$, K, or $\alpha_{\min}$.*

Since the congestion function (7) is a special case of (22), and the illustrative model introduced in Section 2 is a special case of JEA model, Lemma 13 will imply Lemma 3.

We define below a growth condition for congestion functions. Lemma 14 will imply that if the congestion function satisfies this growth condition (with certain parameters), then Lemma 13 holds.

Condition 4 (Growth condition for congestion functions)**.** *We say the congestion functions $(f_j(\cdot))_{j \in V}$ satisfy the growth condition with parameters $(\alpha, K_1, M_1, M_2) \in \mathbb{R}^4_{++}$ if the following holds:*

1. *For each $j \in V$, $f_j(\cdot)$ is strictly increasing and continuously differentiable. Moreover,*
   - *(a) For any $K > K_1$, $f_j(\bar{q}_j) \le f_k(\bar{q}_k)$ (i) for any $k \in V$ if $q_j = 0$, and (ii) for any $j \in V$ if $q_k = d_k$, $k \in V_{\mathrm{b}}$.*
   - *(b) For any $j, k \in V$, we have $f_j\left(\frac{\bar{d}_j}{\sum_{j \in V} \bar{d}_j}\right) = f_k\left(\frac{\bar{d}_k}{\sum_{j \in V} \bar{d}_j}\right)$.*
2. *Define*

$$\mathcal{B}(\mathbf{f}) \triangleq \left\{ \bar{\mathbf{q}} \in \Omega : \max_{j \in V} \left| f_j\left(\frac{\bar{d}_j}{\sum_{j \in V} \bar{d}_j} \right) - f_j(\bar{q}_j) \right| \le 4m \right\} .$$

Denote $\bar{\mathcal{B}}(\mathbf{f}) \triangleq \Omega \backslash \mathcal{B}(\mathbf{f})$.

(a) *For any* $K > K_1$, $\forall \bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$,

$$\alpha \left(\max_{j \in V} \left| f_j \left(\frac{\bar{d}_j}{\sum_{j \in V} \bar{d}_j} \right) - f_j(\bar{q}_j) \right| - 2m \right)^+ \geq \frac{1}{2\tilde{K}} \cdot \max_{j \in V} f'_j(\bar{q}_j) + \mathbb{1}\{q_j = 0 \text{ or } d_j, \exists j\} . \tag{53}$$

(b) *Let* $F(\bar{\mathbf{q}})$ *be the antiderivative of* $\mathbf{f}(\bar{\mathbf{q}}) \triangleq (f_j(\bar{q}_j))_{j \in V}$, *we have* $\sup_{\mathbf{q}, \mathbf{q}' \in \Omega}(F(\bar{\mathbf{q}}) - F(\bar{\mathbf{q}}')) \leq M_1$.

(c) *We have* $\sup_{\bar{\mathbf{q}} \in \mathcal{B}(\mathbf{f})} \max_{j \in V} f'_j(\bar{q}_j) \leq M_2$.

(d) *If* $\exists j \in V$ *such that* $q_j = 0$ *or* $q_j = d_j$, *then* $\bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$.

Lemma 14. *In the JEA setting, if the congestion functions* $(f_j(\cdot))_{j \in V}$ *satisfy the growth conditions (Condition 4) with parameters* $(\alpha_{\min}, K_1, M_1, M_2)$, *then for* $K \geq K_1$,

$$\mathcal{V}_2 + \mathcal{V}_3 + \mathcal{V}_4 \leq M_2 \cdot \frac{1}{\tilde{K}} , \tag{54}$$

where $\mathcal{V}_2, \mathcal{V}_3, \mathcal{V}_4$ *were defined in Lemma 9 and* $\tilde{K}$ *was defined in* (21).

Proof of Lemma 14. Recall that

$$\mathcal{V}_2 + \mathcal{V}_3 + \mathcal{V}_4 = \frac{1}{2\tilde{K}} \cdot \max_{j \in V} f'_j(\bar{q}_j[t]) + \left(W^{\text{SPP}} - g^t_{\text{JEA}}(\mathbf{f}(\bar{\mathbf{q}}[t])) \right) + \mathbb{1}\left\{ q_j[t] = 0 \text{ or } d_j, \exists j \in V \right\} .$$

For $\bar{\mathbf{q}} \in \mathcal{B}(\mathbf{f})$, since the congestion functions satisfy Condition 4, we have $\mathcal{V}_4 = 0$. By definition of W^{SPP} we have $\mathcal{V}_3 \leq 0$. As a result, it follows from Condition 4 that

$$\mathcal{V}_2 + \mathcal{V}_3 + \mathcal{V}_4 \leq \frac{1}{2\tilde{K}} \cdot \sup_{\bar{\mathbf{q}} \in \mathcal{B}(\mathbf{f})} \max_{j \in V} f'_j(\bar{q}_j[t]) = M_2 \cdot \frac{1}{\tilde{K}} .$$

For $\bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$, it follows from Lemma 11 that

$$\mathcal{V}_3 \leq \; -\alpha_{\min} \cdot \left[\max_{j \in V} f_j(\bar{q}_j) - \min_{j \in V} f_j(\bar{q}_j) - 2m \right]^+ .$$

Note that

$$\begin{aligned} &\max_{j \in V} \left| f_j \left(\frac{\bar{d}_j}{\sum_{j \in V} \bar{d}_j} \right) - f_j(\bar{q}_j) \right| \\ \leq \; &\max \left\{ \max_{j \in V} f_j(\bar{q}_j) - \min_{j \in V} f_j \left(\frac{\bar{d}_j}{\sum_{j \in V} \bar{d}_j} \right), \max_{j \in V} f_j \left(\frac{\bar{d}_j}{\sum_{j \in V} \bar{d}_j} \right) - \min_{j \in V} f_j(\bar{q}_j) \right\} . \end{aligned}$$

Note that there must exists $j^* \in V$ such that $\bar{q}_{j^*} \leq \frac{\bar{d}_{j^*}}{\sum_{j \in V} \bar{d}_j}$, hence $f_{j^*}(\bar{q}_{j^*}) \leq f_j \left(\frac{\bar{d}_{j^*}}{\sum_{j \in V} \bar{d}_j} \right)$. Because the congestion functions satisfy Condition 4 point 1(d), we have $f_j \left(\frac{\bar{d}_j}{\sum_{j \in V} \bar{d}_j} \right)$ has the same value for all $j \in V$, therefore

$$\begin{aligned}
\max_{j\in V} f_j(\bar{q}_j) - \min_{j\in V} f_j\left(\frac{\bar{d}_j}{\sum_{j\in V}\bar{d}_j}\right) &= \max_{j\in V} f_j(\bar{q}_j) - f_{j^*}\left(\frac{\bar{d}_{j^*}}{\sum_{j\in V}\bar{d}_j}\right) \\
&\le \max_{j\in V} f_j(\bar{q}_j) - f_{j^*}(\bar{q}_{j^*}) \\
&\le \max_{j\in V} f_j(\bar{q}_j) - \min_{j\in V} f_j(\bar{q}_j)\,.
\end{aligned}$$

Similarly, we can show that

$$\max_{j\in V} f_j\left(\frac{\bar{d}_j}{\sum_{j\in V}\bar{d}_j}\right) - \min_{j\in V} f_j(\bar{q}_j) \le \max_{j\in V} f_j(\bar{q}_j) - \min_{j\in V} f_j(\bar{q}_j)\,.$$

Combined, we have

$$\mathcal{V}_3 \le -\alpha_{\min}\cdot\left(\max_{j\in V}\left|f_j\left(\frac{\bar{d}_j}{\sum_{j\in V}\bar{d}_j}\right) - f_j(\bar{q}_j)\right| - 2m\right)^+ .$$

Plugging in Condition 4 point 2(a), we have for $\bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$,

$$\mathcal{V}_2 + \mathcal{V}_3 + \mathcal{V}_4 \le 0\,.$$

Combine the above two cases, we conclude the proof. □

It remains to be shown that the congestion function (22) satisfies Condition 4.

Lemma 15. *The congestion function* (22) *satisfies the growth conditions (Condition 4) with parameters* $(\alpha_{\min}, K_1, M_1, M_2)$ *where*

$$K_1 = \text{poly}\left(m, \bar{\mathbf{d}}, \frac{1}{\alpha_{\min}}\right), \quad M_1 = Cm\,, \quad M_2 = C\frac{1}{\min_{j\in V}\bar{d}_j}\left(\frac{\sum_{j\in V}\bar{d}_j}{\min\{\sum_{j\in V}\bar{d}_j - 1, 1\}}\right)^{3/2}\sqrt{m}\,.$$

Here C is a universal constant that is independent of m, $\bar{\mathbf{d}}$, K and $\alpha_{\min}$.

We delay the proof of Lemma 15 to Appendix C.3. We are now ready to prove Lemma 13.

Proof of Lemma 13. Lemma 13 immediately follows from Lemma 14 and Lemma 15. □

C.2 Proof of Main Theorems

Recall that we proved Theorem 1 in Section 5 using Lemmas 1, 2, and 3. Similarly, we can prove Theorem 2 and 3.

Proof of Theorem 2. We draw inspiration from the proof of Theorem 1, along with some additional work to handle time-varying demand arrival rates, for which we draw upon Lemma 4 and Proposition 2.

Note that for the congestion functions defined in (22), we have $f_j(\bar{q}_j[t]) \le f_k(\bar{q}_k[t])$ when $q_j[t] = 0$ or $q_k[t] = d_k$. Also, the functions are strictly increasing and continuously differentiable. Hence, Lemmas 9, 11, and 13 (the JEA versions of Lemmas 1, 2 and 13) apply to the congestion functions (22).

As in the proof of Theorem 1, we argue as follows: Plugging in Lemma 13 into the bound in Lemma 9 and taking expectation, we obtain

$$W^{\mathrm{SPP}^t} - \mathbb{E}[v^{\mathrm{MBP}}[t]] \le \tilde{K}\left(\mathbb{E}[F(\bar{\mathbf{q}}[t])] - \mathbb{E}[F(\bar{\mathbf{q}}[t+1])]\right) + M_2\frac{1}{\tilde{K}}\,, \tag{55}$$

for all $K \ge K_2 = \mathrm{poly}\left(m, \bar{\mathbf{d}}, \frac{1}{\alpha_{\min}}\right)$, where $M_2 = C_2\frac{\sqrt{m}}{\min_{j\in V}\bar{d}_j}\left(\frac{\sum_{j\in V}\bar{d}_j}{\min\{\sum_{j\in V}\bar{d}_j - 1, 1\}}\right)^{3/2}$ for a universal constant C_2. Consider the first T_0 periods. Take the sum of both sides of the inequality (55) from $t=0$ to $t=T_0-1$, and divide the sum by T_0. This yields

$$\begin{aligned}
\frac{1}{T_0}\sum_{t=0}^{T_0-1} W^{\mathrm{SPP}^t} \;-\; W^{\mathrm{MBP}}_{T_0} &\le \frac{\tilde{K}}{T_0}\left(\mathbb{E}[F(\bar{\mathbf{q}}[0])] - \mathbb{E}[F(\bar{\mathbf{q}}[T_0])]\right) + M_2\frac{1}{\tilde{K}} \\
&\le \frac{\tilde{K}}{T_0}\left(\sup_{\bar{\mathbf{q}},\bar{\mathbf{q}}'\in\Omega}\left(F(\bar{\mathbf{q}}) - F(\bar{\mathbf{q}}')\right)\right) + M_2\frac{1}{\tilde{K}} \\
&\le \frac{\tilde{K}}{T_0}C_1 m + M_2\frac{1}{\tilde{K}}\,,
\end{aligned} \tag{56}$$

for all $K \ge K_2$, where C_1 is a universal constant. Here we used the bound $\sup_{\mathbf{q},\mathbf{q}'\in\Omega}(F(\bar{\mathbf{q}}) - F(\bar{\mathbf{q}}')) \le C_1 m$ from Lemma 15 (specifically the part of the lemma about Condition 4 part 2(b)).

Using Proposition 2 and then Lemma 4, we have

$$\begin{aligned}
L^{\mathrm{MBP}}_{T_0} = W^*_{T_0} - W^{\mathrm{MBP}}_{T_0} &\le W^{\overline{\mathrm{SPP}}} - W^{\mathrm{MBP}}_{T_0} + m\frac{K}{T_0} \\
&\le \eta T_0 m/2 + \frac{1}{T_0}\sum_{t=0}^{T_0-1} W^{\mathrm{SPP}^t} \;-\; W^{\mathrm{MBP}}_{T_0} + m\frac{K}{T_0} \\
&\le \frac{\tilde{K}}{T_0}m(C_1+1) + M_2\frac{1}{\tilde{K}} + \eta T_0\cdot m/2 \\
&\le \frac{K}{T_0}2m(C_1+1) + M_2\frac{1}{K} + T_0\eta\cdot m/2\,.
\end{aligned} \tag{57}$$

where we used (56) in the third inequality, and $K \le \tilde{K} \le 2K$ for all $K \ge K_3 = m^2$ in the last inequality. It remains to choose T_0 appropriately, i.e., to divide the horizon T into intervals of appropriate length. Note that the bound on per period loss (57) is minimized for $T_0 = T_* = 2\sqrt{(C_1+1)K/\eta}$, which makes the first and third terms equal. This observation will guide our choice of T_0.

If $T \le T_*$, we set $T_0 = T$ and we immediately have

$$L^{\mathrm{MBP}}_{T} \le \frac{K}{T}4m(C_1+1) + M_2\frac{1}{K} \qquad \forall T < T_*\,, \tag{58}$$

since the first term is larger than the third term in (57). If $T > T_*$ then we divide T into $\lceil T/T_*\rceil$ intervals of equal length (up to rounding error). In particular, each interval has length $T_0 \in [T_*/2, T_*]$, the first term is again larger than the third term in (57) and so the per period loss in each interval is bounded above by

$$\frac{K}{T_0}4m(C_1+1) + M_2\frac{1}{K} \le \frac{K}{T_*/2}4m(C_1+1) + M_2\frac{1}{K} = \sqrt{\eta K}4m\sqrt{C_1+1} + M_2\frac{1}{K}\,.$$

Since this bound holds for each interval, it holds for the full horizon of length T, i.e.,

$$L_T^{\mathrm{MBP}} \le \sqrt{\eta K} 4m\sqrt{C_1+1} + M_2\frac{1}{K} \qquad \forall T \ge T_* \,. \tag{59}$$

Combining (58) and (59), we obtain that for any $K \ge K_1 \triangleq \max(K_2, K_3)$ and any horizon T, we have

$$L_T^{\mathrm{MBP}} \le \sqrt{\eta K} 4m\sqrt{C_1+1} + M_2\frac{1}{K} + \frac{K}{T}4m(C_1+1) \le M_1\left(\frac{K}{T} + \sqrt{\eta K}\right) + M_2\frac{1}{K}\,,$$

for $M_1 \triangleq 4m(C_1+1)$. Defining $C \triangleq \max(C_2, 4(C_1+1))$ we obtain the bound claimed in the theorem. □

Proof sketch for Theorem 3. The proof is a direct extension of the proof of Theorem 1, and follows from Lemmas 10, 12, and the JPA counterpart of Lemma 13 (which is almost identical to Lemma 13, and was hence omitted). We bound M_1 using Lemma 15. □

Since Condition 4 implies Lemma 13 (using Lemma 14), we have the following general version of Theorem 2 using the exact same proof as that of Theorem 2.

Theorem 4 (General result for the JEA setting)**.** *Consider a set V of $m \triangleq |V| > 1$ nodes, a subset $V_{\mathrm{b}} \subseteq V$ of buffer-constrained nodes with scaled buffer sizes $\bar{d}_j \in (0,1)$ $\forall j \in V_{\mathrm{b}}$ satisfying*[24] *$\sum_{j\in V}\bar{d}_j > 1$, and a minimum connectivity $\alpha_{\min} > 0$. Consider any congestion functions $(f_j(\cdot))_{j\in V}$ that satisfy Condition 4 with parameters $(\alpha = \alpha_{\min}, K_1, M_1, M_2) \in \mathbb{R}^4_{++}$. Then for any horizon T, any $K \ge K_1$, and any sequence of demand arrival rates $(\boldsymbol{\phi}^t)_{t=0}^{T-1}$ which varies η-slowly (for some $\eta \in [0,2]$) and pickup and dropoff neighborhoods $\mathcal{P}$ and $\mathcal{D}$ such that $(\boldsymbol{\phi}^t, \mathcal{P}, \mathcal{D})$ is $\alpha_{\min}$-strongly connected (Condition 2) for all $t \le T-1$, we have*

$$L_T^{\mathrm{MBP}} \le 4(M_1+m)\cdot\left(\frac{K}{T} + \sqrt{\eta K}\right) + M_2\cdot\frac{1}{K}\,.$$

In the following subsection, we will show examples of alternate congestion functions that satisfy Condition 4 and obtain the corresponding parameters K_1, M_1 and M_2.

C.3 Validating Condition 4 for Congestion Functions

In this section, we prove Lemma 15. We will go a step further and show that Lemma 15 holds several other congestion functions.

Recall the congestion function defined in (22): let $V_{\mathrm{b}} \subset V$ be the subset of buffer-constrained nodes with scaled buffer sizes $\bar{d}_j \in (0,1)$, and

$$f_j(\bar{q}_j) = \sqrt{m}\cdot C_b\cdot\left(\left(1 - \frac{\bar{q}_j}{\bar{d}_j}\right)^{-\frac{1}{2}} - \left(\frac{\bar{q}_j}{\bar{d}_j}\right)^{-\frac{1}{2}} - D_b\right), \qquad \forall j \in V_{\mathrm{b}}$$
$$f_j(\bar{q}_j) = -\sqrt{m}\cdot\bar{q}_j^{-\frac{1}{2}}, \qquad \forall j \in V\backslash V_{\mathrm{b}}$$

Here C_b and D_b are normalizing constants chosen as follows. Define $\epsilon \triangleq \frac{\delta_K}{\tilde{K}}$ (where δ_K and $\tilde{K}$ were defined in (21)). Let $h_{\mathrm{b}}(\bar{q}) \triangleq (1-\bar{q})^{-\frac{1}{2}} - \bar{q}^{-\frac{1}{2}}$ and $h(\bar{q}) \triangleq -\bar{q}^{-\frac{1}{2}}$. Define $C_b \triangleq \frac{h(\epsilon) - h(1/\sum_{j\in V}\bar{d}_j)}{h_{\mathrm{b}}(\epsilon) - h_{\mathrm{b}}(1/\sum_{j\in V}\bar{d}_j)}$

[24] Recall that we define $\bar{d}_j \triangleq 1$ for all $j \in V\backslash V_{\mathrm{b}}$.

and $D_b \triangleq h_{\mathrm{b}}(1/\sum_{j\in V} \bar{d}_j) - C_b^{-1} h(1/\sum_{j\in V} \bar{d}_j)$. These definitions ensure that Condition 4 point 1(b) holds, and are useful in establishing Condition 4 point 1(a).

Proof of Lemma 15. (The proof of this lemma involves a lot of notations and computation. For readability, we use the following simplifying notation (with a slight abuse of notation): for $x_a, y_a \in \mathbb{R}_+$ where $a \in \mathcal{A} \subset \mathbb{Z}_+$, $\{x_a\} = O(\{y_a\})$ ($\{x_a\} = \Omega(\{y_a\})$, resp.) means that there exists a universal constant $C > 0$ that does not depend on $m, K, \bar{\mathbf{d}}$, or $\alpha_{\min}$ such that $x_a \le C y_a$ ($x \ge C y_a$, resp.) for each $a \in \mathcal{A}$. We say $\{x_a\} = \Theta(\{y_a\})$ if $\{x_a\} = O(\{y_a\})$ and $\{x_a\} = \Omega(\{y_a\})$.) Denote $\bar{d}_\Sigma \triangleq \sum_{j\in V} \bar{d}_j$, $\bar{d}_g \triangleq \min\{1, \sum_{j\in V} \bar{d}_j - 1\}$, $\bar{d}_{\min} \triangleq \min_{j\in V} \bar{d}_j$. Recall that $\bar{d}_j \in (0,1)$ for any $j \in V_{\mathrm{b}}$, and that $\bar{d}_\Sigma > 1$.

- Point 1. It is not hard to see that the congestion functions $(f_j(\bar{q}_j))_{j\in V}$ are strictly increasing and continuously differentiable. For any $K > 0$, we have $f_j(\bar{q}_j) = f_k(\bar{q}_k)$ for any $j, k \in V$ if $q_j = q_k = 0$. As a result, if $q_j = 0$, we have $f_j(\bar{q}_j) \le f_k(\bar{q}_k)$ for any $k \in V$. It can be easily verified that Point 1(b) is also satisfied by any $K > 0$.

  It remains to be shown that there exists $K_1 < \infty$ such that for $K \ge K_1$, we have $f_j(\bar{q}_j) \le f_k(\bar{q}_k)$ for any $j \in V$ if $q_k = d_k$ and $k \in V_{\mathrm{b}}$. To this end, if suffices to check the inequality $f_j(\bar{q}_j) \le f_k(\bar{q}_k)$ for $q_j = d_j$, $q_k = d_k$ where $j \in V \backslash V_{\mathrm{b}}$ and $k \in V_{\mathrm{b}}$: In this case, we have $f_j(\bar{q}_j) \le 0$; for $K = \Omega(\max\{\bar{d}_\Sigma^2, \frac{\bar{d}_\Sigma^2}{\bar{d}_g^2}\})$, we have $C_b = \Theta(1)$, $D_b = O(\sqrt{\frac{\bar{d}_\Sigma}{\bar{d}_g}})$ hence $f_k(\bar{q}_k) = \Omega(\sqrt{m}\frac{K^{1/4}}{\bar{d}_g^{1/2}}) \ge 0$. Therefore point 1 is satisfied for $K_1 = O(\max\{\bar{d}_\Sigma^2, \frac{\bar{d}_\Sigma^2}{\bar{d}_g^2}\}) = O(\frac{\bar{d}_\Sigma^2}{\bar{d}_g^2})$.

- Point 2(a). For $\mathbf{q}$ such that $\bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$ and $0 < q_j < d_j$ for any $j \in V$, we have, by definition of $\bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$,

  $$\text{LHS of (53)} \ge 2m\alpha\,.$$

  On the other hand, we have for $K = \Omega(\frac{\bar{d}_\Sigma^2}{\bar{d}_g^2})$, we have $C_b = \Theta(1)$ hence

  $$\text{RHS of (53)} = O\left(\frac{1}{K} \cdot \sqrt{m} \cdot K^{3/4} \cdot \bar{d}_{\min}^{-1} \bar{d}_g^{-3/2}\right)$$

  Here the RHS of (53) is maximized when $q_j = 0$ or $q_j = d_j$. Therefore (53) holds for $K \ge K_1 = \Omega\left(\max\left\{\frac{\bar{d}_\Sigma^2}{\bar{d}_g^2}, \frac{1}{m^2\alpha^4\bar{d}_{\min}^4\bar{d}_g^6}\right\}\right)$. For $\mathbf{q}$ such that $\bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$ and $q_j = 0$ or d_j for some $j' \in V$, we have

  $$\text{LHS of (53)} \ge \alpha\sqrt{m} \cdot \Omega\left(K^{1/4} - \sqrt{\bar{d}_\Sigma}\right)\,,$$

  which is obtained by plugging in $q_{j'}$. For $K = \Omega(\frac{\bar{d}_\Sigma^2}{\bar{d}_g^2})$, we also have

  $$\text{RHS of (53)} = O\left(\frac{1}{K} \cdot \sqrt{m} \cdot K^{3/4} \cdot \bar{d}_{\min}^{-1} \bar{d}_g^{-3/2} + 1\right)\,,$$

  Using the analysis above, for $K \ge \Omega\left(\frac{m^2}{\bar{d}_{\min}^4 \bar{d}_g^6}\right)$, the first term in the parentheses is $O(1)$. In this case we have RHS of (53) $= O(1)$. Therefore (53) holds for

$$K = \Omega\left(\max\left\{\frac{1}{m^2\alpha^4\bar{d}_{\min}^4\bar{d}_g^6}, \frac{m^2}{\bar{d}_{\min}^4\bar{d}_g^6}, \frac{1}{\alpha^2\bar{d}_g^3}, \frac{\bar{d}_\Sigma^2}{\bar{d}_g^2}\right\}\right).$$

Combined, (53) holds for $K_1 = O\left(\max\left\{\frac{1}{m^2\alpha^4\bar{d}_{\min}^4\bar{d}_g^6}, \frac{m^2}{\bar{d}_{\min}^4\bar{d}_g^6}, \frac{1}{\alpha^2\bar{d}_g^3}, \frac{\bar{d}_\Sigma^2}{\bar{d}_g^2}\right\}\right)$.

- Point 2(b). Note that for $K = \Omega(\frac{\bar{d}_\Sigma^2}{\bar{d}_g^2})$,

$$\begin{aligned}
&\sup_{\mathbf{q},\mathbf{q}'\in\Omega^K}\left(F(\bar{\mathbf{q}}) - F(\bar{\mathbf{q}}')\right) \\
&\leq \max\{C_b, 1\}\cdot O\left(\sqrt{m}\sup_{\mathbf{q},\mathbf{q}'\in\Omega^K}\left(\sum_{j\in V}\sqrt{\bar{d}_j}\left(-\sqrt{\bar{q}_j} - \sqrt{\bar{d}_j - \bar{q}_j} + \sqrt{\bar{q}_j'} + \sqrt{\bar{d}_j - \bar{q}_j'}\right)\right)\right) \\
&\leq O\left(\sqrt{m}\max_{\mathbf{q}'\in\Omega^K}\sum_{j\in V}\sqrt{\bar{d}_j}\left(\sqrt{\bar{q}_j'} + \sqrt{\bar{d}_j - \bar{q}_j'}\right)\right) \\
&= O(m)\,.
\end{aligned}$$

  Hence

$$M_1 = \mathrm{poly}(m) = O(m)\,.$$

- Point 2(c). Note that for $\bar{\mathbf{q}} \in \mathcal{B}_{\mathbf{f}}$, we have $\bar{q}_j = \Theta\left(\frac{\bar{d}_j}{\bar{d}_\Sigma}\right)$, hence

$$M_2 = \max_{\bar{\mathbf{q}}\in\mathcal{B}_{\mathbf{f}}}\max_{j\in V}|f'(\bar{q}_j)| \leq \frac{1}{\bar{d}_{\min}}\left(\frac{\bar{d}_\Sigma}{\bar{d}_g}\right)^{3/2} O(\sqrt{m})\,.$$

  For the special case where $V_{\mathrm{b}} = \emptyset$ hence $\bar{d}_j = 1$ for all $j \in V$, we have $\bar{d}_\Sigma = m$, $\bar{d}_{\min} = 1$, $\bar{d}_g = 1$ and $M_2 = O(m^2)$.

- Point 2(d). Note that for $\bar{\mathbf{q}} \in \mathcal{B}_{\mathbf{f}}$, we have $\bar{q}_j = \Theta\left(\frac{\bar{d}_j}{\bar{d}_\Sigma}\right)$, hence point 2(d) holds.

□

In the following (Lemma 16), we verify Condition 4 for two congestion functions other than the one given in (22).

Let $V_{\mathrm{b}} \subset V$ be the subset of buffer-constrained nodes with scaled buffer sizes $\bar{d}_j \in (0, 1)$, and define

- *Logarithmic congestion function.*

$$\begin{aligned}
f_j(\bar{q}_j) &= c\cdot C_b\cdot\left(\log\left(\frac{\bar{q}_j}{\bar{d}_j}\right) - \log\left(1 - \frac{\bar{q}_j}{\bar{d}_j}\right) - D_b\right), && \forall j \in V_{\mathrm{b}} \\
f_j(\bar{q}_j) &= c\cdot\log\bar{q}_j\,, && \forall j \in V\backslash V_{\mathrm{b}}
\end{aligned} \tag{60}$$

Here C_b and D_b are normalizing constants chosen as follows. Define $\epsilon \triangleq \frac{\delta_K}{\tilde{K}}$ (where δ_K and $\tilde{K}$ were defined in (21)). Let $h_{\mathrm{b}}(\bar{q}) \triangleq \log \bar{q} - \log(1-\bar{q})$ and $h(\bar{q}) \triangleq \log \bar{q}$. Define $C_b \triangleq \frac{h(\epsilon)-h(1/\sum_{j\in V}\bar{d}_j)}{h_{\mathrm{b}}(\epsilon)-h_{\mathrm{b}}(1/\sum_{j\in V}\bar{d}_j)}$ and $D_b \triangleq h_{\mathrm{b}}(1/\sum_{j\in V}\bar{d}_j) - C_b^{-1}h(1/\sum_{j\in V}\bar{d}_j)$. Here $c = \Omega\left(\max\{\frac{1}{\alpha}, m\}\right)$.

- *Linear congestion function.*

$$f_j(\bar{q}_j) = c \cdot \frac{\bar{q}_j}{\bar{d}_j}\,, \qquad \forall j \in V\,, \tag{61}$$

where $c = \Omega\left(\frac{\bar{d}_\Sigma}{\alpha}\right)$.

Lemma 16. *Let $\bar{d}_\Sigma \triangleq \sum_{j\in V}\bar{d}_j$, $\bar{d}_g \triangleq \min\{1, \sum_{j\in V}\bar{d}_j - 1\}$ and $\bar{d}_{\min} \triangleq \min_{j\in V}\bar{d}_j$. The congestion functions* (60) *and* (61) *satisfy the growth conditions (Condition 4) with parameters* (α, K_1, M_1, M_2) *where*

- *Logarithmic congestion function:*

$$K_1 = C\cdot\max\left\{\frac{c^2}{\bar{d}_{\min}^2\bar{d}_g^2 m^2\alpha^2}, \frac{c^2}{\bar{d}_{\min}^2\bar{d}_g^2}, \frac{\bar{d}_\Sigma^2}{\bar{d}_{\min}^2\bar{d}_g^2}\right\}\,, \quad M_1 = C\cdot c\cdot\log m\,, \quad M_2 = C\cdot\frac{\bar{d}_\Sigma}{\bar{d}_{\min}}\cdot c$$

  for a universal constant $C > 0$.
- *Linear congestion function:*

$$K_1 = C\cdot\max\left\{\frac{c}{m\alpha}, \bar{d}_\Sigma^2\right\}\,, \quad M_1 = C\cdot c\,, \quad M_2 = C\cdot m^2$$

  for a universal constant $C > 0$.

Proof of Lemma 16. **Logarithmic function.** Point 1 in Condition 4 is obvious. Now we verify the other points one by one:

- Point 1. It is not hard to see that the congestion functions $(f_j(\bar{q}_j))_{j\in V}$ are strictly increasing and continuously differentiable. For any $K > 0$, we have $f_j(\bar{q}_j) = f_k(\bar{q}_k)$ for any $j, k \in V$ if $q_j = q_k = 0$. As a result, if $q_j = 0$, we have $f_j(\bar{q}_j) \le f_k(\bar{q}_k)$ for any $k \in V$. It can be easily verified that Point 1(b) is also satisfied by any $K > 0$.

  It remains to be shown that there exists $K_1 < \infty$ such that for $K \ge K_1$, we have $f_j(\bar{q}_j) \le f_k(\bar{q}_k)$ for any $j \in V$ if $q_k = d_k$ and $k \in V_{\mathrm{b}}$. To this end, if suffices to check the inequality $f_j(\bar{q}_j) \le f_k(\bar{q}_k)$ for $q_j = d_j$, $q_k = d_k$ where $j \in V\backslash V_{\mathrm{b}}$ and $k \in V_{\mathrm{b}}$: In this case, we have $f_j(\bar{q}_j) \le 0$; for $K = \Omega(\max\{\bar{d}_\Sigma^2, \frac{\bar{d}_\Sigma^2}{\bar{d}_g^2}\})$, we have $C_b = \Theta(1)$, $D_b = O(\log\frac{\bar{d}_\Sigma}{\bar{d}_g})$ hence $f_k(\bar{q}_k) = \Omega(c\cdot\log\frac{\sqrt{K}}{\bar{d}_g}) \ge 0$. Therefore point 1 is satisfied for $K_1 = O(\max\{\bar{d}_\Sigma^2, \frac{\bar{d}_\Sigma^2}{\bar{d}_g^2}\}) = O(\frac{\bar{d}_\Sigma^2}{\bar{d}_g^2})$.

- Point 2(a). For $\mathbf{q}$ such that $\bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$ and $0 < q_j < d_j$ for any $j \in V$, we have, by definition of $\bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$,

$$\text{LHS of (53)} \ge 2m\alpha\,.$$

  On the other hand, we have for $K = \Omega(\frac{\bar{d}_\Sigma^2}{\bar{d}_g^2})$, we have $C_b = \Theta(1)$, hence

$$\text{RHS of (53)} = O\left(\frac{1}{K}\cdot c\cdot C_b\cdot\frac{\sqrt{K}}{\bar{d}_{\min}\bar{d}_g}\right)\,.$$

Here the RHS of (53) is maximized when $q_j = 0$ or $q_j = d_j$. Therefore (53) holds for $K_1 = O\left(\max\left\{\frac{\bar{d}_\Sigma^2}{\bar{d}_g^2}, \frac{c^2}{\bar{d}_{\min}^2 \bar{d}_g^2 m^2 \alpha^2}\right\}\right)$.

For $\mathbf{q}$ such that $\bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$ and $q_j = 0$ or d_j for some $j' \in V$, we have

$$\text{LHS of (53)} \geq \alpha c \cdot \Omega\left(\frac{1}{2}\log K - \log(\bar{d}_{\min}\bar{d}_g) - \log \bar{d}_\Sigma\right),$$

which is obtained by plugging in $q_{j'}$. We also have

$$\text{RHS of (53)} \leq O\left(\frac{1}{K} \cdot c \cdot C_b \cdot \frac{\sqrt{K}}{\bar{d}_{\min}\bar{d}_g} + 1\right),$$

Using the analysis above, for $K = \Omega\left(\max\left\{\frac{\bar{d}_\Sigma^2}{\bar{d}_g^2}, \frac{c^2}{\bar{d}_{\min}^2 \bar{d}_g^2}\right\}\right)$, the first term in the parentheses is $O(1)$. In this case we have RHS of (53) $\leq O(1)$. Therefore when $c = \Omega\left(\frac{1}{\alpha}\right)$, (53) holds for $K = \Omega\left(\left(\frac{\bar{d}_\Sigma}{\bar{d}_{\min}\bar{d}_g}\right)^2\right)$.

Combined, for $c = \Omega\left(\frac{1}{\alpha}\right)$, Point 2(a) holds for $K_1 = \Omega\left(\max\left\{\frac{c^2}{\bar{d}_{\min}^2 \bar{d}_g^2 m^2 \alpha^2}, \frac{c^2}{\bar{d}_{\min}^2 \bar{d}_g^2}, \frac{\bar{d}_\Sigma^2}{\bar{d}_{\min}^2 \bar{d}_g^2}\right\}\right)$.

- Point 2(b). Note that

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

  where the inequality follows from the fact that $\bar{q}_j, \bar{d}_j - \bar{q}_j \in (0,1)$ hence $\bar{q}_j \log \bar{q}_j < 0$ and $(\bar{d}_j - \bar{q}_j)\log(\bar{d}_j - \bar{q}_j) < 0$. Hence

$$M_1 = \text{poly}(c, m) = O(c \cdot \log m).$$

- Point 2(c). For $\bar{\mathbf{q}} \in \mathcal{B}_\mathbf{f}$, we have $\frac{\bar{d}_j}{d_\Sigma} e^{-\frac{4m}{c \cdot C_b}} \leq \bar{q}_j \leq \frac{\bar{d}_j}{d_\Sigma} e^{\frac{4m}{c \cdot C_b}}$. Choose $c \geq \frac{8m}{C_b} = \Omega(m)$, we have

$$M_2 = \max_{\bar{\mathbf{q}} \in \mathcal{B}_\mathbf{f}} \max_j |f'(\bar{q}_j)| \leq \max_{\bar{\mathbf{q}} \in \mathcal{B}_\mathbf{f}} \max_{j \in V} |\bar{q}_j^{-1}| \leq \text{poly}(c, \bar{\mathbf{d}}) = \frac{\bar{d}_\Sigma}{\bar{d}_{\min}} \cdot O(c).$$

- Point 2(d). Note that $\bar{\mathbf{q}} \in \mathcal{B}_\mathbf{f}$, we have $\frac{\bar{d}_j}{d_\Sigma} e^{-\frac{4m}{c \cdot C_b}} \leq \bar{q}_j \leq \frac{\bar{d}_j}{d_\Sigma} e^{\frac{4m}{c \cdot C_b}}$. Given the choice of c derived in the last bullet point, we know point 2(d) holds.

Linear Function. Now we consider the linear congestion function.

- Point 1. It is easy to verify Point 1, therefore we omit the proof.
- Point 2(a). For $\mathbf{q}$ such that $\bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$ and $0 < q_j < d_j$ for any $j \in V$, we have, by definition of $\bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$,

$$\text{LHS of (53)} \geq 2m\alpha\,.$$

We also have

$$\text{RHS of (53)} \leq O\left(\frac{1}{K} \cdot c\right)$$

Therefore (53) holds for $K = \Omega\left(\frac{c}{m\alpha}\right)$. For $\mathbf{q}$ such that $\bar{\mathbf{q}} \in \bar{\mathcal{B}}(\mathbf{f})$ and $q_j = 0$ or d_j for some $j' \in V$, we have

$$\text{LHS of (53)} \geq \alpha c \cdot \Omega\left(\frac{1}{\bar{d}_\Sigma} - \frac{1}{\sqrt{K}}\right)\,,$$

which is obtained by plugging in $q_{j'}$. We also have

$$\text{RHS of (53)} \leq O\left(\frac{1}{K} \cdot c + 1\right)\,,$$

for $K = \Omega(c)$, the first term in the parenthese is $O(1)$, therefore we have RHS of $(53) = O\left(1\right)$. Choosing $c = \Omega\left(\frac{\bar{d}_\Sigma}{\alpha}\right)$, then (53) holds for $K = \Omega(\bar{d}_\Sigma^2)$.

Combined, when choosing $c = \Omega\left(\frac{\bar{d}_\Sigma}{\alpha}\right)$, Point 2(a) holds for $K_1 = O\left(\max\{\frac{c}{m\alpha}, \bar{d}_\Sigma^2\}\right)$.
- Point 2(b). Note that

$$\begin{aligned}
&\sup_{\mathbf{q},\mathbf{q}' \in \Omega^K} \left(F(\bar{\mathbf{q}}) - F(\bar{\mathbf{q}}')\right) \\
&= O\left(c \sup_{\mathbf{q},\mathbf{q}' \in \Omega^K} \left(\sum_{j \in V} \bar{d}_j^{-1} \left(\bar{q}_j^2 - (\bar{q}_j')^2\right)\right)\right) \\
&\leq O\left(c \max_{\mathbf{q}' \in \Omega^K} \sum_{j \in V} \bar{d}_j^{-1} \left(\bar{q}_j^2\right)\right) \\
&= O(c)\,.
\end{aligned}$$

Hence

$$M_1 = \text{poly}(c) = O(c)\,.$$

- Point 2(c). For $\bar{\mathbf{q}} \in \mathcal{B}_{\mathbf{f}}$, we have $\bar{q}_j = \Theta\left(\frac{\bar{d}_j}{d_\Sigma}\right)$, hence

$$M_2 = \max_{\bar{\mathbf{q}} \in \mathcal{B}_{\mathbf{f}}} \max_{j \in V} |f'(\bar{q}_j)| = O(c)\,.$$

- Point 2(d). For $\bar{\mathbf{q}} \in \mathcal{B}_{\mathbf{f}}$, we have $\bar{q}_j = \Theta\left(\frac{\bar{d}_j}{d_\Sigma}\right)$. Hence point 2(d) holds.

$\square$

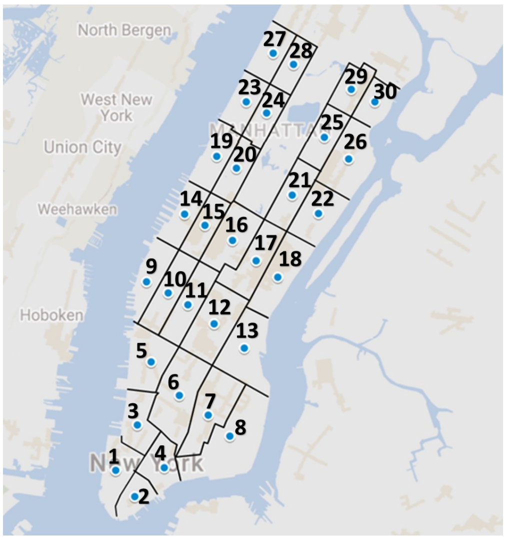

Figure 3: A 30 location model of Manhattan below 110-th street, excluding the Central Park. (tessellation is based on Buchholz (2015))

D Appendix to Section 7.1

In this section we provide a full description of our simulation environment and the benchmark we employ.

D.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[25] k.

• *Graph topology.* We consider a 30-location model of Manhattan below 110-th street excluding Central Park (see Figure 3), 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). 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).[26] We estimate travel times between location pairs using Google Maps.[27]

• *Number of cars, and steady state upper bound.*

[25]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.

[26]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.

[27]We extract the pairwise travel time between region centroids (marked by the dots in Figure 3) using Google Maps, denoted by $\hat{D}_{ij}$'s $(i, j = 1, \cdots, 30)$. We use $\hat{D}_{jk}$ 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 [28] of $\tilde{D}_{ij} = \max\{\hat{D}_{ij}, 2 \text{ minutes}\}$. The average travel time across all demand is 13.1 minutes, and the average pickup time is about 4 minutes (it is policy dependent).

— *Excess supply.* We use as a baseline the fluid requirement K_{fl} on number of cars needed to achieve optimal payoff. A simple workload conservation argument (using Little's Law) gives the fluid requirement as follows. Applying Little's Law, if the optimal solution $\mathbf{z}^*$ of SPP (32)-(34) is realized as the average long run assignment, the mean number of cars who are currently occupied, i.e. serving or picking up customers is $\sum_{j,k\in V}\sum_{i\in\mathcal{P}(j)} D_{ijk}\cdot z^*_{ijk}$, for $D_{ijk} \triangleq \tilde{D}_{ij} + \hat{D}_{jk}$, where $\tilde{D}_{ij}$ is the pickup time from i to j and $\hat{D}_{jk}$ 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^{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^{SPP}. A tighter upper bound on the steady state performance is then the value of the SPP (32)-(34) with the additional supply constraint

$$\sum_{j,k\in V}\sum_{i\in\mathcal{P}(j)} D_{ijk}\cdot z_{ijk} \le K\,.$$

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, Freund, and Lykouris 2016, Ozkan and Ward 2016): Let $\mathbf{z}^*$ be a solution of SPP. When a type (j,k) demand arrives at location j, the randomized fluid-based policy dispatches from location $i \in \mathcal{P}(j)$ with probability z^*_{ijk}/ϕ_{jk}.
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.

D.2 The Excess Supply Case

We simulate the (stationary) system from 8 a.m. to 12 p.m. with 100 randomly generated initial states[29]. The simulation results on performance are shown in Figure 4. 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 105% of W^{SPP} (here W^{SPP} is again an upper bound on the steady state performance[30]), while the static policy and greedy policy only achieve 65% and 68% of W^{SPP}, respectively. The performance of the static policy converges very slowly to W^{SPP}, leading to poor transient performance.[31] The performance of the greedy policy quickly deteriorates over time because it ignores the flow balance constraints and creates huge geographical imbalances

[29]We first uniformly sample 100 points from the simplex $\{\mathbf{q} : \sum_{i\in V} q_i = K\}$, 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.

[30]W^{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^{SPP} as a performance measure, with the understanding that it may exceed 1 at early times.

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

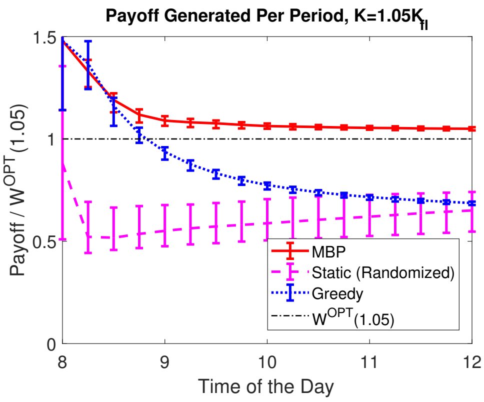

Figure 4: Per period payoff under the MBP policy, static fluid-based policy and greedy policy (with 90% confidence intervals), relative to W^{SPP}.

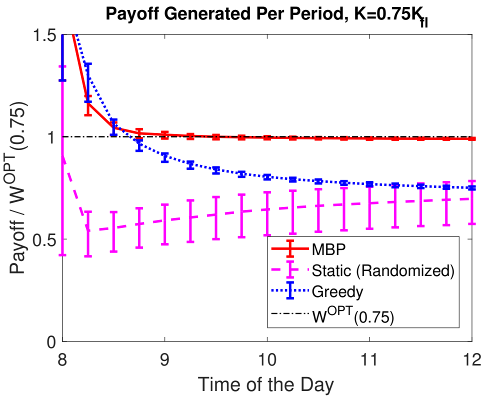

Figure 5: Per period payoff under the modified MBP policy, static fluid-based policy and greedy policy (with 90% confidence intervals), relative to $W^{\text{SPP}}(0.75)$, the value of SPP along with constraint (30) for $K = 0.75K_{\text{fl}}$.

in supply availability.

D.3 The Scarce Supply Case

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

Figure 5 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 99% of the benchmark $W^{\text{SPP}}(0.75)$ over 4 hours, while the static policy and greedy policy only achieves 69% 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 (31) in the SPP along with (31) (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.