Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Blind Dynamic Resource Allocation in Closed Networks via Mirror Backpressure

Yash Kanoria, Pengyu Qian

To cite this article:
Yash Kanoria, Pengyu Qian (2023) Blind Dynamic Resource Allocation in Closed Networks via Mirror Backpressure. Management Science

Published online in Articles in Advance 27 Sep 2023

. https://doi.org/10.1287/mnsc.2023.4934

https://pubsonline.informs.org/journal/mnsc

MANAGEMENT SCIENCE
Articles in Advance, pp. 1–18
ISSN 0025-1909 (print), ISSN 1526-5501 (online)

Blind Dynamic Resource Allocation in Closed Networks via Mirror Backpressure

Yash Kanoria,[a] **Pengyu Qian**[b,*]

[a] Decision, Risk and Operations Division, Columbia Business School, New York, New York 10027; [b] Daniels School of Business, Purdue University, West Lafayette, Indiana 47907

*Corresponding author

Contact: ykanoria@columbia.edu, https://orcid.org/0000-0002-7221-357X (YK); qianp@purdue.edu, https://orcid.org/0000-0002-1759-1009 (PQ)

Received: August 14, 2020
Revised: February 27, 2022; March 22, 2023
Accepted: March 31, 2023
Published Online in Articles in Advance: September 27, 2023

https://doi.org/10.1287/mnsc.2023.4934

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 that 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 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 in time). Under mild conditions, we propose MBP policies that are provably near optimal. Specifically, our policies lose at most $O\left(\frac{K}{T}+\frac{1}{K}+\sqrt{\eta K}\right)$ 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 demand process' average rate of change per customer arrival. An adaptation of MBP is found to perform well in numerical experiments based on data from NYC Cab.

History: Accepted by Gabriel Weintraub, revenue management and market analytics.
Funding: Y. Kanoria was supported by the National Science Foundation's Division of Civil, Mechanical, and Manufacturing Innovation [Grant CMMI-1653477].
Supplemental Material: The data files and online appendices are available at https://doi.org/10.1287/mnsc.2023.4934.

Keywords: control of queueing networks • backpressure • mirror descent • no-underflow constraint

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 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 (Braverman et al. 2019, Banerjee et al. 2021). 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 *the system parameters (demand arrival rates) are precisely known*, and most existing works assume *time invariant* parameters.

In this paper, we relax *both* assumptions. We propose a family of *simple* 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 time-varying demand arrival rates. Strong performance in simulations backs up our theoretical findings.

1.1. Informal Description of Our Model

Our main setting is one where the control levers include entry control and flexible assignment of resources, with time-varying demand arrival rates. Later we allow dynamic pricing control and show that our machinery and guarantees extend seamlessly. For simplicity, we introduce here the special case of our main model with entry control only. We consider a closed queueing network that consists of a set of nodes (locations) V, and a fixed number K of supply units that circulate in the system. Demand units with different origin-destination pairs (j,k)

arrive stochastically over slotted time with some time-varying arrival rates which are unknown to the controller. The controller dynamically decides whether to admit each incoming demand unit. Each control decision to admit a demand unit has two effects: It generates a certain *payoff* w_{jk} depending on the origin-destination pair of the demand unit, and it causes a supply unit to *relocate* from the origin j to the destination k instantaneously. 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 in its pick-up neighborhood, 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 Online Appendix H.1. The intuition is that some nodes have no available supply an $\Omega(1)$ fraction of the time in steady state under the greedy policy. Furthermore, if demand arrival rates are imperfectly known, any state independent policy (such as that of Banerjee et al. 2021) generically suffers a steady-state optimality gap per demand unit of $\Omega(1)$ (Banerjee et al. 2018, proposition 4; 2021).

1.2. Our Control Policy

The system state at time t is the vector of queue lengths $\mathbf{q}[t] = [q_1[t], \ldots, q_{|V|}[t]]^\top$, which sums up to the total supply $\mathbf{1}^\top \mathbf{q}[t] = K$; we work with a normalized queue length vector $\overline{\mathbf{q}}$ satisfying $\mathbf{1}^\top \overline{\mathbf{q}}[t] = 1$. Our proposed *mirror backpressure* (MBP) policy makes entry control decisions according to the following simple rule: Admit a demand unit with origin node j and destination node k if and only if the *score* $w_{jk} + f(\overline{q}_j[t]) - f(\overline{q}_k[t]) \geq 0$ and $\overline{q}_j[t] > 0$. Here, $f(\cdot) \triangleq -\sqrt{m} \cdot \overline{q}_j^{-\frac{1}{2}}$ is a suitably chosen *congestion function*, a monotone increasing function that causes the policy to be generous in allowing use of supply from long queues while protecting supply in near-empty queues. See Figures 1 and 2 for illustrations. The MBP policy is agnostic to demand arrival rates.

Figure 1. Our MBP Policy Admits a Demand with Origin j and Destination k if and Only if the Previously Illustrated Score Is Nonnonnegative and j Has at Least One Supply Unit

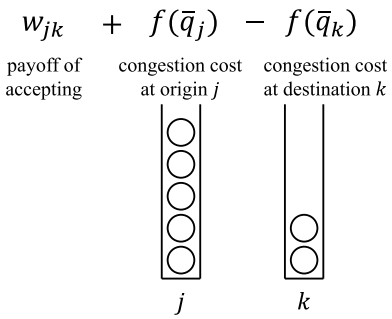
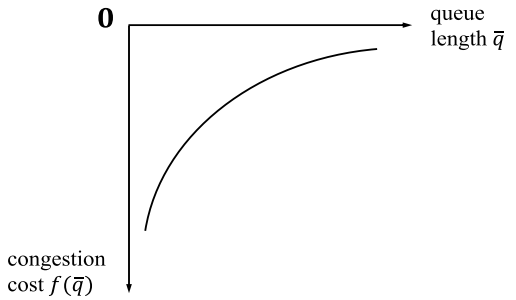


Figure 2. Example of a Congestion Function (a Mapping from Queue Lengths to Congestion Costs) That Aggressively Protects Supply Units in Near-Empty Queues



1.3. Performance Guarantee

We show that, under a mild connectivity assumption on the network, the MBP policy is near optimal. Specifically, we show that our policies lose payoff (per demand unit) at most $O\left((K/T) + (1/K) + \sqrt{\eta K}\right)$ 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, and η is the demand process' average rate of change per customer arrival. Our result is *nonasymptotic*, that is, our performance guarantee holds for finite K and T and thus covers both transient and steady state performance. In particular, for stationary demand arrivals, taking $T \to \infty$, we obtain a steady-state optimality gap of $O(1/K)$. 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 with the number of supply units K. Our policies retain their good performance if the demand arrival rate's average rate of change over K periods, that is, the term ηK, is small. We find that our bound is invariant to system size in a relevant scaling regime (the large market regime) where the number of supply units K increases proportionally to the demand arrival rates (see the discussion after Theorem 1). In this regime, $T^{\text{real}} \triangleq T/K$ is an invariant as $K \to \infty$, which can be interpreted as the time horizon measured in physical time. Let $\zeta \triangleq \eta K$ be the average rate of change of demand with respect to physical time. We can rewrite our bound on the optimality gap as

$$O\left(\frac{1}{T^{\text{real}}} + \frac{1}{K} + \sqrt{\zeta}\right) \overset{K\to\infty}{\longrightarrow} O\left(\frac{1}{T^{\text{real}}} + \sqrt{\zeta}\right).$$

1.4. Motivation for Our Control Policy

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 that maximizes the myopic payoff inclusive of congestion costs. Concretely, in the special case where entry control is the only leverage, 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, and so on (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. However, this property breaks down when the so-called "no-underflow constraint" binds, making it challenging to use backpressure in our setting.

The *no-underflow constraint* is that each decision to admit a demand unit needs to be backed by an available supply unit at the pick-up node of the demand. This constraint binds in our setting under backpressure because we model nonzero payoffs from serving a customer, as a result of which the congestion-adjusted myopic payoff can be positive even if the origin queue is empty (see Online Appendix H.2 for a discussion). Moreover, several popular workarounds to this issue fail in our setting (see Section 1.7).

Our Mirror Backpressure (MBP) policy generalizes the celebrated backpressure (BP) policy. Whereas BP uses the queue lengths as congestion costs, MBP employs a flexibly chosen *congestion function* $f(\cdot)$ to translate from queue lengths to congestion costs. 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 use congestion functions which moreover increase steeply as the queue length approaches buffer capacity, to prevent buffer overflow (Section 5.1).

1.5. Analytical Approach

We show 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, which generalizes the SGD property of backpressure. We develop a general machinery to prove performance guarantees for MBP: We use the antiderivative of the chosen congestion function as the Lyapunov function in our analysis and adapt the Lyapunov drift method from the network control literature to obtain sharp bounds on the suboptimality caused by the no-underflow constraint. Our analysis exploits the structure of the platform's continuum limit optimization problem (see Section 4). Our work fits into the broad literature on the control of stochastic processing networks (Harrison 2003).

1.6. Applications

Our models include a number of ingredients that are central in many applications. We illustrate its versatility by discussing the application to shared transportation systems (Section 6) and the application to scrip systems (Online Appendix G). These applications and the relevant settings in the paper are summarized in Table 1.

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). For ride-hailing, our joint pricing-assignment (JPA) model is relevant in regions such as North America, and our entry-assignment control model is relevant in regions where dynamic pricing is undesirable like in China. We perform simulations of ride-hailing and find that our MBP policy, suitably adapted to account for positive travel times, performs well (Section 6.1).

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), for example, 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 the agent 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 Online Appendix G, 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.

1.7. Literature Review

1.7.1. MaxWeight/Backpressure Policy. Backpressure (also known as MaxWeight; 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), utility maximization (Eryilmaz and Srikant 2007), and so on. Attractive features of MaxWeight/backpressure

Table 1. Summary of Applications of Our Model

Application	Control lever	Corresponding setting in this paper
Ride-hailing in the United States and Europe	Pricing and dispatch	Joint pricing-assignment
Ride-hailing in China	Admission and dispatch	Joint entry-assignment
Bike sharing	Reward points	Pricing (finite buffer queues)
Scrip systems	Admission and provider selection	Joint entry-assignment

policies include their simplicity and provably good performance and that arrival/service rate information is not required beforehand.

The main challenge in using backpressure in settings with payoffs is the no-underflow constraints, as described earlier. Several works make strong assumptions to ensure the constraint does not bind: For example, Dai and Lin (2005) assume that the network satisfies a so-called extreme allocation available (EAA) condition; 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 no-underflow constraint does bind, but the payoffs are generated only by the output nodes. In our setting, payoffs are essential (there is value generated by serving a customer) and can be generated by any node. Therefore, the no-underflow constraint binds, and none of the aforementioned assumptions hold for our network. Another workaround is a machinery that introduces *virtual queues* (Jiang and Walrand 2009). The idea is to introduce a "fake" supply unit into the network each time the constraint binds to preserve the SGD property of backpressure. However, in open queueing networks, these fake supply units eventually leave the system and therefore have a small effect (under appropriate assumptions). In our closed network setting, these fake supply units, once created, never leave and therefore would build up in the system, irreparably damaging performance.

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 et al. (2018), which shows the large deviations optimality of "scaled" MaxWeight policies.

Similar to MBP, several works use nonlinear functions of queue lengths for decision making to improve on the performance of Backpressure in a variety of contexts. Walton (2015) proposes concave switching policies that generalize backpressure to address a weakness of backpressure in fixed route multihop networks, namely, that the number of queues it needs to maintain grows rapidly in network size. Neely (2006) uses exponential functions of queue length as congestion functions to achieve the optimal delay-utility tradeoff. Gupta and Radovanović (2020) use nonlinear functions of the state variables in the context of online stochastic bin-packing to obtain distribution-oblivious algorithms with sublinear additive suboptimality. Gupta and Radovanović (2020) also identify the connection to mirror descent as we do for MBP.

1.7.2. Mirror Descent. Mirror descent (MD) is a generalization of the gradient descent algorithm for optimization (Nemirovsky and Yudin 1983, Beck and Teboulle 2003). Recently, there have been several works that use online MD to solve other online decision-making problems, including the k-server problem (Bubeck et al. 2018) and various online packing and covering problems (Agrawal and Devanur 2014, Gupta and Molinaro 2016).

1.7.3. Applications: Shared Transportation, Scrip Systems. Most of the ride-hailing literature studied controls that require the exact knowledge of system parameters: Özkan and Ward (2020) 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, and Banerjee et al. (2021) adopted the Gordon-Newell closed queueing network model and considered various controls that maximize throughput, welfare, or revenue. Balseiro et al. (2021) 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 that are sufficiently smooth in space and time in a model with no demand stochasticity. Banerjee et al. (2018), who assume a near balance condition[1] on demands and equal pickup costs, may be the only paper in this space that does not require knowledge of system parameters. Compared with Banerjee et al. (2021), who obtain a steady-state optimality gap of $O\left(\frac{1}{K}\right)$ (in the absence of travel times) assuming *perfect* knowledge of demand arrival rates that 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 $O((K/T)+(1/K)+\sqrt{\eta K})$ for a finite number of arrivals T and average changes of up to η per arrival in demand arrival rates. Some of these papers are able to formally handle travel delays: Braverman et al. (2019) and Banerjee et al. (2018, 2021) prove theoretical results

for the setting with independent and identically distributed (i.i.d.) geometric/exponential travel delays; Ma et al. (2019) consider deterministic travel delays. Conversely, Balseiro et al. (2021) ignores travel delays in their theory and later heuristically adapt their policy to accommodate travel delay (the present paper follows a similar approach). Özkan and Ward (2020) is the only paper among these that (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 Online Appendix G.

1.7.4. Other Related Work. A related stream of research studies online stochastic bipartite matching (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 (Gallego and Van Ryzin 1994, Talluri and Van Ryzin 2006, Bumpensanti and Wang 2020). 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.

1.8. Organization of the Paper

The remainder of our paper is organized as follows. Section 2 presents our main model of joint entry-assignment control with time-varying demand arrival rates and the platform objective. Section 3 introduces the MBP policy and presents our main theoretical result, that is, a performance guarantee for the MBP policies. Section 4 outlines the proof of our main result. In Section 5, we provide MBP policies for the joint pricing-assignment control setting, demonstrating the versatility of our approach. In Section 6, we discuss the applications to shared transportation systems.

1.8.1. 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 $\mathbb{R}^n_+$ to denote the nonnegative orthant, and $\mathbb{R}^n_{++}$ to denote the positive orthant. We use $\mathbf{e}_i$ to denote the ith unit column vector with the ith coordinate being one and all other coordinates being zero, and $\mathbf{1}$ ($\mathbf{0}$) to denote the all one (zero) column vector, where the dimension of the vector will be indicated in the superscript when it is not clear from the context.

2. Model: Joint Entry-Assignment Control

In this section, we formally define our model of joint entry-assignment control in closed queueing networks. We consider a finite-state Markov chain model with slotted time $t = 0, 1, 2, \ldots$, where a fixed number (denoted by K) of identical *supply units* circulate among 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 thus far.

2.1. Queues (System State)

At each node $j \in V$, there is an infinite-buffer queue of supply units. (Section 5.1 shows how to seamlessly incorporate finite-buffer queues.) The *system state* is the vector of queue lengths at time t, which we denote by $\mathbf{q}[t] = [q_1[t], \ldots, q_m[t]]^\top$. 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 $\Omega \triangleq \{\mathbf{q} : \mathbf{q} \in \mathbb{R}^m_+, \mathbf{1}^\top \mathbf{q} = 1\}$.

2.2. Demand Types and Time-Varying Arrival Process

We assume exactly one demand unit (customer) arrives at each period t and denote its abstract *type* by $\tau[t] \in \mathcal{T}$, and the type for the demand unit is drawn from distribution $\boldsymbol{\phi}^t = (\phi^t_\tau)_{\tau \in \mathcal{T}}$, independent of demands in earlier periods.[2] The demand arrival rate (i.e., type distribution) can be time varying. Importantly, the system can observe the type of the arriving demand at the beginning of each time slot, but *the probabilities (arrival rates)* $\boldsymbol{\phi}^t$ *are not known.* Thus, we substantially relax the assumption in previous works that the system has exact knowledge of demand arrival rates (Özkan and Ward 2020, Balseiro et al. 2021, Banerjee et al. 2021).

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.)

2.3. Temporal Uncertainty of Demand Arrival Rates

We define the following notion of η-*slowly varying demand* that characterizes the average amount of change of demand arrival rates over a finite time horizon.

Definition 1. We say that demand arrival rates vary η-slowly over a finite horizon T if

$$\frac{1}{T-1}\sum_{t=1}^{T-1} \|\boldsymbol{\phi}^{t+1} - \boldsymbol{\phi}^t\|_1 \le \eta.$$

2.4. Control and Payoff

At time t, after observing the demand type $\tau[t]=\tau$, the system makes a decision

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

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 (1) 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 $v[t]=w_{j\tau k}$. 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 nonnegative by definition, we require the following *no-underflow constraint* to be met at any t:

$$x_{j\tau k}[t]=0\quad\text{if}\quad q_j[t]=0. \tag{2}$$

As a convention, let $x_{j\tau' k}=0$ if $\tau'\neq\tau$.

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

2.5. System Dynamics and Objective

The dynamics of system state $\mathbf{q}[t]$ is as follows:

$$\mathbf{q}[t+1]=\mathbf{q}[t]+\sum_{\tau\in\mathcal{T},j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)}(-\mathbf{e}_j+\mathbf{e}_k)x_{j\tau k}[t], \tag{3}$$

that is, a supply unit is relocated from j to k. 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[3]

$$W_T^\pi\triangleq\min_{\mathbf{q}\in\Omega_K}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[v^\pi[t]\,|\,\mathbf{q}[0]=\mathbf{q}],$$
$$W_T^*\triangleq\sup_{\pi\in\mathcal{U}}\max_{\mathbf{q}\in\Omega_K}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[v^\pi[t]\,|\,\mathbf{q}[0]=\mathbf{q}]. \tag{4}$$

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

$$W^\pi\triangleq\liminf_{T\to\infty}W_T^\pi,\qquad W^*\triangleq\limsup_{T\to\infty}W_T^*. \tag{5}$$

We measure the performance of a control policy π by its finite- and infinite-horizon per-customer *optimality gap* ("loss"), defined, respectively, as

$$L_T^\pi=W_T^*-W_T^\pi\qquad\text{and}\qquad L^\pi=W^*-W^\pi. \tag{6}$$

We consider the worst-case initial system state when evaluating a given policy and the best initial state for the optimal benchmark; see (4). 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 primitives $(\{\boldsymbol{\phi}^t\}_{t\le T},\mathcal{P},\mathcal{D})$.

Condition 1 (Strong Connectivity of $(\{\boldsymbol{\phi}^t\}_{t\le T},\mathcal{P},\mathcal{D})$). *For any demand arrival rates $\boldsymbol{\phi}$, define the connectedness of triple $(\boldsymbol{\phi},\mathcal{P},\mathcal{D})$ as*

$$\alpha(\boldsymbol{\phi},\mathcal{P},\mathcal{D})\triangleq\min_{S\subsetneq V,S\neq\emptyset}\sum_{\tau\in\mathcal{P}^{-1}(S)\cap\mathcal{D}^{-1}(V\setminus S)}\phi_\tau. \tag{7}$$

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 any $1\le t\le T$, $(\boldsymbol{\phi}^t,\mathcal{P},\mathcal{D})$ is strongly connected, namely, that $\alpha(\boldsymbol{\phi}^t,\mathcal{P},\mathcal{D})>0$.

The strong connectivity of $(\boldsymbol{\phi},\mathcal{P},\mathcal{D})$ is equivalent to requiring that with (stationary) demand arrival rates $\boldsymbol{\phi}$, for every ordered pair of nodes (j,k), there is a sequence of demand types with positive arrival rates and corresponding pick-up and drop-off nodes that would take a supply unit from j eventually to k.

We conclude this section with the observation that the main assumption of Banerjee et al. (2018) is automatically violated in our setting.

Remark 1. The complete resource pooling (CRP) condition imposed in Banerjee et al. (2018, assumption 3) is automatically violated in the following subclass of our model. Consider our setup including Condition 1, where each demand type $\tau=(i,j)$ corresponds to an origin-destination pair and that $\mathcal{P}(i,j)=\{i\}$, $\mathcal{D}(i,j)=\{j\}$. The CRP condition can be stated as follows: for each subset of nodes $S\subsetneq V,S\neq\emptyset$, the "net demand" $\mu_S\triangleq\sum_{i\in S}\sum_{j\in V\setminus S}\phi_{ij}$ is less than the "net supply" $\lambda_S\triangleq\sum_{j\in V\setminus S}\sum_{i\in S}\phi_{ji}$, that is, $\mu_S<\lambda_S$. Clearly, any demand arrival rates $\boldsymbol{\phi}$ violate CRP, because if $\mu_S<\lambda_S$ for some $S\subsetneq V,S\neq\emptyset$, then this means that $\mu_{V\setminus S}>\lambda_{V\setminus S}$ (given that $\mu_{V\setminus S}=\lambda_S$ and $\lambda_{V\setminus S}=\mu_S$ by definition), that is, CRP is violated.

3. 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. MBP Policies

We propose a family of online control policies that we call MBP policies. Each member of the MBP family is specified by a mapping of normalized queue lengths (which we will define later) $\mathbf{f}(\overline{\mathbf{q}}) : \Omega \to \mathbb{R}^m$, where $\mathbf{f}(\overline{\mathbf{q}}) \triangleq [f(\overline{q}_1), \ldots, f(\overline{q}_m)]^\top$ and f is a monotone increasing function.[4] 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.

In this section we will state our main result for the congestion function:

$$f(\overline{q}_j) \triangleq -\sqrt{m} \cdot \overline{q}_j^{-\frac{1}{2}}, \tag{8}$$

and postpone the results for other choices of congestion functions to Online Appendix D (see also Remark 2). We will later clarify the precise role of the congestion function and show that it is related to the mirror map in mirror descent (Beck and Teboulle 2003). 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 instance, we use a different congestion function for the setting in Section 5.1 where there are additional buffer capacity constraints.

For technical reasons, we need to keep $\overline{\mathbf{q}}$ in the *interior* of the normalized state space Ω; that is, we need to ensure that all normalized queue lengths remain positive. This is achieved by defining the normalized queue lengths $\overline{\mathbf{q}}$ as

$$\overline{q}_i \triangleq \frac{q_i + \delta_K}{\tilde{K}} \quad \text{for} \quad \delta_K \triangleq \sqrt{K} \quad \text{and} \quad \tilde{K} \triangleq K + m\delta_K. \tag{9}$$

This definition leads to $\mathbf{1}^\top \overline{\mathbf{q}} = 1$ and therefore $\overline{\mathbf{q}} \in \Omega$.

Our proposed MBP policy for the joint entry-assignment control problem is given in Algorithm 1. MBP serves a demand of type τ using a supply unit at j^* and relocate it to k^* if and only if

$$(j^*, k^*) = \underset{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)}{\arg\max}\ w_{j\tau k} + f(\overline{q}_j) - f(\overline{q}_k), \tag{10}$$

and that $w_{j^*\tau k^*} + f(\overline{q}_{j^*}) - f(\overline{q}_{k^*})$ is nonnegative, *and* the origin node j^* has at least one supply unit (see Figure 1 for illustration of the score in Section 1). The score in (10) is nonnegative if and only if the payoff $w_{j\tau k}$ of serving the demand outweighs the difference of congestion costs (given by $f(\overline{q}_k)$ and $f(\overline{q}_j)$) between the dropoff node k and the pickup node 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 (Figures 1 and 2). The policy is not only completely blind but also semilocal; that is, it only uses the queue lengths at the origin and destination. The congestion cost (8) increases with queue length (as required) and furthermore decreases 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 no-underflow Constraint (2).

Algorithm 1 (MBP Policy for Joint Entry-Assignment Control)
At the start of period t, the system observes demand type $\tau[t] = \tau$:

$(j^*, k^*) \leftarrow \arg\max_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} w_{j\tau k} + f(\overline{q}_j[t]) - f(\overline{q}_k[t])$
if $w_{j^*\tau k^*} + f(\overline{q}_{j^*}[t]) - f(\overline{q}_{k^*}[t]) \geq 0$ **and** $q_{j^*}[t] > 0$ **then**
  $x_{j^*\tau k^*}[t] \leftarrow 1$, that is, serve the incoming demand using a supply unit from j^* and relocate it to k^*;
else
  $x_{j^*\tau k^*}[t] \leftarrow 0$, that is, drop the incoming demand;
end

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

3.2. Performance Guarantee for MBP Policies

We now formally state the main performance guarantee of our paper for the joint entry-assignment control model introduced in Section 2. We will outline the proof in Section 4 and extend the result to the dynamic pricing setting in Section 5.

Theorem 1. *Consider a set of m nodes and any sequence of demand arrival rates $\{\boldsymbol{\phi}^t\}_{t \leq T}$ that satisfy Condition 1 and vary η-slowly (Definition 1). Define $\alpha_{\min} \triangleq \min_{1 \leq t \leq T} \alpha(\boldsymbol{\phi}^t) > 0$. Then there exists $K_1 = \mathrm{poly}(m, (1/\alpha_{\min}))$, and a universal constant $C < \infty$, such that the following holds.*[5] *For the congestion function $f(\cdot)$ defined in (8), for any $K \geq K_1$, the following finite-horizon guarantee holds for Algorithm 1:*

$$L_T^{\mathrm{MBP}} \leq M_1 \left(\frac{K}{T} + \sqrt{\eta K} \right) + M_2 \frac{1}{K},$$
$$\text{for } M_1 \triangleq Cm \text{ and } M_2 \triangleq Cm^2.$$

Corollary 1. *When the demand arrivals are stationary ($\eta = 0$), for any $K \geq K_1$, the following infinite-horizon guarantee holds for Algorithm 1:*

$$L^{\mathrm{MBP}} \leq M_2 \frac{1}{K}, \qquad \text{for } M_2 = Cm^2.$$

Remark 2. In Section 5, we obtain results similar to Theorem 1 for the dynamic pricing setting (Theorem 3). In Online Appendix D (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.

There are several attractive features of the performance guarantee provided by Theorem 1 for the simple and practically appealing MBP 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 $\{\boldsymbol{\phi}^t\}$, so it is a very attractive feature of the policy that it does not need any estimate of $\{\boldsymbol{\phi}^t\}$ whatsoever. It is worth noting that the consequent bound of $O(1/K)$ on the steady state optimality gap remarkably matches that provided by Banerjee et al. (2021) even though MBP requires *no* knowledge of $\boldsymbol{\phi}$, whereas the policy of Banerjee et al. (2021) requires *exact* knowledge of $\boldsymbol{\phi}$. (However, our constant is quadratic in the number of nodes m, whereas the constant in the other paper is linear in m.) As shown in Banerjee et al. (2018, proposition 4), if the estimate of demand arrival rates is imperfect, any state independent policy (such as that of Banerjee et al. (2021)) generically suffers a long run (steady state) per customer optimality gap of $\Omega(1)$ (as $K \to \infty$).

(2) **Guarantee on transient performance.** In contrast with Banerjee et al. (2021), who provide 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 with 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 nontrivial fraction of cars are busy at any time (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, whereas $T = \Theta(K \cdot T^{\text{real}})$ because there are $\Theta(K)$ arrivals per unit of physical time, and $\zeta \triangleq \eta K$ is the average rate of change of $\boldsymbol{\phi}$ with respect to physical time Hence, we can rewrite our performance guarantee as

$$W_T^* - W_T^{\text{MBP}} \le M\left(\frac{1}{T^{\text{real}}} + \frac{1}{K} + \sqrt{\zeta}\right) \overset{K\to\infty}{\longrightarrow} M\left(\frac{1}{T^{\text{real}}} + \sqrt{\zeta}\right),$$

which is small when $T^{\text{real}} \to \infty$ and $\zeta = o(1)$.

(3) **Guarantee for time-varying arrivals.** Our bound shows that MBP is near optimal when the demand's average rate of change is small ($\eta K = o(1)$) and that the performance guarantee of MBP degrades gracefully as ηK increases. If the demand arrival rates remain stationary for blocks of time, for example, the first half of the horizon has one stationary arrival rate matrix and the second half of the horizon has another stationary arrival rate matrix, then applying Corollary 1 to each contiguous block of time with stationary demand could yield a better guarantee than directly applying Theorem 1 to the entire horizon.

(4) **Flexibility in the choice of congestion function.** Because of the richness of the class of congestion functions covered in Online Appendix D that generalize Theorem 1, the system controller now has the additional flexibility to choose a suitable congestion function $f(\cdot)$. 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 6.1.

4. Proof of Theorem 1

In this section, we provide the key propositions and 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 (Neely 2010).

We first define a sequence of deterministic optimization problems that arise in the continuum limit: the *static planning problem* (SPP) (Harrison 2003, Dai and Lin 2005), whose values we use to upper bound the optimal finite (and infinite) horizon per period W_T^* (and W^*) defined in (4) and (5). The SPP is a linear program (LP) defined for any demand arrival rates $\boldsymbol{\phi}$:

$$\text{SPP}(\boldsymbol{\phi}): \ \text{maximize}_{\mathbf{x}} \sum_{\tau\in\mathcal{T}, j\in\mathcal{P}(\tau), k\in\mathcal{D}(\tau)} w_{j\tau k}\cdot\phi_\tau\cdot x_{j\tau k} \tag{11}$$

$$\text{s.t.} \sum_{\tau\in\mathcal{T}, j\in\mathcal{P}(\tau), k\in\mathcal{D}(\tau)} \phi_\tau\cdot x_{j\tau k}(\mathbf{e}_j - \mathbf{e}_k) = \mathbf{0} \quad \text{(flow balance)}, \tag{12}$$

$$\sum_{j\in\mathcal{P}(\tau), k\in\mathcal{D}(\tau)} x_{j\tau k} \le 1,\ x_{j\tau k} \ge 0, \quad \forall j,k\in V,\ \tau\in\mathcal{T}.\ \text{(demand constraint)}. \tag{13}$$

One interprets $x_{j\tau k}$ as the fraction of type τ demand that is served by pickup location j and dropoff location k, and the objective (11) as the rate at which payoff is generated under the fractions $\mathbf{x}$. In the SPP (11–13), 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 12) and that $\mathbf{x}$ are indeed fractions (Constraint 13). Let $W^{\text{SPP}(\boldsymbol{\phi})}$ be the optimal value of SPP($\boldsymbol{\phi}$). The following proposition formalizes that, the optimal finite horizon per customer payoff W_T^* cannot be much larger than $W^{\text{SPP}(\overline{\boldsymbol{\phi}})}$, where $\overline{\boldsymbol{\phi}} \triangleq \frac{1}{T}\sum_{t=0}^{T-1}\boldsymbol{\phi}^t$.

Proposition 1. *For any horizon $T < \infty$, any K and any starting state $\mathbf{q}[0]$, the finite horizon and steady-state average payoff W_T^*, W^* are upper bounded as*

$$W_T^* \le W^{\text{SPP}(\overline{\boldsymbol{\phi}})} + m\frac{K}{T}, \quad W^* \le W^{\text{SPP}(\overline{\boldsymbol{\phi}})}. \tag{14}$$

We obtain the finite horizon upper bound to W_T^* in (14) by slightly relaxing the flow constraint (12) to accommodate the fact that flow balance need not be exactly satisfied over a finite horizon. The proof is in Online Appendix A.

Our proposed MBP policy and its analysis is closely related to the (partial) dual of the SPP($\boldsymbol{\phi}^t$):

$$\text{minimize}_{\mathbf{y}}\, g^t(\mathbf{y}), \quad \text{for } g^t(\mathbf{y}) \triangleq \sum_{\tau\in\mathcal{T}} \phi_\tau^t \max_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} (w_{j\tau k} + y_j - y_k)^+. \tag{15}$$

Here, $(x)^+ \triangleq \max\{0,x\}$, and $\mathbf{y}$ are the dual variables corresponding to the flow balance constraints (12) and have the interpretation of "congestion costs" (Neely 2010); that is, y_j can be thought of as the "cost" of having one extra supply unit at node j. In fact, as we describe in Online Appendix I, MBP has the attractive property that it executes stochastic mirror descent (Beck and Teboulle 2003) on (15) (where the inverse mirror map is the antiderivative of the congestion function $\mathbf{f}(\cdot)$).

Our main result, Theorem 1, is directly driven by the following proposition that connects the performance of MBP and the benchmark in the nonstationary environment.

Proposition 2. *Consider the setting in Theorem 1. Then there exists $K_1 = \text{poly}(m, (1/\alpha_{\min}))$, and a universal constant $C < \infty$, such that the following holds. For the congestion function $f(\cdot)$ defined in (8), for any $K \geq K_1$, and any $0 < \Delta_T < T$ the following guarantees hold for Algorithm 1:*

$$L_T^{\text{MBP}} \leq M_1 \frac{K}{\Delta_T} + M_2 \frac{1}{K} + \Delta_T m\eta, \quad \text{for } M_1 \triangleq Cm \text{ and } M_2 \triangleq Cm^2.$$

We illustrate the high-level structure of the proof in Figure 3. We introduce a "batch size" quantity Δ_T; Δ_T is only used for the purpose of analysis and is *not* part of our algorithms. The loss of MBP policy in Δ_T periods can be decomposed into three terms:

(1) The first term (from left to right in Figure 3), is $\frac{1}{T}\sum_{t=0}^{T-1} W^{\text{SPP}(\boldsymbol{\phi}^t)} - W_T^{\text{MBP}}$. We call it the *policy gap*. We bound the policy gap using a Lyapunov analysis in Proposition 3. The antiderivative of the congestion function $\mathbf{f}(\cdot)$ serves as the Lyapunov function. Our choice of $\mathbf{f}(\cdot)$ ensures that the policy gap is (provably) small despite the no-underflow constraints.

(2) The second term, $W^{\text{SPP}(\bar{\boldsymbol{\phi}})} - (1/\Delta_T)\sum_{t=0}^{\Delta_T-1} W^{\text{SPP}(\boldsymbol{\phi}^t)}$ arises from the temporal variation of demand; therefore, we call it the *variation gap* (it is zero for stationary demand). We bound the variation gap in Proposition 4 using graph-theoretic analysis of the sensitivity of the SPP to $\boldsymbol{\phi}$.

(3) The third term, $W_T^* - W^{\text{SPP}(\bar{\boldsymbol{\phi}})}$, has already been bounded in Proposition 1.

Intuitively, the quantities in Figure 3 highlight the following tradeoff. When Δ_T is large, the policy gap is small; when Δ_T is small, the variation gap is small. Theorem 1 follows by balancing this tradeoff and setting $\Delta_T = \Theta(\min(\sqrt{K/\eta}, T))$, which we prove in Online Appendix D.

4.1. Bounding the Policy Gap

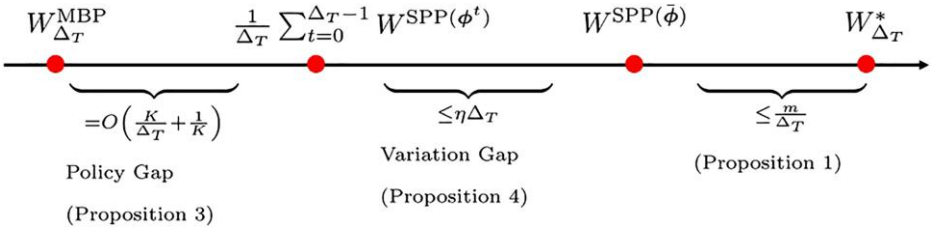
In this section, we prove the following proposition that bounds the policy gap under MBP.

Proposition 3. *Consider a set of m nodes and any sequence of demand arrival rates $\{\boldsymbol{\phi}^t\}_{t\leq T}$ that satisfy Condition 1. Recall that $\alpha_{\min} = \min_{0\leq t\leq T}\alpha(\boldsymbol{\phi}^t) > 0$. Then there exists $K_1 = \text{poly}(m, (1/\alpha_{\min}))$, and a universal constant $C < \infty$, such that the following holds. For the congestion function $f(\cdot)$ defined in (8), for any $K \geq K_1$, the following guarantees hold for Algorithm 1:*

$$\frac{1}{T}\sum_{t=0}^{T-1} W^{\text{SPP}(\boldsymbol{\phi}^t)} - W_T^{\text{MBP}} \leq M_1 \frac{K}{T} + M_2 \frac{1}{K}, \quad \text{for } M_1 \triangleq Cm \text{ and } M_2 \triangleq Cm^2.$$

Notably, this proposition holds for *arbitrary* sequences of demand arrival rates satisfying Condition 1. The proof is based on Lyapunov analysis. A Lyapunov function will allow us to decompose the expected payoff from the next arrival into a combination of the objective, change in potential, and a per-period loss. We use the antiderivative of $\mathbf{f}(\cdot)$ as our Lyapunov function; for

Figure 3. (Color online) Roadmap of the Proof of Proposition 2



the congestion function f in (8), this is

$$F(\overline{\mathbf{q}}) \triangleq -2\sqrt{m}\sum_{j\in V}\sqrt{\overline{q}_j}. \tag{16}$$

The motivation for our choice of Lyapunov function comes from the "drift-plus-penalty" framework in the network control literature (Neely 2006, 2010; Gupta and Radovanović 2020). We generalize and refine the framework by adding a new term in the analysis, which allows us to bound the suboptimality contributed by underflow.

Recall that $W^{\mathrm{SPP}(\boldsymbol{\phi}^t)}$ is the optimal value of SPP (11–13) with demand arrival rate $\boldsymbol{\phi}^t$, $v^{\mathrm{MBP}}[t]$ denotes the payoff collected under the MBP policy in the tth period, and $g^t(\cdot)$ is the dual problem (15). We have the following key lemma (proved in Online Appendix B.1).

Lemma 1 (Suboptimality of MBP in One Period). *Consider any congestion function $f(\cdot)$ that is strictly increasing and continuously differentiable. We have the following decomposition (recall that $\tilde{K} = K + m\sqrt{K}$):*

$$\begin{aligned} &W^{\mathrm{SPP}(\boldsymbol{\phi}^t)} - \mathbb{E}[v^{\mathrm{MBP}}[t]\,|\,\overline{\mathbf{q}}[t]] \\ &\le \underbrace{\tilde{K}\left(F(\overline{\mathbf{q}}[t]) - \mathbb{E}[F(\overline{\mathbf{q}}[t+1])\,|\,\overline{\mathbf{q}}[t]]\right)}_{\mathcal{V}_1 \text{ change in potential}} \\ &\quad + \underbrace{\frac{1}{2\tilde{K}}\max_{j\in V}\max_{\overline{q}\in\left[\overline{q}_j[t]-\frac{1}{\tilde{K}},\overline{q}_j[t]+\frac{1}{\tilde{K}}\right]}|f'(\overline{q})|}_{\mathcal{V}_2 \text{ loss due to stochasticity}} \\ &\quad + \underbrace{\left(W^{\mathrm{SPP}(\boldsymbol{\phi}^t)} - g^t(\mathbf{f}(\overline{\mathbf{q}}[t]))\right)}_{\mathcal{V}_3 \text{ dual optimality gap}} + \underbrace{\mathbb{1}\{q_j[t]=0, \exists j\in V\}}_{\mathcal{V}_4 \text{ loss due to underflow}}. \end{aligned} \tag{17}$$

In Lemma 1, the left-hand side (LHS) of (17) is the suboptimality incurred by MBP (benchmark against the value of SPP($\boldsymbol{\phi}^t$)) in a single period. On the right-hand side (RHS) of (17), $\mathcal{V}_1$ and $\mathcal{V}_2$ come from the standard Lyapunov drift argument; $\mathcal{V}_3$ is the negative of the dual suboptimality at $\mathbf{y} = \mathbf{f}(\overline{\mathbf{q}}[t])$, and hence it is always nonpositive; amd $\mathcal{V}_4$ is the payoff loss because of underflow.

For the change in potential $\mathcal{V}_1$, it forms a telescoping series when summing over many periods and will therefore remain bounded. Hence, as we average over many periods, we have that $\mathcal{V}_1$ tends to zero.

We now proceed to upper bound $\mathcal{V}_2 + \mathcal{V}_3 + \mathcal{V}_4$ on the right-hand side of (17). We outline our analysis in the following. Observe that the terms $\mathcal{V}_2$ and $\mathcal{V}_4$ are nonnegative, whereas $\mathcal{V}_3$ is nonpositive; thus, the goal is to show that $\mathcal{V}_3$ compensates for $\mathcal{V}_2 + \mathcal{V}_4$. First, $\mathcal{V}_2$ is large when there exist very short queues (because the congestion function (8) changes rapidly only for short queue lengths), and $\mathcal{V}_4$ is nonzero 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 (15). In Lemma 2, we provide an upper bound for $\mathcal{V}_3$ that becomes more negative as the shortest queue length decreases. We prove Lemma 2 in Online Appendix B.2 by using complementary slackness for the SPP (11–13).

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

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

The following lemma bounds $\mathcal{V}_2 + \mathcal{V}_3 + \mathcal{V}_4$. The proof is in Online Appendix B.3. (In fact, we prove a general version of the lemma that applies to all congestion functions that satisfy certain growth conditions formalized in Condition 3 in Online 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 (8) and any $\boldsymbol{\phi}$ with connectedness $\alpha(\boldsymbol{\phi}) > 0$. Then there exists $K_1 = \mathrm{poly}(m, (1/\alpha(\boldsymbol{\phi})))$ such that for $K \ge K_1$ and a universal constant $C > 0$,*

$$\mathcal{V}_2 + \mathcal{V}_3 + \mathcal{V}_4 \le M_2\frac{1}{\tilde{K}} \quad \text{for } M_2 \triangleq Cm^2. \tag{18}$$

Recall that $\tilde{K} = K + m\sqrt{K}$.

Putting Lemmas 1 and 3 together leads to the following proof of Proposition 3. The main idea is to use the Lyapunov drift argument of Neely (2010), namely, to sum the expectation of (17) (the bound in Lemma 1) over the first T time steps.

Proof of Proposition 3. Plugging in Lemma 3 into (17) in Lemma 1 and taking expectation, we obtain that there exists $K_1 = \mathrm{poly}(m, (1/\alpha_{\min}))$ such that for any $0 \le t \le T$,

$$\begin{aligned} W^{\mathrm{SPP}(\boldsymbol{\phi}^t)} - \mathbb{E}[v^{\mathrm{MBP}}[t]] &\le \tilde{K}\left(\mathbb{E}[F(\overline{\mathbf{q}}[t])] - \mathbb{E}[F(\overline{\mathbf{q}}[t+1])]\right) \\ &\quad + M_2\frac{1}{\tilde{K}} \quad \text{for } K \ge K_1, \end{aligned} \tag{19}$$

where $\tilde{K} = K + m\sqrt{K}$. Take the sum of both sides of Inequality (19) from $t = 0$ to $t = T-1$ and divide the sum by T. This yields

$$\begin{aligned} &\frac{1}{T}\sum_{t=0}^{T-1} W^{\mathrm{SPP}(\boldsymbol{\phi}^t)} - W_T^{\mathrm{MBP}} \\ &\le \frac{\tilde{K}}{T}\left(\mathbb{E}[F(\overline{\mathbf{q}}[0])] - \mathbb{E}[F(\overline{\mathbf{q}}[T])]\right) + M_2\frac{1}{\tilde{K}} \\ &\le \frac{\tilde{K}}{T}\sup_{\overline{\mathbf{q}}_1,\overline{\mathbf{q}}_2\in\Omega}\left(F(\overline{\mathbf{q}}_1) - F(\overline{\mathbf{q}}_2)\right) + M_2\frac{1}{\tilde{K}} \quad \text{for } K \ge K_1. \end{aligned}$$

Let $M_1 \triangleq \sup_{\overline{\mathbf{q}}_1,\overline{\mathbf{q}}_2\in\Omega}(F(\overline{\mathbf{q}}_1)-F(\overline{\mathbf{q}}_2))$. Observe that the function $F(\overline{\mathbf{q}})$ given in (16) is negative $F(\overline{\mathbf{q}})\le 0$ for all $\overline{\mathbf{q}}\in\Omega$ and is a convex function that achieves its minimum at $\overline{\mathbf{q}}=\frac{1}{m}\mathbf{1}$. Hence,

$$M_1 \le 0 - \inf_{\overline{\mathbf{q}}\in\Omega} F(\overline{\mathbf{q}}) = -F\left(\frac{1}{m}\mathbf{1}\right) = 2m.$$

Therefore, the policy gap of MBP is upper bounded by $M_1(\tilde{K}/T)+M_2(1/\tilde{K})$, where $M_1=Cm$, $M_2=Cm^2$, and C does not depend on m, K, or $\alpha_{\min}$. Moreover, $\tilde{K}=K+m\sqrt{K}\in[K,2K]$ taking $K_1\ge m^2$. This concludes the proof. □

We conclude with some informal intuition as to why MBP with congestion function $\mathbf{f}(\cdot)$ given in (8) and normalized queue lengths $\overline{\mathbf{q}}$ defined in (9) ensures a small policy gap. MBP could run into two issues due to the no-underflow constraints: (i) The queue lengths corresponding to the optimal dual variables lie outside of the state space; and (ii) the Lyapunov drift could be positive at certain "boundary states," that is, states where some of the queues are empty. For issue (i), although the range of normalized queue length $\overline{\mathbf{q}}$ belongs to $[0,1]$, the range of $f(\overline{\mathbf{q}})$ goes to $(-\infty,-\sqrt{m})$ as $K\to\infty$. As a result, for large enough K, there exists $\overline{\mathbf{q}}\in\Omega$ such that $\mathbf{f}(\overline{\mathbf{q}})$ corresponds to the optimal dual variables.[6] For issue (ii), first, this problem only occurs when there exists empty queues. At these states, the dual suboptimality at $\mathbf{f}(\overline{\mathbf{q}})$ is large because $f(0)\to-\infty$ as $K\to\infty$, which creates a negative Lyapunov drift that "pushes" $\mathbf{f}(\overline{\mathbf{q}})$ toward the optimal dual variable. This corresponds to the intuition that MBP is aggressive in preserving supply units in near-empty queues. In contrast, we show in Online Appendix H.2 that regular backpressure (i.e., linear $\mathbf{f}(\cdot)$) may fail to address the two issues mentioned previously, leading to a large policy gap.

4.2. Bounding the Variation Gap

We have the following result that bounds the variation gap.

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

$$\frac{1}{T}\sum_{t=1}^{T} W^{\mathrm{SPP}(\boldsymbol{\phi}_t)} \ge W^{\mathrm{SPP}(\overline{\boldsymbol{\phi}})} - m\eta T. \tag{20}$$

The proof is based on sensitivity analysis of the linear program $\mathrm{SPP}(\overline{\boldsymbol{\phi}})$ via flow decomposition (Williamson 2019). In essence, our approach involves decomposing the feasible flow for $\mathrm{SPP}(\overline{\boldsymbol{\phi}})$ into directed cycles and subsequently reducing it to satisfy constraints in $\mathrm{SPP}(\boldsymbol{\phi}^t)$. We prove Proposition 4 in Online Appendix C.

4.3. Proof of Proposition 2

Using Proposition 3 and considering the first Δ_T periods, we obtain that there exists $K_2=\mathrm{poly}(m,(1/\alpha_{\min}))$, and a universal constant $C<\infty$, such that the following holds. For the congestion function $f(\cdot)$ defined in (8), for any $K\ge K_2$, the following guarantees hold for Algorithm 1:

$$\frac{1}{\Delta_T}\sum_{t=0}^{\Delta_T-1} W^{\mathrm{SPP}(\boldsymbol{\phi}^t)} - W^{\mathrm{MBP}}_{\Delta_T} \le M_1\frac{K}{\Delta_T}+M_2\frac{1}{K},$$
$$\text{for } M_1\triangleq Cm \quad \text{and } M_2\triangleq Cm^2. \tag{21}$$

Suppose the demand varies η_1-slowly on $[0,\Delta_T]$. Using Proposition 1 and then Proposition 4, we have

$$\begin{aligned} L^{\mathrm{MBP}}_{\Delta_T} &= W^{*}_{\Delta_T} - W^{\mathrm{MBP}}_{\Delta_T} \\ &\le \left(W^{\mathrm{SPP}(\overline{\boldsymbol{\phi}})}+m\frac{K}{\Delta_T}\right) - W^{\mathrm{MBP}}_{\Delta_T} \\ &\le \left(m\eta_1\Delta_T+\frac{1}{\Delta_T}\sum_{t=0}^{\Delta_T-1}W^{\mathrm{SPP}(\boldsymbol{\phi}^t)}\right)+m\frac{K}{\Delta_T}-W^{\mathrm{MBP}}_{\Delta_T} \\ &\le \left(Cm\frac{K}{\Delta_T}+M_2\frac{1}{K}\right)+m\eta_1\Delta_T+m\frac{K}{\Delta_T} \\ &\le \frac{K}{\Delta_T}m(C+1)+M_2\frac{1}{K}+m\eta_1\Delta_T. \end{aligned} \tag{22}$$

Here, we used (21) in the third inequality. Suppose the demand varies η_ℓ-slowly in batch ℓ, and because the demand varies η-slowly over the whole horizon $[0,T]$, we must have $\eta=(1/\lfloor T/\Delta_T\rfloor)\sum_{\ell=1}^{\lfloor T/\Delta_T\rfloor}\eta_\ell$. Now take average on both sides of (22) over $\lfloor T/\Delta_T\rfloor$ batches, we obtain the result.

5. Generalizations and Extensions

In this section, we consider two general settings: one with finite buffer queues and one allows JPA. We show that the extended models enjoy similar performance guarantees to that in Theorem 1 under mild conditions on the model primitives.

5.1. 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 $d_j=\overline{d}_jK$ for some scaled buffer size $\overline{d}_j\in(0,1)$. (If $\overline{d}_j\ge 1$, the buffer size exceeds the number of supply units $d_j\ge K$ and there is no constraint as a result, i.e., $j\notin V_{\mathrm{b}}$.) We will find it convenient to define $\overline{d}_j=1$ 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 $\sum_{j\in V}\overline{d}_j>1$. The normalized state space

will be

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

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

$$\overline{q}_j \triangleq \frac{q_j + \overline{d}_j \delta_K}{\tilde{K}} \quad \text{for} \quad \delta_K = \sqrt{K} \quad \text{and}$$
$$\tilde{K} \triangleq K + \left(\sum_{j \in V} \overline{d}_j\right) \delta_K. \tag{23}$$

One can verify that $\overline{\mathbf{q}} \in \Omega$ for any feasible state $\mathbf{q}$. When $\overline{d}_j = 1$ for all $j \in V$, the definition of $\overline{q}_j$ in (23) reduces to the one in (9). 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:

$$f_j(\overline{q}_j) \triangleq \begin{cases} \sqrt{m} \cdot C_b \cdot \left(\left(1 - \frac{\overline{q}_j}{\overline{d}_j}\right)^{-\frac{1}{2}} - \left(\frac{\overline{q}_j}{\overline{d}_j}\right)^{-\frac{1}{2}} - D_b \right), & \forall j \in V_{\mathrm{b}}, \\ -\sqrt{m} \cdot \overline{q}_j^{-\frac{1}{2}} & \forall j \in V \setminus V_{\mathrm{b}}. \end{cases} \tag{24}$$

Here C_b and D_b are normalizing constants[7] chosen to ensure that (i) for all $j,k \in V$, we have that $f_j(\overline{q}_j) = f_k(\overline{q}_k)$ when both queues are empty $q_j = q_k = 0$; and (ii) for all $j,k \in V_{\mathrm{b}}$, we have that $f_j(\overline{q}_j) = f_k(\overline{q}_k)$ when both queues are full $q_j = d_j$, $q_k = d_k$. (We state the results for other choices of congestion functions in Online Appendix C.)

Note that $f_j(\cdot)$ in (24) is identical to $f(\cdot)$ in (8) for $j \notin V_b$, that is, (24) is a generalization of (8) to the case where some queues have buffer constraints. The intuitive reason (24) is a suitable congestion function is that it enables MBP to focus on queues that 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.

We have the following result for the finite-buffer setting. The proof is in Online Appendix D.

Theorem 2. *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 $\overline{d}_j \in (0,1)$ $\forall j \in V_{\mathrm{b}}$ satisfying[8] $\sum_{j \in V} \overline{d}_j > 1$. Consider any sequence of demand arrival rates $(\boldsymbol{\phi}^t)_{t \le T}$ that satisfy Condition 1 and vary η-slowly (Definition 1). Recall that $\alpha_{\min} = \min_{0 \le t \le T} \alpha(\boldsymbol{\phi}^t)$. Then there exists $K_1 = \mathrm{poly}\left(m, \overline{\mathbf{d}}, \frac{1}{\alpha_{\min}}\right)$, and a universal constant $C < \infty$, such that the following holds. For the congestion function $\mathbf{f}(\cdot)$ defined in (24), for any $K \ge K_1$, the following guarantees hold for Algorithm 1:*

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

for $M_1 = Cm$,

$$M_2 = C \frac{1}{\min_{j \in V} \overline{d}_j} \left(\frac{\sum_{j \in V} \overline{d}_j}{\min\{\sum_{j \in V} \overline{d}_j - 1, 1\}}\right)^{3/2} \sqrt{m}.$$

5.2. JPA Setting

In this section, we consider the JPA setting and design the corresponding MBP policy. The platform's control problem is to set a price for each demand origin-destination pair and decide an assignment at each period to maximize payoff. Our model here will be similar to that of Banerjee et al. (2021), except that the platform does *not* know demand arrival rates, and we allow a finite horizon. 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 2. For simplicity, we assume that the demand type distribution $\boldsymbol{\phi} = (\phi_\tau)_{\tau \in \mathcal{T}}$ is time invariant and that all buffers have infinite capacity in this section.

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

$$(x_{j\tau k}[t])_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} \in \{0,1\}^{|\mathcal{P}(\tau)| \cdot |\mathcal{D}(\tau)|} \quad \text{such that} \quad \sum_{j \in \mathcal{P}(\tau), k \in \mathcal{D}(\tau)} x_{j\tau k}[t] \le 1. \tag{25}$$

As before, we require $x_{j\tau k}[t] = 0$ if $q_j[t] = 0$.

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), then a supply unit relocates based on $x_{j\tau k}[t]$. 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. 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 (Gallego and Van Ryzin 1994).

Condition 2. *Assumptions on demand functions.*

(1) *Assume*[9] $F_\tau(p_\tau^{\min}) = 0$ *and that* $F_\tau(p_\tau^{\max}) = 1$.

(2) *Each demand type's willingness-to-pay is nonatomic 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.)*

(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 *that is realized (i.e., willing to pay the price offered).*

As a consequence of Condition 2, 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 that 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 that 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 is (see Online Appendix E for the statement of SPP and the derivation of its dual)

$$\text{minimize}_{\mathbf{y}}\ g_{\text{JPA}}(\mathbf{y}) \quad \text{for } g_{\text{JPA}}(\mathbf{y}) \triangleq \sum_{\tau\in\mathcal{T}} \phi_\tau \max_{\{0\le\mu_\tau\le 1\}} \left(r_\tau(\mu_\tau) + \mu_\tau \max_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} (-c_{j\tau k} + y_j - y_k) \right). \tag{26}$$

Once again, the MBP policy (Algorithm 2) 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 (26) of $g_{\text{JPA}}(\cdot)$, and the assignment decision of MBP achieves the inner argmax in the definition (26) 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 optimization problem for computing $\mu_\tau[t]$ is a one-dimensional concave maximization problem (Condition 2, part 3); hence, $\mu_\tau[t]$ can be efficiently computed.

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 $F_\tau(\cdot)$ are exactly known to the platform; it may be possible to relax this assumption via a modified policy that "learns" the $F_\tau(\cdot)$; however, pursuing this direction is beyond the scope of the present paper.

Algorithm 2 (MBP Policy for JPA)

At the start of period t, the system observes $\tau[t] = \tau$
$(j^*, k^*) \leftarrow \arg\max_{j\in\mathcal{P}(\tau),k\in\mathcal{D}(\tau)} \{-c_{j\tau k} + f_j(\overline{q}_j[t]) - f_k(\overline{q}_k[t])\}$;
if $q_{j^*}[t] > 0$ **then**
  $\mu_\tau[t] \leftarrow \arg\max_{\mu_\tau\in[0,1]} \{r_\tau(\mu_\tau) + \mu_\tau \cdot (-c_{j^*\tau k^*} + f_{j^*}(\overline{q}_{j^*}[t]) - f_{k^*}(\overline{q}_{k^*}[t]))\}$;
  $p_\tau[t] \leftarrow F_\tau^{-1}(\mu_\tau[t])$;
  $x_{j^*\tau k^*}[t] \leftarrow 1$, that is, if the incoming demand stays, serve it by pick up from j^* and drop off at k^*, otherwise do nothing;
else
  $x_{j^*\tau k^*}[t] \leftarrow 0$, that is, drop the incoming demand;
end
The queue lengths update as

$$\overline{\mathbf{q}}[t+1] = \overline{\mathbf{q}}[t] - \frac{1}{K} x_{j^*\tau k^*}[t](\mathbf{e}_{j^*} - \mathbf{e}_{k^*}).$$

We have the following performance guarantee for Algorithm 2, analogous to Theorem 1.

Theorem 3. *Fix a set* V *of* $m = |V| > 1$ *nodes, 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 1 (strong connectivity), and willingness-to-pay distributions* $(F_\tau)_{\tau\in\mathcal{T}}$ *that satisfy Condition 2. Then there exist* $K_1 < \infty$, $M_1 = Cm$, *and* $M_2 = Cm^2$ *for some universal constant* $C > 0$ *such that for the congestion function* $f(\cdot)$ *defined in (8), the following guarantee holds for Algorithm 2. For any horizon* T *and for any* $K \ge K_1$, *we have*

$$L_T^{\text{MBP}} \le M_1 \frac{K}{T} + M_2 \frac{1}{K}, \quad \text{and} \quad L^{\text{MBP}} \le M_2 \frac{1}{K}.$$

We outline the proof of Theorem 3 in Online Appendix E.

6. 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, whereas supply units and demand units correspond to vehicles and customers, respectively.

6.1. Dynamic Incentive Program for Bike Sharing Systems

Chung et al. (2018) explain that Citi Bike's Bike Angel incentive program works as follows: there are two types of bike stations at any time: the incentivized ones and neutral ones; depending on the origin and the destination stations of a trip, different amounts of points are awarded to the rider. The points have monetary values. The system objective is to minimize out-of-stock and out-of-bike events. Therefore, to view it as an application of our JPA model, we can view the amount of points awarded for a certain trip as (the negative of) price of this trip; the customers have a demand function denoting their response to reward points (i.e., negative of prices); and the value the platform derives from a

ride equals customer utility and/or revenue generated (which is a constant) minus the cash value of points awarded. By using a JPA-based MBP policy, the platform can dynamically set the number of reward points for each origin-destination pair. 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 5.1. 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%).[10] As a result, we expect our control insights to retain their power despite the presence of delays. (Indeed, we will numerically demonstrate in Section 6.1 that this is the case in the ride-hailing 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.

6.2. Online Control of Ride-Hailing Platforms

Ride-hailing platforms make dynamic decisions to optimize their objectives (e.g., revenue, welfare). 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 JEA model[11] studied in Section 2 and JPA model studied in Section 5. Again, a concern may be that travel delays play a significant role in ride-hailing, whereas delays are ignored in our theory. In the following section, we summarize a numerical investigation of ride-hailing focusing on entry and assignment controls only (a full description is provided in Online Appendix F). We find that MBP performs well despite the presence of travel delays. To address the case where the available supply is scarce, we heuristically adapt MBP to incorporate the Little's law constraint (Section 6.3.1).

6.3. Numerical Investigation of the Application to Ride-Hailing

In Section 6.3.1, we examine the performance of MBP policy when there are travel delays using numerical experiments. The simulation environment we study is inspired by ride-hailing and leverages demand estimates deduced from NYC yellow cab data (Buchholz 2022) and travel times from Google Maps. In Section 6.3.2, we provide the summary of simulations that study the performance of MBP policy in large networks. In the interest of space, we provide only the key findings of our simulations here and defer a full description of the simulation environment and various technical details to Online Appendix F.

6.3.1. Travel Delays and the Supply-Aware MBP Policy.

In the following, we investigate the performance of the MBP policy when there are travel delays. Similar to our main setting[12] in Section 2, we allow the platform two control levers: entry control and assignment/dispatch control. 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 retain our simplifying assumption that drivers do not relocate in the absence of a passenger. We consider the following two cases:

(1) *Excess supply.* The number of cars in the system is slightly (5%) above the "fluid requirement" (see Online Appendix F.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"; that is, there are not enough cars to realize the optimal solution of static planning problem (11)–(13) under instantaneous relocation (even if we ignore stochasticity).

We compare our MBP policy to three state-of-the-art policies in literature: the fluid-based policy in Banerjee et al. (2021), the utility-delay optimal algorithm (UDOA) in Neely (2006), and the deficit MaxWeight (DMW) policy in Jiang and Walrand (2009). The UDOA policy is in fact a member of the MBP family of policies, with exponential congestion function $f(q) = \omega \cdot (e^{\omega(q-q_0)} - e^{\omega(q_0-q)})$ for suitable $\omega, q_0 > 0$. See Online Appendix F for a detailed description of these benchmark policies.

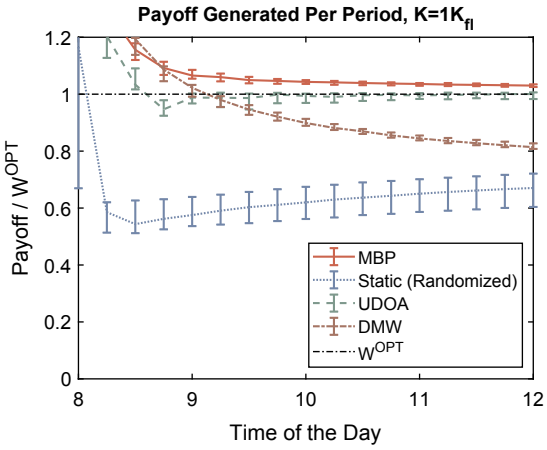
6.3.1.1. Summary of Findings.

We make a natural modification of the MBP policy (with Congestion Function (8)) to account for finite travel times; specifically, we use a *supply-aware MBP* policy that estimates and uses a shadow price of keeping a vehicle (supply unit) occupied for one unit of time.[13] This policy is described at the end of this section.

6.3.1.2. Excess Supply Case.

We simulate the (stationary) system from 8 a.m. to 12 p.m. with 100 randomly generated initial states.[14] The simulation results on performance are shown in Figure 4. The results show that MBP policy significantly outperforms both the DMW policy and the fluid-based policy and consistently outperforms the UDOA policy: The average payoff under MBP over four hours is about 105% of W^{SPP} (here W^{SPP} is again an upper bound on the steady-state performance[15]), whereas UDOA, DMW, and the fluid-based policy achieve 100%, 81%, and 68% of W^{SPP}, respectively. The performance of the static policy converges very slowly to W^{SPP}, leading to

Figure 4. (Color online) Per Period Payoff Under the MBP, UDOA, DMW, and Fluid-Based Policy, Relative to W^{SPP}

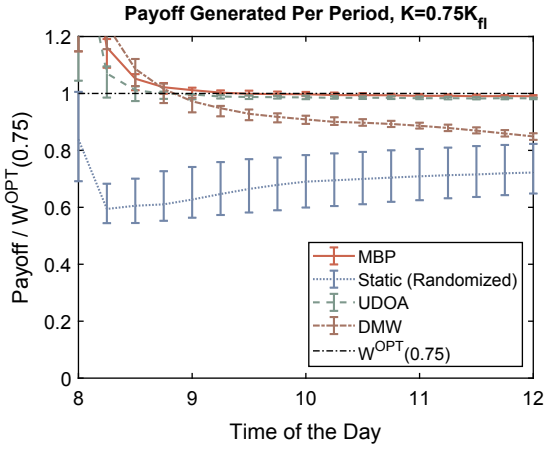


Note. For each data point, we run 100 experiments; the error bars represent the 90% confidence intervals.

poor transient performance.[16] The performance of the DMW policy deteriorates over time because the "fake packets" it generates accumulate in the system.

6.3.1.3. Scarce Supply Case. In the scarce supply case, for example, $K = 0.75K_{\mathrm{fl}}$, no policy can achieve a stationary performance of W^{SPP}; rather, we have a steady-state upper bound of $W^{\mathrm{SPP}}(0.75) \approx 0.86W^{\mathrm{SPP}}$ for this K, where $W^{\mathrm{SPP}}(0.75)$ is the value of the problem given by (11)–(13) together with the supply constraint (27). Figure 5 shows that the MBP policy again vastly outperforms the DMW policy and the fluid-based policy and has similar performance to the UDOA policy in

Figure 5. (Color online) Per Period Payoff Under the (Modified) MBP, (Modified) UDOA, (Modified) DMW, and Fluid-Based Policy, Relative to $W^{\mathrm{SPP}}(0.75)$, the Value of SPP Along with Constraint (27) for $K = 0.75K_{\mathrm{fl}}$



Note. For each data point, we run 100 experiments; the error bars represent the 90% confidence intervals.

the scarce supply case. MBP generates average per period payoff that is 99% of the steady-state upper bound over four hours, whereas the UDOA, DMW, and fluid-based policy achieve 98%, 85%, and 74%, respectively, of the steady-state upper bound 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 (28) in the supply-aware SPP (11–13 along with 28); 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, whereas the performance of DMW deteriorates.

6.3.1.4. Supply-Aware MBP Policy. 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 by Braverman et al. (2019) and Banerjee et al. (2021) 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 (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

$$\sum_{j,k\in V}\sum_{i\in\mathcal{P}(j)} D_{ijk}\cdot z_{ijk} \le K, \tag{27}$$

which simply encodes that the average number of cars occupied at any time cannot exceed K. 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

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

If $w_{i^*jk} + f(\overline{q}_{i^*}[t]) - f(\overline{q}_k[t]) - v[t]D_{i^*jk} \ge 0$ **and** $q_{i^*}[t] > 0$, dispatch from i^*, **else** Drop,

where we define the tightened supply constraint as

$$\sum_{j,k\in V}\sum_{i\in\mathcal{P}(j)} D_{ijk}\cdot z_{ijk} \le 0.95K, \tag{28}$$

where the coefficient of K is the flexible "utilization" parameter that we have set at 0.95, meaning that we are

aiming to keep 5% vehicles free on average, system-wide.[17] Here $v[t]$ is the current estimate of the shadow price for the "tightened" version of supply Constraint (28). We use the congestion function given in (8), that is, $f_j(\overline{q}_j) = \sqrt{m} \cdot \overline{q}_j^{-1/2}$, in our numerical simulations. An important detail here is that the queue lengths are normalized by the estimated number of *free* cars $K - 0.95K = 0.05K$ instead of K. We update $v[t]$ as

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

An iteration of supply-aware MBP is equivalent to executing a (dual) stochastic mirror descent step on the supply-aware SPP (11)–(13) along with (28).

6.3.2. MBP Policy in Large Networks. Recall that, in Corollary 1, the steady-state optimality gap of MBP is shown to be $O\left(\frac{m^2}{K}\right)$ for Congestion Function (8). Compared with the $O\left(\frac{m}{K}\right)$ bound for the fluid-based policy proved in Banerjee et al. (2021), our bound for MBP has the same dependence on K but worse dependence on m. A natural question is whether the worse dependence on m reflects poorer performance or if it is a proof artifact. We conduct numerical experiments in Online Appendix F.2 to study this question.

6.3.2.1. Summary of Findings. We construct a family of instances that has the same total demand rate, but different network sizes m. We compare the performance of our MBP policy with the fluid-based policy in Banerjee et al. (2021) for different values of fleet size K and network size (i.e., number of locations) m. The results demonstrate that MBP consistently outperforms the fluid-based policy in steady state across different choices of m and K. Also, the steady-state suboptimality of MBP appears to scale as m/K (and not m^2/K, which was the scaling of our formal upper bound on the optimality gap).

7. 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 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 that are stationary or vary slowly over time, losing at most $O((K/T) + (1/K) + \sqrt{\eta K})$ 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 average rate of change in demand arrival rates per customer arrival. We considered a variety of 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 that is given exact demand arrival rates and horizon length T in advance. Consider the joint entry-assignment setting (Section 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 $\tilde{f}_j(\overline{q}_j) = y_j^* + f(\overline{q}_j)$. 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.

2. Another promising direction is to pursue the viewpoint that there is an MBP policy that (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 $F(\overline{\mathbf{q}})$ (for the average payoff maximization dynamic programming problem) with respect to $\overline{q}_j$; see Chapter 7.4 of Bertsekas (1995) for background on dynamic programming. Thus, estimates of the relative value function $F(\overline{\mathbf{q}})$ can guide the choice of congestion function.

Endnotes

[1] For a more detailed discussion on this condition and its connection to the present paper, please refer to Remark 1 in Section 2.

[2] Analyses of i.i.d. unit demand arrivals have been shown to generalize easily to more general arrival processes, for example, Markovian arrivals with bounded demand units per period as in Huang and Neely (2009), although at a significant notational burden. Given the aforementioned precedent, we reason that the cost of carrying the reader through this generalization exceeds the benefit of doing so and assume i.i.d. unit demand arrivals throughout the paper.

[3] In (4), the expectations are taken over the randomness in arrivals and (possibly) control decisions and that the supremum is well defined because the payoffs are bounded from above.

[4] 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 $\mathbf{f}(\overline{\mathbf{q}}) \triangleq [f_1(\overline{q}_1), \ldots, f_m(\overline{q}_m)]^\top$ for any monotone increasing (f_j). Here it suffices to consider a single congestion function $f(\cdot)$, whereas in Section 5.1, we will use queue-specific congestion functions $f_j(\cdot)$.

[5] Here "poly" indicates a polynomial. The constant C is universal in the sense that it does not depend on K, m, or $\alpha_{\min}$.

[6] Optimal dual variables for (15) is nonunique, because if $\mathbf{y}^*$ is optimal, $\mathbf{y}^* + \theta\mathbf{1}$ is also optimal for any $\theta \in \mathbb{R}$. Therefore, we can always find an optimal dual variable that corresponds to $\mathbf{f}(\overline{\mathbf{q}})$ where $\overline{\mathbf{q}} \in \Omega$.

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

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

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

[10] The report at https://nacto.org/bike-share-statistics-2017/ tells us that U.S. dock-based systems produced an average of 1.7 rides/bike/day, whereas 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 minutes for casual users. In other words, for most systems, each bike was used less than one 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 less than 20% even during rush hours).

[11] 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.

[12] 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, that is, $\mathcal{T} = V \times V$. 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 dropoff; therefore, $\mathcal{D}(j,k) = \{k\}$. 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. (2021) for the stationary demand setting.

[13] To make the comparison fair, we modify the UDOA and DMW policies using the same heuristic approach, as the original UDOA and DMW policies do not take into account the travel delays.

[14] 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.

[15] 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 one at early times.

[16] For example, the average payoff generated by static policy in the last hour of a 20-hour period is $0.96W^{\mathrm{SPP}}$.

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

References

Adan I, Weiss G (2012) A loss system with skill-based servers under assign to longest idle server policy. *Probability Engrg. Inform. Sci.* 26(3):307–321.

Agarwal N, Ashlagi I, Azevedo E, Featherstone CR, Karaduman Ö (2019) Market failure in kidney exchange. *Amer. Econom. Rev.* 109(11):4026–4070.

Agrawal S, Devanur NR (2014) Fast algorithms for online stochastic convex programming. Krauthgamer R, ed. *Proc. 26th Annual ACM-SIAM Sympos. on Discrete Algorithms* (SIAM, Philadelphia), 1405–1424.

Balseiro SR, Brown DB, Chen C (2021) Dynamic pricing of relocating resources in large networks. *Management Sci.* 67(7):4075–4094.

Banerjee S, Freund D, Lykouris T (2021) Pricing and optimization in shared vehicle systems: An approximation framework. *Oper. Res.* 70(3):1783–1805.

Banerjee S, Kanoria Y, Qian P (2018) Dynamic assignment control of a closed queueing network under complete resource pooling. Preprint, submitted March 13, https://arxiv.org/abs/1803.04959.

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

Bertsekas DP (1995) *Dynamic Programming and Optimal Control*, vol. 1 (Athena Scientific, Belmont, MA).

Braverman A, Dai JG, Liu X, Ying L (2019) Empty-car routing in ridesharing systems. *Oper. Res.* 67(5):1437–1452.

Bubeck S, Cohen MB, Lee YT, Lee JR, Ma¸dry A (2018) K-server via multiscale entropic regularization. *Proc. 50th Annual ACM SIGACT Sympos. on Theory of Comput.* (ACM, New York), 3–16.

Buchholz N (2022) Spatial equilibrium, search frictions, and dynamic efficiency in the taxi industry. *Rev. Econom. Stud.* 89(2):556–591.

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

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

Caldentey R, Kaplan EH, Weiss G (2009) Fcfs infinite bipartite matching of servers and customers. *Adv. Appl. Probability* 41(3):695–730.

Chung H, Freund D, Shmoys DB (2018) Bike angels: An analysis of Citi bike's incentive program. *Proc. 1st ACM SIGCAS Conf. on Comput. and Sustainable Societies* (ACM, New York), 5.

Dai JG, Lin W (2005) Maximum pressure policies in stochastic processing networks. *Oper. Res.* 53(2):197–218.

Dai JG, Lin W (2008) Asymptotic optimality of maximum pressure policies in stochastic processing networks. *Ann. Appl. Probability* 18(6):2239–2299.

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

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

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

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

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

Gupta A, Molinaro M (2016) How the experts algorithm can help solve lps online. *Math. Oper. Res.* 41(4):1404–1431.

Gupta V, Radovanović A (2020) Interior-point-based online stochastic bin packing. *Oper. Res.* 68(5):1474–1492.

Harrison JM (2003) Brownian models of open processing networks: Canonical representation of workload. *Ann. Appl. Probability* 13(1):390–393.

Huang L, Neely MJ (2009) Delay reduction via Lagrange multipliers in stochastic network optimization. *Proc. 7th Internat. Sympos. on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks* (IEEE, Piscataway, NJ), 1–10.

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

Jiang L, Walrand J (2009) Stable and utility-maximizing scheduling for stochastic processing networks. *Proc. 47th Annual Allerton Conf. on Comm., Control, and Comput.* (IEEE, Piscataway, NJ), 1111–1119.

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

Jordan WC, Graves SC (1995) Principles on the benefits of manufacturing process flexibility. *Management Sci.* 41(4):577–594.

Ma H, Fang F, Parkes DC (2019) Spatio-temporal pricing for ridesharing platforms. *Proc. ACM Conf. on Econom. and Comput.*, 583–583.

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

Neely MJ (2006) Super-fast delay tradeoffs for utility optimal fair scheduling in wireless networks. *IEEE J. Selected Areas Comm.* 24(8):1489–1501.

Neely MJ (2010) Stochastic network optimization with application to communication and queueing systems. *Synthesis Lectures Comm. Networks* 3(1):1–211.

Nemirovsky AS, Yudin DB (1983) Problem complexity and method efficiency in optimization. *Problem Complexity and Method Efficiency in Optimization* (John Wiley & Sons, New York).

Özkan E, Ward AR (2020) Dynamic matching for real-time ride sharing. *Stochastic Systems* 10(1):29–70.

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

Stolyar AL (2004) Maxweight scheduling in a generalized switch: State space collapse and workload minimization in heavy traffic. *Ann. Appl. Probability* 14(1):1–53.

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

Talluri KT, Van Ryzin GJ (2006) *The Theory and Practice of Revenue Management*, vol. 68 (Springer Science & Business Media, Boston).

Tassiulas L, Ephremides A (1992) Stability properties of constrained queueing systems and scheduling policies for maximum throughput in multihop radio networks. *IEEE Trans. Automated Control* 37(12):1936–1948.

Walton N (2015) Concave switching in single-hop and multihop networks. *Queueing Systems* 81(2):265–299.

Williamson DP (2019) *Network Flow Algorithms* (Cambridge University Press, Cambridge, UK).