Ride-Hailing Networks with Strategic Drivers: The Impact of Platform Control Capabilities on Performance

Philipp Afèche[*] Zhe Liu[†] Costis Maglaras[‡§]

February 16, 2018

Abstract

Problem Definition: Motivated by ride-hailing platforms such as Uber, Lyft and Gett, we study the problem of matching demand (riders) with self-interested capacity (drivers) over a spatial network. We focus on the performance impact of two operational platform control capabilities, demand-side admission control and supply-side repositioning control, from the platform's, the drivers' and the riders' viewpoints, considering the interplay with two practically important challenges: (i) Significant spatial demand imbalances prevail for extended periods of time; and (ii) drivers are self-interested and strategically decide whether to join the network, and if so, when and where to reposition when not serving riders.
Academic/Practical Relevance: This matching problem considers the practically essential yet theoretically underexplored interplay of operational features and incentives in ride-hailing platforms: spatial demand imbalances, transportation times, queueing, and driver incentives and repositioning decisions.
Methodology: We develop and analyze the steady-state behavior of a novel game-theoretic fluid model of a two-location, four-route loss network.
Results: First, we fully characterize and compare the steady-state system equilibria under three control regimes, from minimal platform control to centralized admission and repositioning control. Second, we provide a new result on the interplay of platform admission control and drivers' repositioning decisions: It may be optimal to strategically reject demand at the low-demand location even if drivers are in excess supply, to induce repositioning to the high-demand location. We provide necessary and sufficient conditions for this policy. Third, we derive tight upper bounds on the gains in platform revenue and drivers' profits due to platform controls; these are more significant under moderate capacity and significant cross-location demand imbalance.
Managerial Implications: Our results contribute important guidelines on the optimal operations of ride-hailing networks. Our model can also inform the design of driver compensation structures that support more centralized network control.

Keywords: ride-hailing, control, network, matching, strategic drivers, demand imbalance

[*]Rotman School of Management, University of Toronto (afeche@rotman.utoronto.ca)
[†]Graduate School of Business, Columbia University (zliu18@gsb.columbia.edu)
[‡]Graduate School of Business, Columbia University (c.maglaras@gsb.columbia.edu)
[§]We are grateful to Cyndia Yu, Victoria Lin and Aruna Prasad for excellent research assistance.

1 Introduction

We are motivated by the emergence of ride-hailing platforms such as Uber, Lyft and Gett, that face the problem of matching service supply (drivers) with demand (riders) over a spatial network. We study the impact of operational platform controls on the equilibrium performance of such networks, focusing on the interplay with two practically important challenges for this matching problem: (i) Significant demand imbalances prevail across network locations for extended periods of time, as commonly observed in urban areas during rush hour (see Figures 1 and 2), so that the natural supply of drivers dropping off riders at a location either falls short of or exceeds the demand for rides originating at this location. These mismatches adversely affect performance as they lead to lost demand, to drivers idling at low-demand locations, and/or to drivers repositioning, i.e., traveling without serving a rider, from a low- to a high-demand location. (ii) Drivers are self-interested and decide strategically whether to join the network, and if so, when and where to reposition, trading off the transportation time and cost involved against their matching (queueing) delay at their current location. These decentralized supply decisions may not be optimal for the overall network.

Motivated by these challenges, we consider two platform control levers, demand-side *admission control* and supply-side *capacity repositioning*. Admission control allows the platform to accept or reject rider requests based on their destinations, thereby affecting not only the supply of cars through the network, but also the queueing delays of drivers to be matched at lower-demand locations, and in turn, their decisions to reposition to higher-demand locations. Repositioning control allows the platform to direct drivers to go where they are needed the most, rather than having to incentivize them to do so, thereby alleviating the demand/supply imbalance in the network.

To evaluate these control levers we study the steady-state behavior of a deterministic fluid model of a two-location ride-hailing loss network with four routes (two for local and two for cross-location traffic) in a game-theoretic framework. Demand for each route is characterized by an arrival rate and a deterministic travel time. Three parties interact through this network. Riders generate demand for each route, paying a fixed price per unit travel time. Prices are fixed throughout, and for simplicity assumed to be route-independent, though this is not necessary for our analysis. Drivers decide, based on their outside opportunity cost and their equilibrium expected profit rate from participation, whether to join the network, and if so, whether to reposition from one location to the other, rather than idling. Drivers have heterogeneous opportunity costs but homogenous transportation costs; therefore, participating drivers are homogeneous in their strategies and in the eyes of the platform. The platform receives a fixed commission (fraction) of the fare paid by the rider and seeks to maximize its commission revenue. We consider three control regimes: (i) *Centralized Control* with respect to admission and driver repositioning decisions; (ii) *Minimal Control*: no admission control and decentralized (driver) repositioning decisions; (iii) *Optimal Admission Control* at each location and decentralized repositioning.

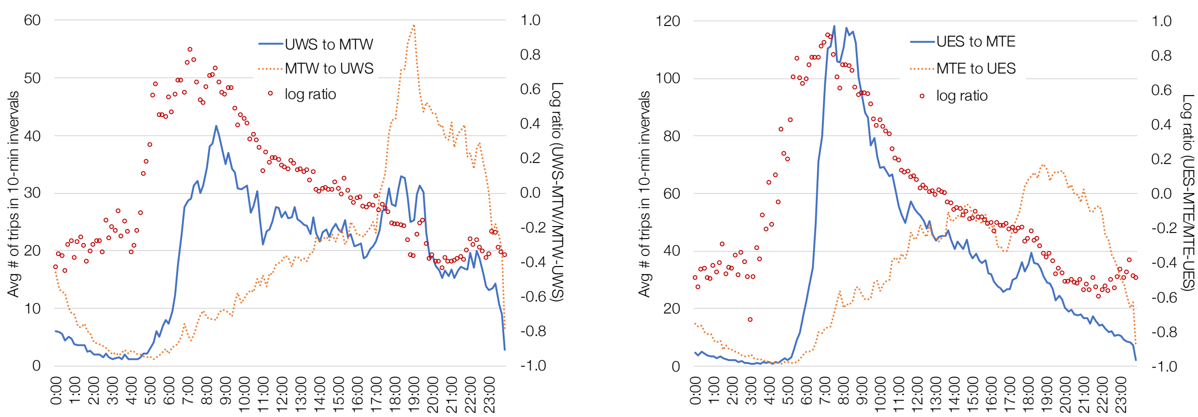
Main Results and Contributions. First, we propose a novel model that accounts for the network structure and flow imbalances, the driver incentives, and the interplay of queueing, transportation times, and driver repositioning decisions—all essential features of ride-hailing platforms.

Second, we fully characterize the steady-state system equilibria, including the drivers' repositioning incentive-compatibility conditions, for the three control problems outlined above, relying on the analysis of equivalent capacity allocation problems. The solutions provide insights on how and why platform control impacts the system performance, both financially, in terms of platform revenue and per-driver profits, and operationally, in terms of rider service levels and driver participation, repositioning and queueing. One immediate finding is that when capacity is moderate, accepting rider requests in a pro-rata (or FIFO) manner need not be optimal, and neither is the practice of accepting rider requests at a location as long as it has available driver supply.

Third, we glean new insights on the interplay between the platform's admission control and drivers' strategic repositioning decisions. Specifically, we show that the platform may find it optimal to *strategically reject demand* at the low-demand location, even though there is an excess supply of drivers, so as to induce repositioning to the high-demand location. We provide intuitive necessary and sufficient conditions for this policy feature in terms of the network characteristics and the driver economics. This deliberate degradation of the rider service level at one location yields more efficient repositioning, and in turn a higher service level at the other location. Thus, operational levers, as opposed to location-specific or origin-destination pricing, can affect repositioning behavior.

Fourth, we derive upper bounds on the performance benefits that the platform and the drivers enjoy due to increased platform control capability. These bounds show that these benefits can be very significant for the platform, of the order of 50%, 100% or even larger improvements, especially when the network operates in a moderate capacity regime, so there are tangible supply/demand trade-offs across the network, and when there are significant cross-location demand imbalances. Per-driver benefits are also most significant when capacity is moderate, but are less significant in their magnitude because driver participation decisions are endogenous—so, if platform controls increase the per-driver profit significantly at a given participation level, more drivers choose to participate, reducing the equilibrium per-driver benefit.

Flow Imbalances: Example Manhattan. We illustrate the magnitude and duration of the demand imbalances alluded to above with data on flow imbalances at the route- and location-level, for taxi rides in Manhattan. These flow imbalances are derived from the publicly available New York City TLC (Taxi & Limousine Commission) Trip Record Data.[1] These data record for each taxi ride the start and end time, the geo pickup and dropoff locations, and the fare paid. Although the data report censored demand (i.e., only realized trips but not rider demand that was not filled), we believe the (uncensored) demand imbalances are likely of the same order of magnitude as the (censored) flow imbalances; indeed, in the high-traffic direction the flow likely underestimates demand by a larger margin than in the opposite direction. We further note that, although our data do not include trips on ride-hailing platforms (this information is not public), these platforms likely experience similar imbalances.



(a) Upper West Side–Midtown West, May 2016

(b) Upper East Side–Midtown East, May 2016

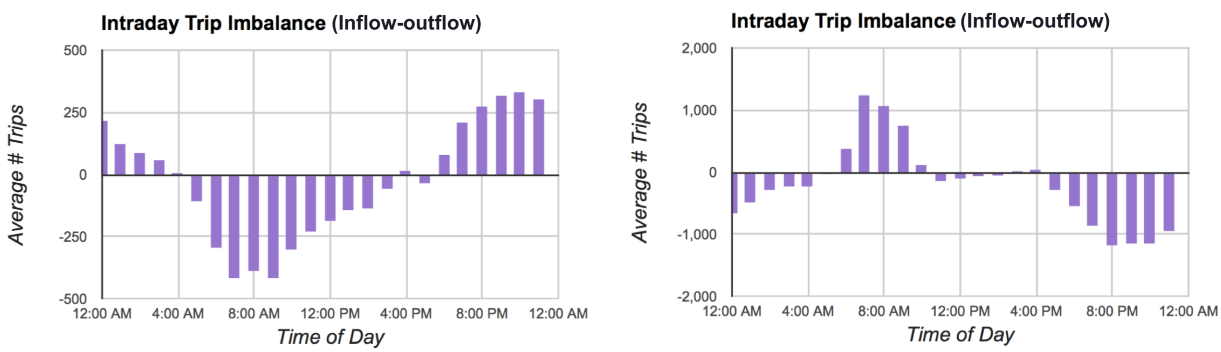
Figure 1: Route-level flow imbalances in Manhattan

Figure 1 illustrates the route-level realized *flow* imbalances for two origin-destination pairs in Manhattan, NY. Chart (a) shows the average number of trip originations per 10 minute-interval between Upper West Side (UWS) and Midtown West (MW) (to and fro), and the logarithm of the ratio of number of trip originations between the two locations per time bin; for each time bin the average is computed over all weekdays for one month. Chart (b) shows the same quantities between the Upper East Side and Midtown East. We observe a pronounced imbalance of almost one order of magnitude (about 10x) in the morning rush hour and about half an order of magnitude (about 3x) in the evening rush hour in the reverse direction. Therefore, focusing on origin-destination pairs in the network, we observe strong flow imbalances, so the cross traffic between such two locations will not replenish sufficient capacity for the originating trips in the high-demand location (UWS in the

[1] http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml

morning) and will likely oversupply the low-demand location (MW in the morning).

Figure 2 shows that the realized flow imbalances persist even after aggregation to the location-level. That is, the flow of other trips being completed in the high-demand location, e.g., morning arrivals into the UWS from other locations, is insufficient to supply the necessary capacity for the UWS to MW trips: Charts (a) and (b) show the net imbalances (total number of drop-offs minus total number of pickups) at two locations in Manhattan, over the course of a day. These net imbalances are statistically significant and exhibit strong and different intraday patterns.



(a) Upper West Side, May 2016

(b) Midtown, May 2016

Figure 2: Location-level flow imbalances in Manhattan

Taken together, these observations on route- and location-level flow imbalances suggest that self-interested strategic drivers actively reposition towards high-demand locations so as to supply the needed capacity that would otherwise not be available at high-demand locations/time-periods. They also suggest that platform controls may be able to use destination information to direct capacity where it is needed (something that does not happen in the taxi market), and to incentivize or completely control a more efficient driver repositioning strategy.

Finally, we note that both Figure 1 and Figure 2 show that imbalance periods persist for multiple time bins, lasting typically for a couple of hours. To contrast, the typical transportation times between these location pairs are of the order of 10-15min, suggesting that network transients may settle down quickly relative to the duration of these imbalances, which, in part, motivates our focus on the steady state fluid model equations as opposed to a study of the transient process itself.

The paper proceeds as follows. This section concludes with a brief literature review. In §2 we outline the system model and formulate the three optimization problems that correspond to the three control regimes mentioned above. §3 studies the centralized control regime, and §4 studies the two control regimes where drivers make self-interested repositioning decisions. Finally, §5 derives upper bounds on the performance benefits to the platform and to the drivers due to increased platform control, and offers some numerical illustrations.

Related Literature. This paper is related to the growing literature on the sharing economy, with a specific focus on the operations of ride-hailing platforms. With a couple of exceptions discussed below, this literature can be broadly grouped into two streams: papers that study i) a single location (or the system as a whole) with driver incentives, and ii) a network of locations without strategic drivers or driver incentives. In contrast to these two streams, this paper studies a network model with driver incentives.

In single-location models with incentives, the strategic driver decision is whether to enter the system; driver repositioning is a network consideration that is not captured in these models. Several papers focus on the impact of *surge or dynamic pricing* and its impact on equilibrium performance over static pricing policies. Banerjee et al. (2016) study a stochastic single-location queueing model with price-sensitive riders (demand) and drivers (capacity) and find (consistent with the literature on revenue maximization in large scale queues) that static pricing is asymptotically (first order) optimal. Cachon et al. (2017) compare five pricing schemes, with increasing flexibility in setting consumer price and provider wage, and find that dynamic (surge) pricing significantly increases the platform's profit and benefits drivers and riders. Castillo et al. (2017) show analytically and empirically that surge pricing avoids the "wild goose chase" effect when capacity is scarce—the phenomenon where drivers have to travel far distances to pick up customers. Chen and Sheldon (2015) use Uber data to show empirically that surge pricing incentivizes drivers to work longer and hence increases systemwide efficiency. Hall et al. (2017) in a different empirical study based on Uber data, find that the driver supply is highly elastic to wage changes; a fare hike only has a short-lived effect on raising the driver hourly earnings rate, and is eventually offset by endogenous entry. As mentioned above, our results confirm this empirical phenomenon.

Taylor (2017) and Bai et al. (2017) both study the price and wage decisions of on-demand service platforms that serve delay-sensitive customers and earnings-sensitive independent providers. Taylor (2017) examines the impact of delay sensitivity and provider independence on the optimal price and wage, and shows that the results are different when uncertainty in customer and provider valuations prevails. Bai et al. (2017) extend the model in Taylor (2017) by also considering the impact of demand and service rate and potential capacity, as well as total welfare (in addition to platform profit) as the system performance measures. Some papers consider strategic, self-interested agents (capacity) without queueing considerations. Gurvich et al. (2016) allow the platform to determine the capacity size (staffing), system-state-contingent compensation and a cap on effective capacity.

Hu and Zhou (2017) study the optimal price and wage decisions of an on-demand matching platform. They show that a commission contract can be optimal or near-optimal under market uncertainty. Benjaafar et al. (2015) study peer-to-peer product sharing, where individuals with varying usage levels make decisions about whether or not to own; the roles of market participants' as provider or consumer are endogenous, unlike in the above papers (and ours).

In the second stream that studies ride-hailing networks without incentives, papers primarily either focus on routing of cars or on matching riders to drivers. Most papers rely on the analysis of an approximating deterministic fluid model. An important paper is Braverman et al. (2017); they prove an asymptotic limit theorem that provides justification for the use of a deterministic fluid network model (such as the one in this paper), and then study the transient optimization problem of empty car repositioning under centralized control. In contrast to this paper, Braverman et al. (2017) do not consider admission control and focus on centralized repositioning decisions. The underlying model is the BCMP product-form network. Iglesias et al. (2017) also consider centralized matching and repositioning decisions in the context of a BCMP network model. Similar to our paper, both papers model a closed loss network. Banerjee et al. (2017) study optimal dynamic pricing of a vehicle sharing network, and show that state-independent (demand) prices derived through a convex relaxation are near-optimal when capacity grows large; the paper provides explicit approximation guarantees for systems with finite size. He et al. (2017) study the problem of designing the geographical service region for urban electric vehicle sharing systems.

Hu and Zhou (2016) consider dynamic matching for a two-sided, heterogeneous-type, discrete-time system with random arrivals and abandonment. They provide conditions under which the optimal matching policy is a priority rule. Ozkan and Ward (2017) study dynamic matching on a network, and through a linear programming relaxation propose an asymptotically optimal matching policy that outperforms the commonly-used heuristic of matching a rider with the nearest available driver. Their model differs from the one in Hu and Zhou (2016) in that it i) assumes probabilistic rather than deterministic matching, and ii) establishes that customers and drivers in the same location may not be matched. Feng et al. (2017) compare customers' average waiting time under two booking systems, on-demand versus street hailing, assuming all trips occur on a circle. Caldentey et al. (2009), Adan and Weiss (2012) and Bušić et al. (2013), among others, study multi-class matching in the context of the infinite bipartite matching model. In the broader dynamic queue matching context, Gurvich and Ward (2014) prove the asymptotic optimality of a discrete review matching policy for a multi-class double-sided matching queue.

In contrast to these two streams, this paper studies a network model with strategic supply, i.e., drivers decide whether to join the system and where to provide service, given the incentives offered by the platform. Bimpikis et al. (2017) study how a ride-sharing platform with strategic drivers should price demand and compensate drivers across a network to optimize its profits, and show that the platform's profits and consumer surplus increase when demands are more balanced across the network; they focus on a discrete-time multi-period model without driver queueing effects, where driver movement between any two locations is costless and takes one time-period. Guda and Subramanian (2017) consider strategic surge pricing and market information sharing in a two-period model with two local markets without cross-location demand (there is no queueing and driver travel takes one period). Buchholz (2017) empirically analyzes, using the NYC taxi data, the dynamic spatial equilibrium in the search and matching process between strategic taxi drivers and passengers. His counterfactual analysis shows that even under optimized pricing, performance can further be improved significantly by more directed matching technology, which supports the value of studying the impact of operational controls.

In contrast to Bimpikis et al. (2017), Guda and Subramanian (2017), and Buchholz (2017), we focus on the impact of operational controls, as opposed to pricing, on system performance, and provides insights and guidelines on the optimal operations of ride-hailing networks. Consistent with the pricing results of Bimpikis et al. (2017), we find that cross-network demand imbalances crucially affect performance, and that the impact of operational platform controls increases in such imbalances.

2 Model and Problem Formulations

We consider a deterministic fluid model of a ride-hailing network in steady state. Three parties interact through this network, riders generate demand for rides, drivers supply capacity for the rides, and the platform is instrumental in matching supply with demand. §2.1 introduces the operational and financial model primitives; §2.2 describes the information that is available to the parties and the controls that determine how supply is matched with demand; §2.3 formulates the optimization problems for the three control regimes that we study in this paper. Finally, §2.4 explains how to reformulate the problems specified in §2.3 in terms of capacity allocation decisions.

Such an approximating fluid model was rigorously justified in Braverman et al. (2017) for a system where drivers were not acting strategically; an adaptation of their arguments could be used in our setting as well. We will focus directly on a set of (motivated) steady state flow equations.

2.1 Model Primitives

We describe the model primitives for the network, the riders, the drivers and the platform.

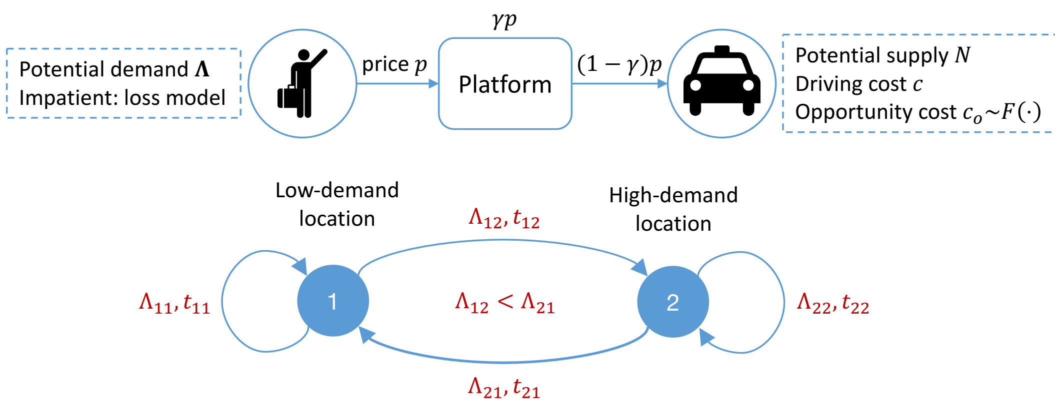


Figure 3: Model primitives

Network. The network has two locations (nodes), indexed by $l = 1, 2$, and four routes (arcs), indexed by lk for $l, k \in \{1, 2\}$. Figure 3 depicts the network schematic and the model primitives that we define in this section. We denote by t_{lk} the travel time on route lk and by t the travel time vector. We impose no restrictions on travel times; specifically, we allow $t_{12} \neq t_{21}$, to reflect, for example, different uptown/downtown routes. The travel times are constant and, in particular, independent of the number of drivers that serve demand for the platform. This assumes that the number of drivers has no significant effect on road congestion and transportation delays.

Riders. Riders generate demand for trips. We assume that the platform charges a fixed price of $\$p$ per unit of travel time, uniform across all routes.; i.e., a rider pays for a route-lk trip a fee of $\$pt_{lk}$. Given the price p, the potential demand rate for route-lk trips is Λ_{lk}, and Λ denotes the potential demand rate vector. The platform keeps a portion γ of the total fee as commission and drivers collect the remainder. Riders are impatient, i.e., rider requests not matched instantly are lost. We focus on the case $\Lambda_{12} \neq \Lambda_{21}$, i.e., the cross-location demands are not balanced; without loss of generality, we make the following assumption.

Assumption 1 (Demand imbalance)**.** $0 < \Lambda_{12} < \Lambda_{21}$.

Drivers. Drivers supply capacity to the network. Let N be the pool of (potential) drivers, each equipped with one car (unit of capacity). Drivers are self-interested and seek to maximize their profit rate per unit time. They decide whether to join the network, and, if so, whether to reposition to the other network location, i.e., travel without serving a rider, in anticipation of a faster match.

Each driver has an idiosyncratic opportunity cost rate, denoted by c_o, that is assumed to be an independent draw from a common distribution F, which is assumed to be continuous with connected support $[0, \infty)$. Drivers join the network if their profit rate per unit time equals or exceeds their outside opportunity cost rate. Therefore, if the per-driver profit rate is \$$x$, then the number of participating drivers, denoted by n, is $n = NF(x) = NP(c_o \leq x)$. The per-driver profit rate x is itself a quantity that emerges in equilibrium and depends on n, the drivers' trip-related earnings and cost, the platform's controls, and the driver decisions, to be specified in §3 and §4.

Participating drivers incur a driving cost rate of \$$c$ per unit of travel time, which is common to all drivers and independent of the car occupancy. Drivers serving rider demand earn revenue at rate \$$\bar{\gamma}p$ per unit travel time, where $\bar{\gamma} = 1 - \gamma$, i.e., the price paid by the rider, net of the platform commission. Therefore, the maximum profit rate that drivers can earn equals $\bar{\gamma}p - c$. However, their actual profit rate is lower if they spend time waiting for riders (accruing zero profit during this wait) and/or repositioning from one location to the other (incurring a cost c per unit time). The following assumption ensures that drivers can earn a positive profit by repositioning[2].

Assumption 2 (Positive profit from repositioning)**.** $ct_{12} < t_{21}(\bar{\gamma}p - c)$ *and* $ct_{21} < t_{12}(\bar{\gamma}p - c)$.

Given Assumption 1, only the first condition in Assumption 2 will prove to be relevant.

Platform. The platform is operated by a monopolist firm that matches drivers with riders with the objective of maximizing its revenue rate.[3] The platform may have two controls: a) demand-side admission control, and b) supply-side capacity repositioning, as detailed in §2.3.

2.2 Matching Supply with Demand

Information. Riders and drivers rely on the platform for matching, that is, potential riders cannot see the available driver capacity, and drivers cannot see the arrivals of rider requests.

The platform knows the model primitives introduced in §2.1, including the potential demand rates Λ, the destination of each trip request, the travel times t, the driving cost c and the opportunity cost rate distribution F. The driver opportunity cost rates are private information, not known by the platform. Therefore, participating drivers are homogenous to the platform. The platform knows the state of the system, i.e., each driver's location, travel direction and status at each point in time.

[2]Assumption 2 also implies that $\bar{\gamma}p - c > 0$, hence $F(\bar{\gamma}p - c) > 0$, so at least some drivers choose to join the network.

[3]The platform could, in practice, incorporate additional considerations in its objective or as control constraints, e.g., to reward market penetration or penalize lost demand.

Riders do not need any network information since they are impatient—they simply arrive with a trip request and leave if their request is not accepted or cannot be fulfilled immediately.

Drivers do not observe the state of the system, but they have (or can infer) the information required to compute their individual expected profit rates, namely: the travel times t; the steady-state delays until they get matched at each location, possibly zero; the destination (routing) probabilities for matches at each location; and the probabilities that they will choose or be instructed to reposition from one location to the other. These delays and the routing and repositioning probabilities are endogenous quantities consistent with the network equilibrium, to be detailed later on.

Admission control. Let $\lambda_{lk} \leq \Lambda_{lk}$ denote the *effective* route-lk demand rate, i.e., the rate of trip requests that are served. Let λ be the vector of effective demand rates and $\Lambda - \lambda$ the vector of lost demand rates. A trip request may be lost either if there is no available driver capacity at the time and location of the request (recall that riders are impatient), or if the platform chooses to reject the request (e.g., based on the requested destination), even though driver capacity is available. In the regimes detailed in §2.3, the platform exercises either no or optimal admission control.

Matching at each location. At each location drivers that become available (i.e., do not reposition upon arrival) join a single queue, to be matched by the platform with riders originating at this location. (Drivers cannot reject the platform's matching requests.) Throughout we assume that the platform matches drivers to accepted ride requests according to a uniform matching policy, such as First-In-First-Out (FIFO) or random service order. Since participating drivers are homogeneous to the platform, this is not restrictive. This implies that in steady state drivers queueing at location l encounter the same waiting time, denoted by w_l, and are directed to route-lk with probability $\frac{\lambda_{lk}}{\lambda_{l1}+\lambda_{l2}}$. Let q_l denote the steady-state number of drivers queueing at location l. By Little's Law we have $w_l = q_l/(\lambda_{l1} + \lambda_{l2})$ for $l = 1, 2$. Let w and q denote, respectively, the vector of steady-state waiting times and queue lengths.

Repositioning of capacity between locations. Let ν_{12} and ν_{21} be the rates of drivers repositioning from location 1 and 2, respectively, and $\nu = (\nu_{12}, \nu_{21})$. Up to three flows emanate from location l. Drivers that are matched with riders leave at rates λ_{l1} and λ_{l2}, and drivers that reposition to location $k \neq l$ leave at rate ν_{lk} (without queueing at location l). Therefore, letting $\eta\,(\lambda, \nu)$ denote the corresponding vector of steady-state repositioning fractions, we have

$$\eta_1\,(\lambda, \nu) = \frac{\nu_{12}}{\lambda_{11} + \lambda_{12} + \nu_{12}} \quad \text{and} \quad \eta_2\,(\lambda, \nu) = \frac{\nu_{21}}{\lambda_{21} + \lambda_{22} + \nu_{21}}. \tag{1}$$

Repositioning decisions are either centralized or decentralized, as detailed in §2.3. Under centralized repositioning the platform controls the repositioning flows ν (e.g., drivers are employees or autonomous vehicles) and the fractions η emerge in response through (1); in this case we assume the drivers are informed about η. Under decentralized repositioning the participating drivers choose the fractions η to maximize their individual profit rates, and the resulting flow rates ν satisfy (1). In both regimes, drivers are homogeneous once they have joined the network, and the resulting repositioning fractions, η, are the same for all participating drivers.

Steady-state system flow balance constraints. Assuming a participating driver capacity equal to n, the effective demand rates λ, repositioning flow rates ν and waiting times w must satisfy: (i) flow balance between locations, $\lambda_{12} + \nu_{12} = \lambda_{21} + \nu_{21}$, and (ii) the capacity constraint $\sum_{l,k=1,2} \lambda_{lk} t_{lk} + (\nu_{12} t_{12} + \nu_{21} t_{21}) + \sum_{l=1,2} w_l (\lambda_{l1} + \lambda_{l2}) = n$, where $\sum_{l,k=1,2} \lambda_{lk} t_{lk}$ is the number of drivers serving riders, $\nu_{12} t_{12} + \nu_{21} t_{21}$ is the number of drivers repositioning between locations, and $\sum_{l=1,2} w_l (\lambda_{l1} + \lambda_{l2})$ is the number of drivers queueing in the two locations.

2.3 Three Control Regimes: Problem Formulations

We study three control regimes referred to as *Centralized Control, Minimal Control* and, *Admission Control*, that differ in terms of (i) whether the platform does or does not exercise admission control, and (ii) whether the platform controls or the drivers control repositioning decisions. Each regime yields an optimization problem in terms of the tuple (λ, ν, w, n), the system flow constraints described in §2.2, and the incentives of the platform and the drivers that we formalize in this section. In all three regimes drivers make their own participation decision by comparing their outside opportunity cost to their per-driver profit rate from joining the system.

Platform revenue. The platform's steady-state revenue rate is given by $\Pi(\lambda) := \gamma p \lambda \cdot t$, where γp is the platform's commission rate per busy driver and $\lambda \cdot t$ is the number of busy drivers.

Driver profit and equilibrium constraints. Drivers are homogeneous once they have joined the network, as explained in §2.2, so all participating drivers achieve the same steady-state profit rate. We can compute the per-driver profit rate in two ways: i) as the per-driver portion of the cumulative driver profits, for the participation equilibrium constraint, and ii) from the perspective of an individual driver circulating through the network, for the repositioning equilibrium constraint.

First, the per-driver profit rate can be computed as follows

$$\pi(\lambda, \nu, n) = \frac{(\bar{\gamma} p - c) \sum_{l,k=1,2} \lambda_{lk} t_{lk} - c(\nu_{12} t_{12} + \nu_{21} t_{21})}{n},$$

where the numerator is the total profit generated by all n participating drivers. A *participation* *equilibrium* requires $n = NF(\pi(\lambda, \nu, n))$.

Second, from the perspective of an infinitesimal driver circulating through the network, her profit rate, denoted by $\widetilde{\pi}(\eta; \lambda, w)$, is a function of her repositioning fractions η, the routing probabilities implied by λ and the delays w in the matching queues. The explicit form of $\widetilde{\pi}(\eta; \lambda, w)$ is discussed in §4.1. The effective demand rates λ and delays w emerge as equilibrium quantities that depend on the platform control and the decisions of all drivers. A participating driver's repositioning strategy is a vector of probabilities that specify for each network location the fraction of times that the driver will, upon arrival, immediately reposition to the other location. Since participating drivers are homogeneous, it suffices to focus on symmetric strategies, where drivers symmetrically choose the fractions η to maximize $\widetilde{\pi}(\eta; \lambda, w)$. In equilibrium, we require that the unique repositioning fractions $\eta(\lambda, \nu)$ induced by the aggregate flow rates (λ, ν) through (1) must agree with individual drivers' profit-maximizing repositioning decisions, i.e.,

$$\eta(\lambda, \nu) \in \arg\max_{\eta} \ \widetilde{\pi}(\eta; \lambda, w). \tag{2}$$

Since every driver chooses $\eta(\lambda, \nu)$, each earns the same profit rate, so that $\widetilde{\pi}(\eta(\lambda, \nu); \lambda, w) = \pi(\lambda, \nu, n)$ for all (λ, ν, w, n) that satisfy the system flow constraints described in §2.2.

Centralized Control (C). In the centralized benchmark the platform has "maximum" control, over both demand admission and driver repositioning decisions. The platform solves:

$$\text{(Problem C)} \quad \max_{\lambda,\nu,w} \quad \mathit{\Pi}(\lambda) \tag{3}$$

$$\text{s.t} \quad \lambda_{12} + \nu_{12} = \lambda_{21} + \nu_{21}, \tag{4}$$

$$\sum_{l,k=1,2} \lambda_{lk} t_{lk} + \nu_{12} t_{12} + \nu_{21} t_{21} + \sum_{l=1,2} w_l (\lambda_{l1} + \lambda_{l2}) = n, \tag{5}$$

$$0 \le \lambda \le \mathit{\Lambda}, \ \nu \ge 0, \ w \ge 0, \tag{6}$$

$$\pi(\lambda, \nu, n) = \frac{(\bar{\gamma} p - c) \sum_{l,k=1,2} \lambda_{lk} t_{lk} - c(\nu_{12} t_{12} + \nu_{21} t_{21})}{n}, \tag{7}$$

$$n = NF(\pi(\lambda, \nu, n)), \tag{8}$$

where (4)–(6) are flow balance conditions and (7) and (8) enforce the participation equilibrium.

Admission Control (A). This regime differs from the centralized benchmark in that repositioning decisions are decentralized, i.e., controlled by drivers. The platform must therefore also account for the repositioning equilibrium constraint (2) and solves:

$$\text{(Problem A)} \quad \max_{\lambda,w} \{ \Pi(\lambda) \ : \ (1)-(2)\,,\ (4)-(8) \} \,. \tag{9}$$

Minimal Control (M). In this regime the platform does not exercise any demand admission control, and drivers control repositioning decisions. The platform simply matches rider trip requests in a pro-rata (or FIFO) manner to drivers, and never turns away requests when there are drivers available to serve them. In addition to (1)–(2) and (4)–(8), the network flows should satisfy the additional conditions (10)–(12) that we discuss next.

First, at each location the effective demand rates are proportional to the corresponding potential demand rates:

$$\frac{\lambda_{l1}}{\Lambda_{l1}} = \frac{\lambda_{l2}}{\Lambda_{l2}}, \quad l = 1, 2, \tag{10}$$

i.e., routes originating at a location receive equal service probabilities (pro rata service).

Second, drivers cannot be repositioning out of location l if the potential rider demand at that location has not been served, i.e.,

$$(\Lambda_{l1} + \Lambda_{l2} - \lambda_{l1} - \lambda_{l2})\nu_{lk} = 0, \quad l = 1, 2, k \neq l, \tag{11}$$

and demand requests originating at a location l can only be lost if this location has no supply buffer, so no drivers are waiting, i.e.,

$$(\Lambda_{l1} + \Lambda_{l2} - \lambda_{l1} - \lambda_{l2})w_l = 0, \quad l = 1, 2. \tag{12}$$

In the Minimal Control Regime, the set of feasible tuples is given by

$$\mathcal{M} = \{ (\lambda, \nu, w, n) \ : \ (1)-(2)\,,\ (4)-(8),\ (10)-(12) \} \,. \tag{13}$$

We will show in §4.2 that for fixed participating capacity n, the set $\mathcal{M}$ is a singleton.

2.4 Reformulation to Capacity Allocation Problems

It is intuitive and analytically convenient to reformulate the above problems in terms of the driver capacities allocated to serving riders, repositioning (without riders), and queueing for riders.

For route lk let S_{lk} denote the offered load of trips, and s_{lk} denote the (effective) capacity serving riders. Let S and s denote the corresponding vectors, $\overline{S} = \sum_{lk} S_{lk}$ the total offered load, and $\overline{s} = \sum_{lk} s_{lk}$ the total service capacity. From Little's Law,

$$S_{lk} = \Lambda_{lk} t_{lk} \quad \text{and} \quad s_{lk} := \lambda_{lk} t_{lk}, \quad l, k \in \{1, 2\}. \tag{14}$$

Let r_{lk} be the capacity repositioning from location l to k, $r = (r_{12}, r_{21})$, and $\overline{r} = r_{12} + r_{21}$, where

$$r_{lk} = \nu_{lk} t_{lk}, \quad l \neq k, \tag{15}$$

and q_l be the capacity queueing at location l. Let $q = (q_1, q_2)$ and $\overline{q} = q_1 + q_2$, where

$$q_l = w_l(\lambda_{l1} + \lambda_{l2}), \quad l = 1, 2. \tag{16}$$

Using (14)–(16) we can transform the formulations of §2.3 into equivalent problems with respect to (s, r, q, n). With some abuse of notation, we write the platform profit function as $\Pi(s) = \gamma p \overline{s}$ instead of $\Pi(\lambda)$, and similarly the per-driver profit functions as $\widetilde{\pi}(\eta; s, q)$ instead of $\widetilde{\pi}(\eta; \lambda, w)$ and $\pi(s, r, n)$ instead of $\pi(\lambda, \nu, n)$, and the repositioning fractions in (1) as $\eta(s, r)$ instead of $\eta(\lambda, \nu)$.

3 Centralized Control (C)

In the centralized control regime (C) the platform controls demand admission and driver repositioning so as to maximize its revenue, and drivers make participation decisions in response to the resulting profit rate. The optimization problem for this regime is (3)–(8), which can be reformulated using (14)–(16) as:

$$\begin{aligned}
&\text{(Problem C)} \quad \max_{s,r,q} \quad && \Pi(s) && (17)\\
&\qquad\qquad\qquad \text{s.t} && \frac{s_{12} + r_{12}}{t_{12}} = \frac{s_{21} + r_{21}}{t_{21}}, && (18)\\
& && \overline{s} + \overline{r} + \overline{q} = n, && (19)\\
& && 0 \leq s \leq S, \; r \geq 0, \; q \geq 0, && (20)\\
& && \pi(s, r, n) = \frac{(\bar{\gamma} p - c)\overline{s} - c\overline{r}}{n}, && (21)\\
& && n = NF(\pi(s, r, n)). && (22)
\end{aligned}$$

We present the solution of (17)–(22) in two steps. First, in Proposition 1 we solve for the optimal capacity allocation assuming an exogenously given capacity of participating drivers, n. Then, we characterize the resulting reduced form of the per-driver profit $\pi(s, r, n)$ in (21) as a function of n, and we establish in Corollary 1 that there exists a unique equilibrium capacity of participating drivers, i.e., the solution of (22). (All proofs are included in the Appendix.)

Proposition 1 (Allocation of fixed driver capacity under regime C)**.** *Consider the problem* (17)–(20)*) for fixed capacity n. Define the constants*

$$n_1^C := \overline{S} - (\Lambda_{21} - \Lambda_{12})\, t_{21} \qquad \textit{and} \qquad n_2^C := \overline{S} + (\Lambda_{21} - \Lambda_{12})\, t_{12}, \tag{23}$$

where $n_1^C < \overline{S} < n_2^C$. The optimal driver capacity utilization has the following structure.
(1) Scarce capacity ($n \le n_1^C$). All drivers serve riders: $\overline{s} = n$; $r = 0$; $q = 0$.
(2) Moderate capacity ($n_1^C < n \le n_2^C$). Drivers serve riders or reposition from the low- to the high-demand location: $\overline{s} + r_{12} = n$ where $r_{12} = t_{12}/\,(t_{12} + t_{21})\,(n - n_1^C)$, $r_{21} = 0$; $q = 0$.
(3) Ample capacity ($n > n_2^C$). Drivers serve all riders, reposition from the low- to the high-demand location, or wait in queue: $\overline{s} = \overline{S}$; $r_{12} = n_2^C - \overline{S}, r_{21} = 0$; $\overline{q} = n - n_2^C$.

Note that n_1^C is the maximum offered load that can be served without repositioning, n_2^C is the minimum capacity level required to serve the total offered load, $\overline{S}$, i.e., with the minimum amount of repositioning, and $n_1^C < \overline{S} < n_2^C$ due to the demand imbalance $\Lambda_{21} > \Lambda_{12}$ (Assumption 1).

The results in Proposition 1 are intuitive. In the scarce-capacity zone 1 ($n \le n_1^C$) drivers are serving riders 100% of the time. For $n = n_1^C$ the platform only loses the excess demand from the high- to the low-demand location, $\Lambda_{21} - \Lambda_{12}$, whereas destination-based admission control allows the platform to selectively direct capacity to serve all local requests (route-22).

In the moderate-capacity zone, the platform uses destination-based admission control and capacity repositioning to optimize performance; all drivers are moving around, so no capacity is "wasted" due to idling. Repositioning is only needed from the low- to the high-demand location.[4]

In the ample-capacity zone 3 ($n_2^C < n$), in addition to serving all riders and repositioning as needed, there is spare capacity that will ultimately idle waiting to be matched with riders. The zone-3 solution in Proposition 1 is not unique; multiple allocations of r and q support serving all demand. However, the solution with $r_{21} = 0$, i.e., no repositioning from location 2 to 1, maximizes the drivers' profit rate for given n, and supports the maximum achievable equilibrium capacity.

[4]The capacity dedicated into repositioning captures, in a discrete, two-location network, the "wild goose chase" phenomenon described in Castillo et al. (2017).

Let $\pi_C(n)$ denote the per-driver profit under the optimal capacity allocation in Proposition 1, as a function of the number of drivers n. Substituting $\overline{s}$ and $\overline{r}$ from Proposition 1 into (21) yields

$$\pi_C(n) = \frac{(\bar{\gamma}p - c)\overline{s} - c\overline{r}}{n} = \begin{cases} \bar{\gamma}p - c, & \text{zone 1 } (n \leq n_1^C), \\ \frac{1}{n}\bar{\gamma}p\left(n_1^C + (n - n_1^C)\frac{t_{21}}{t_{12}+t_{21}}\right) - c, & \text{zone 2 } (n_1^C < n \leq n_2^C), \\ \frac{1}{n}(\bar{\gamma}p\overline{S} - cn_2^C), & \text{zone 3 } (n > n_2^C). \end{cases} \tag{24}$$

This profit rate reflects the drivers' utilization profile: in zone 1 they serve riders all the time (with revenue rate $\bar{\gamma}p$ and cost rate c); in zone 2 they serve riders only a fraction of the time but still drive around all the time; and in zone 3 they also queue a fraction of the time.

It is easy to check that $\pi_C(n)$ is continuously decreasing in n and that $\lim_{n\to\infty} \pi_C(n) = 0$. The participation equilibrium equation (22) simplifies to $n = NF(\pi_C(n))$ and has a unique solution.

Corollary 1 (Driver participation equilibrium under regime C)**.** *Under centralized admission control and repositioning, there exists a unique equilibrium capacity of participating drivers, denoted by n_C^*, which solves $n_C^* = NF(\pi_C(n_C^*))$, where $\pi_C(n)$ is the per-driver profit in (24).*

4 Regimes with Decentralized Repositioning

In §4.1 we provide an explicit characterization of a driver repositioning equilibrium. Subsequently we characterize the system equilibria for two regimes: in §4.2 for Minimal Control (M) where the platform exercises no admission control, and in §4.3 for Admission Control (A) where the platform optimizes over admission control decisions. Finally, in §4.4 we summarize the key differences between the equilibria of all three regimes, C, M, and A.

4.1 Driver Repositioning Equilibrium

Recall from §2.2 that under decentralized repositioning, drivers symmetrically choose their repositioning fractions η to maximize their profit rate $\widetilde{\pi}(\eta; \lambda, w)$, and by (1) and (2) a driver repositioning equilibrium requires $\eta(\lambda, \nu) \in \arg\max_\eta \widetilde{\pi}(\eta; \lambda, w)$. That is, a set of flow rates (λ, ν) and delays w admit a driver repositioning equilibrium if, and only if, the unique repositioning fractions $\eta(\lambda, \nu)$ that are consistent with (λ, ν) are every driver's best response to (λ, w). Using (14)–(16) to map (λ, ν, w) to (s, r, q), we henceforth express the functions $\eta(\lambda, \nu)$ and $\widetilde{\pi}(\eta; \lambda, w)$ as $\eta(s, r)$ and $\widetilde{\pi}(\eta; s, q)$, respectively. Lemma 1 characterizes the profit rate function $\widetilde{\pi}(\eta; s, q)$. Its explicit expression is given in the proof in the Appendix.

Lemma 1 (Per-driver profit rate)**.** *Let the function $T(\eta; s, q)$ denote a driver's expected steady-state cycle time through the network, i.e., the average time between consecutive arrivals to the same location. Let $T^s(\eta; s)$, $T^r(\eta)$ and $T^q(\eta; s, q)$ denote the expected time a driver spends during a cycle serving riders, repositioning and queueing, respectively, where $T^s(\eta; s) + T^r(\eta) + T^q(\eta; s, q) = T(\eta; s, q)$. The drivers' expected steady-state profit rate is an explicit function that satisfies*

$$\widetilde{\pi}(\eta; s, q) = \frac{(\bar{\gamma} p - c)T^s(\eta; s) - cT^r(\eta)}{T^s(\eta; s) + T^r(\eta) + T^q(\eta; s, q)}. \tag{25}$$

We have the following formal definition of a driver repositioning equilibrium.

Definition 1 (Driver repositioning equilibrium)**.** A capacity allocation (s, r, q) forms a driver repositioning equilibrium if, and only if, $\eta\,(s, r)$ is every driver's best response, that is

$$\eta_1(s, r) = \frac{r_{12}}{s_{11}\frac{t_{12}}{t_{11}} + s_{12} + r_{12}} \quad \text{and} \quad \eta_2(s, r) = \frac{r_{21}}{s_{21} + s_{22}\frac{t_{21}}{t_{22}} + r_{21}}, \tag{26}$$

$$\eta(s, r) \in \arg\max_{\eta} \widetilde{\pi}(\eta; s, q). \tag{27}$$

Observe that under Assumption 1, it is not optimal to reposition from the high-demand location (2) to the low-demand location (1). Therefore, we focus hereafter on driver repositioning equilibria with $\eta_2 = 0$ and $r_{21} = 0$. Condition (27) in Definition 1 yields an explicitly defined set of capacity allocations (s, r, q) that admit a driver repositioning equilibrium. We call this the set of *driver-incentive compatible* capacity allocations, denoted by $\mathcal{D}$, and specify it in Proposition 2.

Proposition 2 (Driver-incentive compatibility)**.** *There exists a driver repositioning equilibrium with $\eta_2 = 0$ if, and only if, the capacity allocation (s, r, q) is driver-incentive compatible, that is:*

$$(s, r, q) \in \mathcal{D} := \left\{ (s, r, q) \geq 0 : r_{21} = 0,\ q_1 \begin{cases} \leq q_1^*(s) + k(s)q_2 & \text{if } r_{12} = 0 \\ = q_1^*(s) + k(s)q_2 & \text{if } r_{12} > 0 \end{cases} \right\}, \tag{28}$$

where $q_1^*(s)$ and $k(s)$ are specified in (52), and $q_1^*(s) = k(s) = 0$ if $s_{11} + s_{12} = 0$, and $q_1^*(s), k(s) > 0$ if $s_{11} + s_{12} > 0$. For $(s, r, q) \in \mathcal{D}$ the unique driver repositioning equilibrium is given by (26).

The structure of the driver-incentive compatible capacity allocation set $\mathcal{D}$ is intuitive. Inducing drivers not to reposition from location 1 (i.e., $r_{12} = 0$) requires a relatively short location-1 queue, i.e., $q_1 \leq q_1^*(s) + k(s)q_2$. Inducing drivers to reposition from location 1 (i.e., $r_{12} > 0$) requires appropriately balanced queues in the two locations[5], i.e., $q_1 = q_1^*(s) + k(s)q_2$; in this case, drivers are indifferent between repositioning to location 2 and queueing at location 1, and (26) identifies the unique randomization probability for the corresponding symmetric driver repositioning equilibrium. Proposition 2 also foreshadows the critical role that demand admission control plays in shaping drivers' repositioning incentives, to be discussed in detail in §4.3.

In light of Proposition 2, constraint (28) alone suffices to account for decentralized repositioning in the problem formulations of regimes M and A: Given a driver-incentive capacity allocation, the corresponding repositioning equilibrium fractions are immediately determined by (26).

4.2 Minimal Control (M)

In the minimal control regime (M) the platform exercises no admission control and drivers make their own repositioning and participation decisions. We express the feasible set $\mathcal{M}$, given by (13) as the set of feasible capacity allocations (s, r, q, n). Using (14)–(16) the flow constraints (10)–(12) are equivalent to

$$\begin{aligned} \frac{s_{l1}}{S_{l1}} &= \frac{s_{l2}}{S_{l2}}, \quad l = 1, 2, && (29) \\ (S_{l1} + S_{l2} - s_{l1} - s_{l2})r_{lk} &= 0, \quad l = 1, 2, k \neq l, && (30) \\ (S_{l1} + S_{l2} - s_{l1} - s_{l2})q_l &= 0, \quad l = 1, 2. && (31) \end{aligned}$$

Substituting (28) for (1)–(2) based on Proposition 2, and noting that (4)–(8) correspond to (18)–(22), the set of feasible capacity allocations under minimal control is given by

$$\mathcal{M} = \{(s, r, q, n) \ : \ (18) - (22), \ (29) - (31), \ (28)\}. \tag{32}$$

We specify the system equilibrium in two steps. First, we characterize in Proposition 3 the unique equilibrium capacity allocation of a fixed number n of participating drivers. Then, we establish in Corollary 2 that there exists a unique equilibrium capacity of participating drivers.

Proposition 3 (Allocation of fixed capacity under regime M)**.** *Consider the set $\mathcal{M}$ of feasible capacity allocations, defined in (32), for fixed capacity n. Define the constants*

$$n_1^M := n_1^C - \left(1 - \frac{\Lambda_{12}}{\Lambda_{21}}\right) S_{22}, \quad n_2^M := n_1^M + q_1^*(S), \quad \textit{and} \quad n_3^M := n_2^C + q_1^*(S), \tag{33}$$

[5]Observe that the conditions $r_{12} > 0$ and $q_1 > q_1^*(s) + k(s)q_2$ are mutually incompatible. The latter condition implies that every driver repositions from location 1, but then no location-1 demand is served so that $q_1 = 0$.

where n_1^C and n_2^C are defined in (23) and $n_1^M < n_1^C < \overline{S} < n_2^C < n_3^M$. There is a unique feasible driver capacity utilization (s, r, q) satisfying (18)–(20), (29)–(31) and (28), with the following structure.

(1) Scarce capacity ($n \le n_1^M$). All drivers serve riders: $\overline{s} = n$; $r = 0$; $q = 0$.

(2) Moderate capacity—no repositioning but queueing ($n_1^M < n \le n_2^M$). Drivers serve all riders at the low- and a fraction $\frac{\Lambda_{12}}{\Lambda_{21}}$ of riders at the high-demand location, or queue at the low-demand location: $\overline{s} = n_1^M$ where $s_{1k} = S_{1k}$, $s_{2k} = S_{2k}\frac{\Lambda_{12}}{\Lambda_{21}}$ for $k = 1, 2$; $r = 0$; $q_1 = n - n_1^M < q_1^*(S)$, $q_2 = 0$.

(3) Moderate capacity—repositioning and queueing ($n_2^M < n \le n_3^M$). Drivers serve all riders at the low- and more than a fraction $\frac{\Lambda_{12}}{\Lambda_{21}}$ of riders at the high-demand location, reposition from the low- to the high-demand location, or queue at the low-demand location: $\overline{s} > n_1^M$ where $s_{1k} = S_{1k}$ for $k = 1, 2$; $r_{12} > 0$, $r_{21} = 0$; $q_1 = q_1^*(S)$, $q_2 = 0$.

(4) Ample capacity ($n > n_3^M$). Drivers serve all riders, reposition from the low- to the high-demand location, or queue at both locations in an incentive-compatible split: $\overline{s} = \overline{S}$; $r_{12} = n_2^C - \overline{S}$, $r_{21} = 0$; $q_1 = q_1^*(S) + k(S)q_2$, $q_2 > 0$.

Proposition 3 differs in two ways from Proposition 1 for Centralized Control (regime C). First, because the platform exercises no admission control in regime M, the maximum offered load that can be served without repositioning, given by n_1^M, is lower than under optimal admission control (regime C), i.e., $n_1^M = n_1^C - (1 - \Lambda_{12}/\Lambda_{21})\, S_{22}$. Specifically, with destination-based admission control the platform could direct the scarce capacity to serve all local traffic at the high-demand location. Second, and most importantly, because repositioning is decentralized in regime M, drivers can only be induced to reposition to the high-demand location if there is a sufficiently long queueing delay at the low-demand location. So, at lower capacity (zone 2), the low-demand queue builds up, but is not sufficient to incentivize drivers to reposition to the high-demand location. Only at sufficiently high capacity levels (zone 3) is the queue long enough for drivers to reposition. Since queueing is required to get sufficient repositioning to occur, the minimum capacity level required to serve the total offered load, n_3^M, exceeds the corresponding requirement under centralized control, n_2^C, by exactly the size of the queue that will provide the repositioning incentive.

Let $\pi_M(n)$ denote the per-driver profit under the equilibrium capacity allocation of Proposition 3, as a function of the number of drivers n. (See the Proof of Corollary 2 in the Appendix for an explicit expression of $\pi_M(n)$.) It is easy to verify that $\pi_M(n)$ is continuously decreasing in n and that $\lim_{n\to\infty} \pi_M(n) = 0$. The participation equilibrium equation (22), which simplifies to $n = NF(\pi_M(n))$, therefore has a unique solution.

Corollary 2 (Driver participation equilibrium under regime M)**.** *Under no admission control and decentralized repositioning, there exists a unique equilibrium capacity of participating drivers, denoted by* n_M^**, which solves* $n_M^* = NF(\pi_M(n_M^*))$*.*

4.3 Admission Control (A)

4.3.1 Network Equilibrium

In regime A the platform is free to choose the optimal admission control policy, while drivers make their own repositioning and participation decisions. The corresponding capacity allocation problem, the analog of (9), is given by

$$\text{(Problem A)} \quad \max_{s,r,q} \left\{ \mathit{\Pi}\left(s\right) \ : \ (18) - (22),\ (28) \right\}. \tag{34}$$

Contrasting with (32), the platform's control need not satisfy (29)–(31). We proceed as before in two steps.

Proposition 4 (Allocation of fixed capacity under regime A)**.** *Consider the problem (34) for fixed capacity* n*. Define the constants*

$$n_1^A := n_1^C \qquad and \qquad n_3^A := n_2^C + q_1^*\left(S\right), \tag{35}$$

where n_1^C *and* n_2^C *are defined in (23) and* $n_1^A = n_1^C < \overline{S} < n_2^C < n_3^A$*. There exists a threshold* n_2^A *such that* $n_1^A < n_2^A < n_3^A$ *and the optimal driver capacity utilization has the following structure.*
(1) <u>*Scarce capacity*</u> *(*$n \le n_1^A$*). All drivers serve riders:* $\overline{s} = n;\ r = 0;\ q = 0$*.*
(2) <u>*Moderate capacity—no repositioning but queueing*</u> *(*$n_1^A < n \le n_2^A$*). Drivers serve all riders except a fraction* $1 - \frac{\Lambda_{12}}{\Lambda_{21}}$ *from the high- to the low-demand location, and queue at the low-demand location:* $\overline{s} = n_1^A;\ r = 0;\ q_1 = n - n_1^A < q_1^*\left(s\right),\ q_2 = 0$*.*
(3) <u>*Moderate capacity—repositioning, with or without queueing*</u> *(*$n_2^A < n \le n_3^A$*). Compared to zone 2, drivers serve more riders at the high- but possibly fewer riders at the low-demand location, they reposition from the low- to the high-demand location, and may queue at the low-demand location:* $\overline{s} > n_1^A;\ r_{12} > 0,\ r_{21} = 0;\ q_1 = q_1^*(s) \ge 0,\ q_2 = 0$*.*
(4) <u>*Ample capacity*</u> *(*$n > n_3^A$*): Drivers serve all riders, reposition from the low- to the high-demand location, or queue at both locations in an incentive-compatible split:* $\overline{s} = \overline{S};\ r_{12} = n_2^C - \overline{S},\ r_{21} = 0;\ q_1 = q_1^*(S) + k(S)q_2,\ q_2 > 0$*.*

Proposition 4 differs in two ways from Proposition 3 for Minimal Control (regime M). First, because the platform exercises admission control in regime A, the maximum offered load that can be served without repositioning is the same as under Centralized Control, i.e., $n_1^A = n_1^C$, and exceeds its counterpart under Minimal Control, i.e., $n_1^M < n_1^A$, as discussed in §4.2.

Second and more importantly, whereas under Minimal Control the demand served on each route increases in the total available driver capacity n, this may *not* hold under optimal admission control: Compared to zone 2, in zone 3 the platform may reject, and therefore serve *fewer,* rider requests at the low-demand location even if there are available drivers, and even though there is more capacity in the network. The idea is to make it less attractive for drivers to queue at the low-demand location by decreasing the served demand rate, rather than by relying on the buildup of a long queue. This frees cars to reposition and subsequently serve riders.[6] We term this policy feature *strategic demand rejection* as it reduces the revenue at the low-demand location, in order to incentivize drivers to reposition and generate more revenue at the high-demand location. We elaborate on the rationale and specify optimality conditions for this key result in §4.3.2.

Let $\pi_A(n)$ denote the per-driver profit under the equilibrium capacity allocation of Proposition 4, as a function of the number of drivers n. (See the Proof of Corollary 3 in the Appendix for an explicit expression of $\pi_A(n)$.) The function $\pi_A(n)$ may be discontinuous (if the platform rejects riders at the low-demand location), but it is decreasing with $\lim_{n\to\infty} \pi_A(n) = 0$, and Corollary 3 confirms the existence of a unique participation equilibrium.

Corollary 3 (Driver participation equilibrium under regime A)**.** *Under optimal admission control and decentralized repositioning, there exists a unique equilibrium capacity of participating drivers, denoted by* n_A^**, which satisfies* $NF(\pi_A(n_A^{*+})) \leq n_A^* \leq NF(\pi_A(n_A^{*-}))$*, where* $n_A^{*+} = \lim_{\epsilon\downarrow 0} n_A^* + \epsilon$ *and* $n_A^{*-} = \lim_{\epsilon\downarrow 0} n_A^* - \epsilon$*.*

4.3.2 Strategic Demand Rejection to Induce Driver Repositioning

Proposition 4 states that under moderate-capacity conditions (zone 3, $n_2^A < n \leq n_3^A$), drivers may serve *fewer* riders at the low-demand location (1), compared to when there is less capacity in the network (zone 2). In this case the optimal admission control policy exhibits a somewhat counterintuitive behavior, whereby the platform rejects some or all rider requests in the low-demand location, even though there is an excess supply of drivers, that is, empty cars are leaving and possibly also waiting to get matched at this location. This strategic demand rejection sacrifices revenue at

[6] This property of the optimal platform's control seems to be in contrast to the literature in ride-hailing networks.

the low-demand location, so as to incentivize drivers to reposition from the low- to the high-demand location where they can generate more revenue for themselves and for the platform. Specifically, rejecting rider requests at the low-demand location creates an artificial demand shortage that drivers offset by choosing to reposition more frequently to the high-demand location, rather than joining the low-demand matching queue. The end result is a shorter queue at the low-demand location (the waiting time may increase or decrease). In terms of Proposition 2, rejecting demand at location 1 alters the driver-incentive compatible capacity allocation by reducing the queue-length threshold $q_1^*(s)$, which frees up driver capacity to reposition and serve riders at the high-demand location. By controlling congestion, the platform has an operational lever to incentivize drivers to reposition, as opposed to, for example, increasing their wage.

This control action can only be relevant when capacity is moderate; when capacity is scarce all drivers are busy; when capacity is ample, all riders are served. For the moderate-capacity zone (3), Proposition 5 identifies a necessary and sufficient condition for the optimality of strategic demand rejection in terms of the model primitives. To simplify notation and highlight the structural imbalances, define the following ratios:

$$\rho_1 \coloneqq \frac{S_{11}}{S_{11}+S_{12}},\ \rho_2 \coloneqq \frac{S_{22}}{S_{21}+S_{22}},\ \tau = \frac{t_{21}}{t_{12}},\ \kappa = \frac{c}{\bar{\gamma}p} < 1, \tag{36}$$

where ρ_1 and ρ_2 are the shares of the local-demand offered load at location 1 and 2, respectively, τ is ratio between cross-location travel times, and κ is the ratio of driving cost to drivers' service revenue ("relative driving cost"). Assumption 2 requires that $\kappa < \tau/(1+\tau)$.

Proposition 5 (Optimality of strategic demand rejection in regime A)**.** *Under optimal platform admission control and decentralized repositioning, it is optimal at moderate capacity, i.e., for some* $n \in (n_2^A, n_3^A]$*, to strategically reject rider requests at the low-demand location so as to induce repositioning to the high-demand location, if and only if the following condition holds:*

$$\frac{\Lambda_{12}}{\Lambda_{21}}\frac{1-\rho_1\kappa}{1-\rho_1} < \frac{\tau-(\tau+1-\rho_2)\kappa}{1-\rho_2}\left(\kappa\frac{1+\tau}{\tau}\frac{\tau+1-\rho_2}{\rho_2}-1\right). \tag{37}$$

To gain some intuition we discuss (37) under the simplifying assumption that $\tau = 1$.

Effect of ρ_2*, the share of the local-demand offered load at the high-demand location.*[7] Intuitively, if the local-demand share at the high-demand location ρ_2 is low, then this location is less attractive

[7]For $\tau = 1$ and $\kappa > 0$ the right-hand side (RHS) of (37) decreases in ρ_2 from $+\infty$ for $\rho_2 = 0$ to $-\infty$ as $\rho_2 \to 1$, so that (37) holds for ρ_2 below some threshold.

because of the higher likelihood that a driver will get matched to a rider going back to the low-demand location; drivers have a weak "natural" incentive to reposition to the high-demand location, so encouraging them to do so requires rejecting demand at the low-demand location.

Effect of ρ_1, the share of the local-demand offered load at the low-demand location.[8] Holding ρ_2 fixed, which fixes the RHS of (37), the condition cannot hold for sufficiently large ρ_1, i.e., if the local demand at the low-demand location is dominant: In this case drivers in the low-demand location may get stuck serving local requests and queueing in between, which adversely affects their profit rate and makes repositioning naturally more attractive.

Effect of $\Lambda_{12}/\Lambda_{21}$, the cross-location demand imbalance. The LHS of (37) is positive and decreases to zero as Λ_{21} increases from Λ_{12} to ∞, where $\Lambda_{12}/\Lambda_{21} < 1$ by Assumption 1. Therefore, (37) holds for sufficiently large Λ_{12}, provided the RHS is strictly positive (i.e., the local-demand share at the high-demand location, ρ_2, is below some threshold, as discussed above). Intuitively, more cross-location demand at the high-demand location increases the value of rejecting demand at the low-demand location in order to induce drivers to reposition.

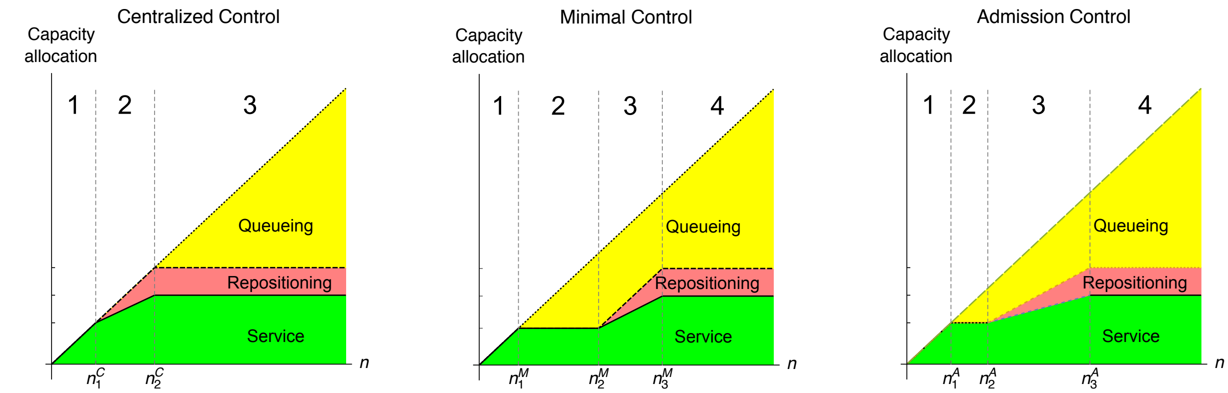
Effect of κ, the relative driving cost. (37) can only hold for sufficiently large $\kappa > 0$. When repositioning becomes significantly more expensive than queueing (for which drivers incur no direct cost), the platform needs to strengthen the incentive for repositioning over queueing at the low-demand location by rejecting demand there.

4.4 Graphical Depiction of Control Regimes C, M, and A

Figures 4 and 5 depict the broad features that are identified in Propositions 1, 3 and 4 for the three control regimes. (1) When capacity is scarce, all drivers are busy serving riders under all three control regimes. (2) When capacity is ample, all riders are served, and the three control regimes agree again. But, importantly, (3) with optimal admission control and centralized repositioning the platform can serve the entire rider demand with less capacity and without any drivers queueing at any location; in contrast, with decentralized repositioning, in order to create the appropriate incentives, drivers need to queue. (4) With admission control, the platform can (*i*) prioritize rider demand at the high-demand location based on their destination, and therefore increase driver utilization in the absence of repositioning, and (*ii*) reject rider demand at the low-demand location to incentivize driver repositioning.

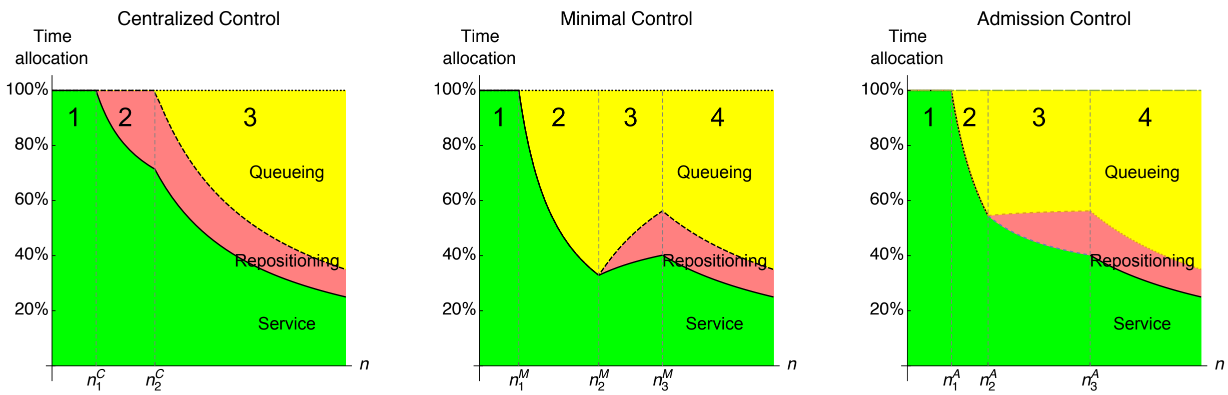
Switching to the driver's view, additional platform control capability ensures that drivers are

[8] The left-hand side (LHS) of (37) increases in ρ_1 from $\Lambda_{12}/\Lambda_{21}$ for $\rho_1 = 0$ to ∞ as $\rho_1 \to 1$, so that condition (37) holds if both local-demand shares, ρ_1 and ρ_2, are below some threshold.



(a) Regime C (b) Regime M (c) Regime A

Figure 4: Optimal capacity allocation in three control regimes



(a) Regime C (b) Regime M (c) Regime A

Figure 5: Driver time allocation in three control regimes

busy for a larger fraction of their time, and thus more profitable.

5 The Impact of Platform Controls on System Performance

In this section we compare the equilibrium performance of the three control regimes $\{M, A, C\}$ from the viewpoint of the platform, the drivers, and the rider service level.

5.1 Ranking of Platform Revenue, Per-Driver Profit, and Driver Capacity

We start by ranking the platform revenue, the per-driver profit, and the driver capacity under the three control regimes. Let $\Pi_X(n)$ denote the platform revenue rate under the equilibrium capacity allocation in regime $X \in \{M, A, C\}$, as a function of the participating driver capacity n.

Proposition 6 establishes that for fixed participating capacity both the platform and the drivers are better off with increasing platform control capabilities.

Proposition 6 (Ranking of equilibrium profits for fixed capacity)**.** *For fixed participating capacity n control capabilities have the following impact on profits:*
(1) The platform revenue rate increases with increasing platform control capability ($M \to A \to C$):

$$\Pi_M(n) \leq \Pi_A(n) \leq \Pi_C(n). \tag{38}$$

(2) Centralized control maximizes the per-driver profit rate:

$$\max\{\pi_M(n), \pi_A(n)\} \leq \pi_C(n). \tag{39}$$

(3) Under decentralized repositioning, if (37) is not satisfied then optimal admission control increases the per-driver profit rate:

$$\pi_M(n) \leq \pi_A(n). \tag{40}$$

Parts (1) and (2) of Proposition 6 are as expected. Part (1) of Proposition 6 also implies that riders benefit from increasing platform control capability: From the riders' viewpoint, an important performance metric is the network-wide service level, defined as the fraction of the total rider demand that is served. This service level equals the ratio of total service capacity to total offered load, i.e., $\overline{s}/\overline{S}$. Since the platform revenue rate is proportional to the total service capacity, i.e., $\Pi(s) = \gamma p \overline{s}$ by (17), the network-wide service level is proportional to the platform revenue rate, and therefore increases with increasing platform control capability by Part (1) of Proposition 6.

Part (3) of Proposition 6 establishes that optimal admission control typically benefits drivers, provided that condition (37) is *not* satisfied. Conversely, if condition (37) holds then by Proposition 5 the platform chooses to strategically reject demand in the low-demand location at moderate capacity levels to induce driver repositioning; as a result, drivers incur higher driving costs and may be worse off than without admission control (the platform is still better off).

Proposition 7 establishes that the equilibrium driver participation increases with increasing platform control capabilities, and as under fixed capacity, so do the resulting platform revenue and per-driver profits. Let $\Pi_X^* := \Pi_X(n_X^*)$ and $\pi_X^* := \pi_X(n_X^*)$ denote, respectively, the platform revenue rate and the per-driver profit rate, under the equilibrium capacity allocation and levels in regime $X \in \{M, A, C\}$.

Proposition 7 (Ranking of equilibrium profits and capacity)**.** *(1) The equilibrium platform revenue rate increases with increasing platform control capability ($M \to A \to C$):*

$$\Pi_M^* \leq \Pi_A^* \leq \Pi_C^*. \tag{41}$$

(2) Centralized control maximizes the equilibrium driver participation and per-driver profit rate:

$$\max\{n_M^*, n_A^*\} \leq n_C^* \quad \textit{and} \quad \{\pi_M^*, \pi_A^*\} \leq \pi_C^*. \tag{42}$$

(3) Under decentralized repositioning, if (37) is not satisfied, then optimal admission control increases drivers' participation and profit rate:

$$n_M^* \leq n_A^* \quad \textit{and} \quad \pi_M^* \leq \pi_A^*. \tag{43}$$

Parts (2) and (3) follow from Parts (2) and (3) of Proposition 6, respectively, and because the marginal opportunity cost function, defined as $c_o(n) := F^{-1}\left(\frac{n}{N}\right)$, strictly increases in the participating capacity n. Part (1) follows because the platform revenue functions $\Pi_C(n)$, $\Pi_M(n)$ and $\Pi_A(n)$ are increasing in n. Propositions 1, 3, and 4 imply that the performance improvements reported in Proposition 7 are strictly positive if, and only if, the equilibrium driver capacity is in the moderate-capacity zone where increased control capability is effective.

Corollary 4 (Conditions for performance gains)**.** *Platform controls strictly improve profits under the following equilibrium conditions:*
(1) Admission control (regime A over M): If (37) is not satisfied then

$$\Pi_M^* < \Pi_A^*,\ \pi_M^* < \pi_A^* \textit{ and } n_M^* < n_A^* \quad \textit{if and only if} \ \ n_M^* \in (n_1^M, n_3^M), \tag{44}$$

where n_1^M and n_3^M are defined in (33).
(2) Centralized repositioning control (regime C over A):

$$\Pi_A^* < \Pi_C^*,\ \pi_A^* < \pi_C^* \textit{ and } n_A^* < n_C^* \quad \textit{if and only if} \ \ n_A^* \in (n_1^A, n_3^A), \tag{45}$$

where n_1^M and n_3^M are defined in (35).

The equilibrium iff capacity conditions in (44) are satisfied when the potential driver supply and the outside opportunity cost distribution are such that (i) more drivers are willing to participate at the maximum profit rate $(\overline{\gamma}p - c)$ than are needed to meet the rider demand that can be served without repositioning (so $n_M^* > n_1^M$), but (*ii*) not enough drivers are willing to participate at the reduced profit rate resulting from the queueing delays and repositioning costs that are necessary to serve all riders (so $n_M^* < n_3^M$). Both of these conditions seem plausible, and as such we expect admission control to benefit the platform and the drivers in practical settings. A similar argument indicates that in practical parameter regimes the iff capacity conditions in (45) also hold, i.e., that centralized repositioning is beneficial.

5.2 Upper Bounds on the Gains in Platform Revenue and Per-Driver Profit

In this section we provide upper bounds on the gains in platform revenue and per-driver profit due to increased platform control as explicit functions of the network primitives ρ_1, ρ_2, and τ defined in (36) across all feasible driver outside opportunity cost rate distributions.

Proposition 8 (Upper bounds on platform revenue gains)**.** *Fix $N \geq n_3^M = n_3^A$.*
(1) Platform revenue gain due to admission control (regime A over M): If (37) is not satisfied,

$$\max_{F(\cdot)} \frac{\Pi_A^* - \Pi_M^*}{\Pi_M^*} \leq \frac{\overline{S}}{n_1^M} - 1 = \left(\frac{\Lambda_{21}}{\Lambda_{12}} - 1\right) \frac{1}{1 + \frac{1-\rho_2}{1-\rho_1}\frac{1}{\tau}}. \tag{46}$$

(2) Platform revenue gain due to centralized repositioning control (regime C over A):

$$\max_{F(\cdot)} \frac{\Pi_C^* - \Pi_A^*}{\Pi_A^*} \leq \frac{\overline{S}}{n_1^A} - 1 = \left(\frac{\Lambda_{21}}{\Lambda_{12}} - 1\right) \frac{1}{1 + \frac{1}{1-\rho_1}\frac{1}{\tau} + \frac{\rho_2}{1-\rho_2}\frac{\Lambda_{21}}{\Lambda_{12}}}. \tag{47}$$

The key insight is that the potential revenue gains from both admission control and centralized repositioning increase as a function of the cross-location demand imbalance $(\Lambda_{21}/\Lambda_{12})$, and the cross-location travel time imbalance $(\tau = t_{21}/t_{12})$. The condition $N \geq n_3^M = n_3^A$ requires that the potential driver supply is large enough to serve all riders under decentralized repositioning. The upper bound on the gain from admission control (the RHS in (46)) is attained if, under minimal control the platform can only meet the rider demand that does not require repositioning, and admission control allows the platform to increase driver participation enough to serves all riders. The upper bound on the gain from repositioning control in (47) has a similar interpretation. The upper bound in (46) exceeds the bound in (47), because regime A allows the platform to serve more riders without repositioning. The bounds in (46) and (47) can be approached arbitrarily closely for specific choices of the opportunity cost distribution $F(\cdot)$, as discussed in Section S2 of the Supplemental Materials.

Table 1 numerically illustrates the magnitude of the upper bounds in (46) and (47) as a function of the cross-location demand imbalance $\Lambda_{21}/\Lambda_{12}$.

cross demand imbalance $(\frac{\Lambda_{21}}{\Lambda_{12}})$	1	2	5	10
from admission control	0%	43%	150%	319%
from central. repositioning	0%	25%	100%	225%

(a) Balanced cross-local demand at low-demand location ($\rho_1 = 0.5$)

$(\frac{\Lambda_{21}}{\Lambda_{12}})$	1	2	5	10
from admission control	0%	53%	189%	407%
from central. repositioning	0%	30%	120%	270%

(b) Imbalanced cross-local demand at low-demand location ($\rho_1 = 0.25$)

Table 1: Upper bounds in (46) and (47) on platform revenue gain ($t_{lk} = 1, \forall lk$, $\Lambda_{12} = \Lambda_{22} = 1$)

Turning attention to the per-driver profit gain, we have the following result.

Proposition 9 (Upper bound on per-driver profit gains)**.** *Fix $N \geq n_3^M = n_3^A$ and assume that (37) is not satisfied. The per-driver profit gain from admission control (under regime A or C) satisfies:*

$$\max_{F(\cdot)} \frac{\pi_A^* - \pi_M^*}{\pi_M^*} = \max_{F(\cdot)} \frac{\pi_C^* - \pi_M^*}{\pi_M^*} \leq \frac{1-\rho_2}{\tau - (1-\rho_2+\tau)\kappa}. \tag{48}$$

In contrast to the bounds on the platform revenue gains in Proposition 8 that can only be attained when more control yields repositioning, the bound in (48) can only be attained in the *absence* of repositioning, i.e., when admission control increases drivers' utilization to 100% with only a small increase in their participation. In the online appendix we illustrate this tension between the drivers' and the platform's gains from control, along with the properties of the opportunity cost distribution $F(\cdot)$ that are required to attain the bound in (48). In a nutshell, if a small change in per-driver profit rate increases the capacity of participating drivers significantly, then the platform may extract significant gains while drivers are only marginally better off; conversely, if it takes a large change in per-driver profit rate to attract incremental participating driver capacity, then drivers extract significant incremental profits, while the platform benefit is moderate.

5.3 Value of Platform Control: Numerical Illustration

We conclude with a numerical illustration of the value of platform control. Taking Minimal Control with commission rate $\gamma = 0$ as the base case, we consider the performance effects of Centralized Control with $\gamma = 0$ and $\gamma = 0.2$. We assume unit travel times, i.e., $t_{lk} = 1$ for $l, k = 1, 2$, fix the price per unit time $\$p = 3$, the driver cost rate $\$c = 1$, the offered load vector $S = (1, 1, x, x)$ and consider the cases $x = 2$ and $x = 5$, corresponding to moderate and large demand imbalance, respectively. We assume that drivers' outside opportunity costs are uniformly distributed on $[0, p - c]$.

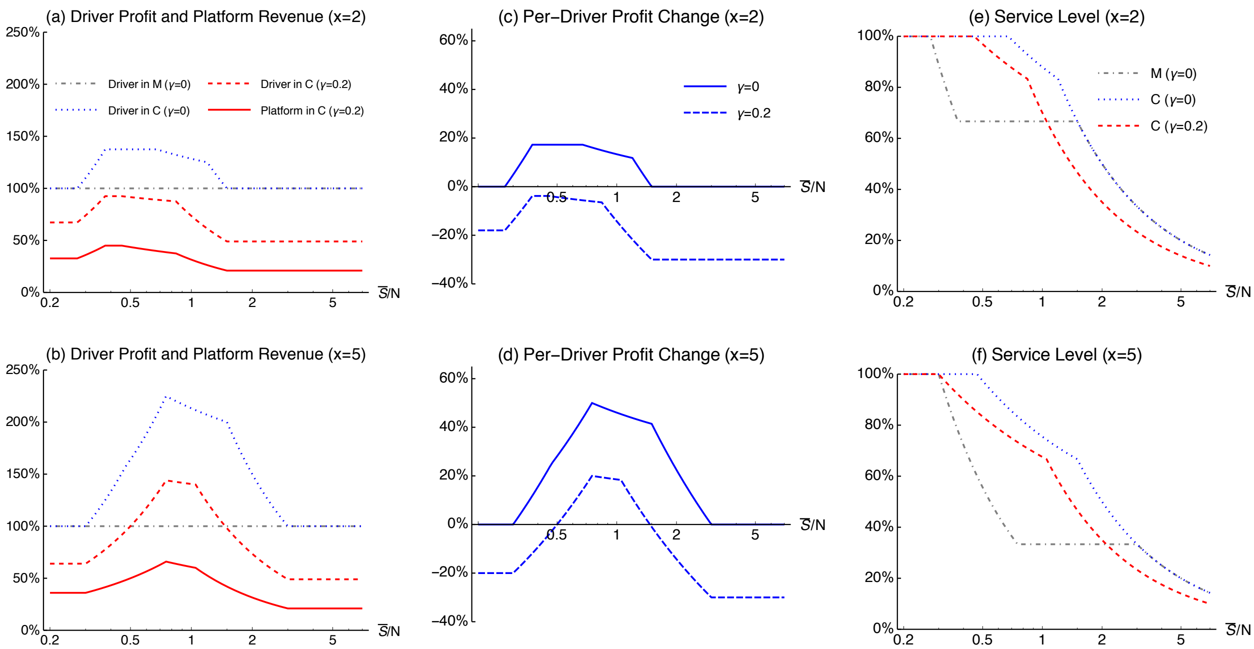


Figure 6: Impact of Centralized vs. Minimal control on performance ($x = S_{2k}/S_{1k}, k = 1, 2$)

The graphs in Figure 6 depict the effects of Centralized Control on four equilibrium performance measures, the total driver profit and platform revenue (left panel), the per-driver profit rate (center panel), and the rider service level (right panel), in the top row for cross-location imbalance $S_{21} = 2$ and in the bottom row for $S_{21} = 5$. These measures are shown as functions of the offered-load-to-driver-pool ratio $\bar{S}/N$, where $\bar{S}$ is fixed. Because $F(p - c) = 1$, all drivers in the pool would participate at the maximum achievable profit rate $p - c$, i.e., if they could serve riders all the time and the platform extracted no commission ($\gamma = 0$). However, due to the cross-location demand imbalance, serving all riders involves repositioning (plus queueing under Minimal Control) which reduces the per-driver profit rate below $p - c$, and requires more participating cars than demand (so $n > \bar{S}$), so that serving all riders in equilibrium is only feasible if $\bar{S}/N < 1$.

We make the following observations, which appear robust to other system parameters (and possibly to other general network structures).

Total driver profit and platform revenue. Panels (a) and (b) show the effects of Centralized Control on the total driver profit and platform revenue, in percentage terms relative to Minimal Control with $\gamma = 0$ (100%). Irrespective of the commission rate, both the drivers and the platform are best off at intermediate values of the offered-load-to-driver-pool ratio $\bar{S}/N$. At intermediate values of $\bar{S}/N$, drivers are always better off under Centralized Control with a zero commission rate ($\gamma = 0$), but are only better off when the cross network imbalance is significant ($x = 5$); it is natural to assume that $\bar{S}/N$ is itself moderate, say in the interval $(.5, 2)$, since otherwise the platform may choose to change the price that riders pay to increase or to reduce demand, respectively.

Per-driver profit vs. total driver profit. Panels (c) and (d) show the effects of Centralized Control on the per-driver profit, in percentage terms relative to Minimal Control with $\gamma = 0$. Comparing with the total driver profits, we see that the change in the per-driver profit is always significantly lower in magnitude relative to the change in total driver profits, because participation is endogenous and large relative profit gains will be spread among an increasing supply of participating drivers, or relative profit losses will be moderated by a decreasing supply.

Rider service level. Panels (e) and (f) show the rider service level. In the Minimal Control regime, the service level drops significantly when the load ratio $\bar{S}/N$ increases; this is aggravated when the demand imbalance is higher. Centralized platform control boosts the service level significantly at moderate values of $\bar{S}/N$. However, at at higher values of $\bar{S}/N$, when capacity is scarce, the adverse effect of the commission ($\gamma = 0.2$) on driver participation may dominate the benefits of Centralized Control, resulting in a lower service level; e.g., see panel (e) for $\bar{S}/N > 1$.

In summary, in the presence of demand imbalance, platform control capabilities can generate significant value for all parties at moderate load ratios, provided that the commission rate is commensurate with the extra gross revenue that better control yields for the drivers.

References

Adan I, Weiss G (2012) Exact fcfs matching rates for two infinite multitype sequences. *Operations Research* 60(2):475–489, URL http://dx.doi.org/10.1287/opre.1110.1027.

Bai J, So KC, Tang CS, Chen X, Wang H (2017) Coordinating supply and demand on an on-demand service platform with impatient customers, URL https://ssrn.com/abstract=2831794, working paper.

Banerjee S, Freund D, Lykouris T (2017) Pricing and optimization in shared vehicle systems: An approximation framework, URL https://arxiv.org/abs/1608.06819, working paper.

Banerjee S, Riquelme C, Johari R (2016) Pricing in ride-share platforms: A queueing-theoretic approach, URL https://ssrn.com/abstract=2568258, working paper.

Benjaafar S, Kong G, Li X, Courcoubetis C (2015) Peer-to-peer product sharing: Implications for ownership, usage and social welfare in the sharing economy, URL https://ssrn.com/abstract=2669823, working paper.

Bimpikis K, Candogan O, Saban D (2017) Spatial pricing in ride-sharing networks, URL https://ssrn.com/abstract=2868080, working paper.

Braverman A, Dai J, Liu X, Ying L (2017) Empty-car routing in ridesharing systems, URL https://arxiv.org/abs/1609.07219v2, working paper.

Buchholz N (2017) Spatial equilibrium, search frictions and efficient regulation in the taxi industry, working paper.

Bušić A, Gupta V, Mairesse J (2013) Stability of the bipartite matching model. *Advances in Applied Probability* 45(2):351–378, ISSN 00018678, URL http://www.jstor.org/stable/43563305.

Cachon GP, Daniels KM, Lobel R (2017) The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Operations Management* 19(3):368–384, URL http://dx.doi.org/10.1287/msom.2017.0618.

Caldentey R, Kaplan EH, Weiss G (2009) Fcfs infinite bipartite matching of servers and customers. *Advances in Applied Probability* 41(3):695–730, ISSN 00018678, URL http://www.jstor.org/stable/27793900.

Castillo JC, Knoepfle DT, Weyl EG (2017) Surge pricing solves the wild goose chase, URL https://ssrn.com/abstract=2890666, working paper.

Chen MK, Sheldon M (2015) Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform, working paper.

Feng G, Kong G, Wang Z (2017) We are on the way: Analysis of on demand ride-hailing systems, working paper.

Guda H, Subramanian U (2017) Strategic surge pricing on on-demand service platforms, URL https://ssrn.com/abstract=2895227, working paper.

Gurvich I, Lariviere M, Moreno A (2016) Operations in the on-demand economy: Staffing services with self-scheduling capacity, URL https://ssrn.com/abstract=2336514, working paper.

Gurvich I, Ward A (2014) On the dynamic control of matching queues. *Stochastic Systems* 4(2):479–523, URL http://dx.doi.org/10.1287/13-SSY097.

Hall JV, Horton JJ, Knoepfle DT (2017) Labor market equilibration: Evidence from uber, URL http://john-joseph-horton.com/papers/uber_price.pdf, working paper.

He L, Mak HY, Rong Y, Shen ZJM (2017) Service region design for urban electric vehicle sharing systems. *Manufacturing & Service Operations Management* 19(2):309–327, URL http://dx.doi.org/10.1287/msom.2016.0611.

Hu M, Zhou Y (2016) Dynamic type matching, URL https://ssrn.com/abstract=2592622, working paper.

Hu M, Zhou Y (2017) Price, wage and fixed commission in on-demand matching, URL https://ssrn.com/

abstract=2949513, working paper.

Iglesias R, Rossi F, Zhang R, Pavone M (2017) A BCMP network approach to modeling and controlling autonomous mobility-on-demand systems, URL https://arxiv.org/abs/1607.04357v3, working paper.

Ozkan E, Ward AR (2017) Dynamic matching for real-time ridesharing, URL https://ssrn.com/abstract=2844451, working paper.

Taylor T (2017) On-demand service platforms, URL https://ssrn.com/abstract=2722308, working paper.

Appendix. Proofs.

Proof of Proposition 1. We start with the following observations about the optimal solution.

(i) *Allocating all capacity n towards serving riders (i.e., $r = 0, q = 0$) is feasible (hence optimal) if and only if $n \leq n_1^C$.* To see this, let

$$S_1^C = (S_{11}, S_{12}, S_{12}\frac{t_{21}}{t_{12}}, S_{22}). \tag{49}$$

Then $n_1^C = S_1^C \cdot 1 = \overline{S} - (\Lambda_{21} - \Lambda_{12})t_{21}$ is the maximum service capacity without repositioning ($r = 0$). Therefore with $r = 0$ and $q = 0$, $\overline{s} \leq n_1^C \Leftrightarrow n \leq n_1^C$ by (19); if $n > n_1^C$, then $r = 0, q = 0$ is not feasible.

(ii) *Allowing for repositioning capacity $r_{12} \geq 0$, the maximum service capacity achievable (assuming n is sufficiently large) is*

$$\begin{cases} n_1^C + r_{12}\frac{t_{21}}{t_{12}}, \ \textit{with } r_{21} = 0 & \textit{if } r_{12} \in [0, n_2^C - \overline{S}] \\ \overline{S}, \ \textit{with } r_{21} = (S_{12} + r_{12})\frac{t_{21}}{t_{12}} - S_{21} > 0 & \textit{if } r_{12} > n_2^C - \overline{S} \end{cases}.$$

To see this, from (18) we have $s_{21} = (s_{12} + r_{12})\frac{t_{21}}{t_{12}} - r_{21}$, hence for $r_{12} \in [0, n_2^C - \overline{S}]$,

$$\max_{0 \leq s \leq S, r_{21} \geq 0} \overline{s} = \max_{0 \leq s \leq S, r_{21} \geq 0} s_{11} + s_{12} + (s_{12} + r_{12})\frac{t_{21}}{t_{12}} - r_{21} + s_{22} = n_1^C + r_{12}\frac{t_{21}}{t_{12}},$$

where the maximum is achieved at $s_{11} = S_{11}, s_{12} = S_{12}, s_{22} = S_{22}, r_{21} = 0$. Note that the service capacity reaches its upper bound $\overline{S}$ when r_{12} reaches $n_2^C - \overline{S}$, i.e., $n_1^C + (n_2^C - \overline{S})\frac{t_{21}}{t_{12}} = \overline{S}$. For $r_{12} > n_2^C - \overline{S}$, the maximum service capacity *stays* at $\overline{S}$ with $s = S$, but $r_{21} = (S_{12} + r_{12})\frac{t_{21}}{t_{12}} - S_{21} > 0$ by (18).

With these observations, we can derive the optimal structure given by the Proposition. Zone (1) follows directly from observation (i). In zone (2) and (3) where $n > n_1^C$, by observation (ii), the optimization problem with least repositioning travel cost (i.e., avoiding unnecessary repositioning capacity) can be simplified as

$$\max_{r_{12}} \left\{ n_1^C + r_{12}\frac{t_{21}}{t_{12}} : \ n_1^C + r_{12}\frac{t_{21}}{t_{12}} + r_{12} \leq n, \ r_{12} \in [0, n_2^C - \overline{S}], \ r_{21} = 0 \right\}.$$

When $n_1^C < n \le n_2^C$ (zone (2)), the inequality constraint is binding and the optimal solution is

$$r_{12} = (n - n_1^C)\frac{t_{12}}{t_{12}+t_{21}},\ r_{21} = 0, \quad s = \underline{s} + \left(0, 0, (n - n_1^C)\frac{t_{21}}{t_{12}+t_{21}}, 0\right), \quad q = 0.$$

For $n > n_2^C$ (zone (3)), the inequality constraint is not binding. With all the demand served ($s = S$), the extra capacity waits in queues and the optimal solution is

$$r_{12} = n_2^C - \overline{S},\ r_{21} = 0, \quad s = S, \quad q \in \left\{(q_1, q_2) : q_1 + q_2 = n - n_2^C\right\}.$$

□

Proof of Lemma 1. Drivers' expected steady-state profit rate under policy $\widetilde{\pi}(\eta; s, q)$ in (25) follows from Renewal Reward Theory. We next derive its specific expression for $\eta = (\eta_1, 0)$. W.L.O.G., we calculate the time functions over cycles starting and ending at the low-demand location (1). Let $p_{lk} = \lambda_{lk}/(\lambda_{l1} + \lambda_{l2})$ denote the probability of serving a lk-ride at location l. The expected service, repositioning and queueing time functions are as follows:

- The expected service time in a cycle is given by

$$T^s(\eta; s) = (1-\eta_1)\left[p_{11}t_{11} + p_{12}\left(t_{12} + \frac{1-p_{21}}{p_{21}}t_{22} + t_{21}\right)\right] + \eta_1\left(\frac{1-p_{21}}{p_{21}}t_{22} + t_{21}\right),$$

  where $\frac{1-p_{21}}{p_{21}}t_{22}$ gives the expected time serving local demand at location 2. This follows from the fact that the number of local rides at location 2 a driver serves ("failures") before picking a ride back to location 1 ("success") follows a geometric distribution of "success" probability p_{21}.

- The expected repositioning time in a cycle is simply $T^r(\eta) = \eta_1 t_{12}$.

- The expected queueing delay in a cycle is given by

$$T^q(\eta; s, q) = (1-\eta_1)\left(W_1 + p_{12}\frac{1}{p_{21}}W_2\right) + \eta_1\frac{1}{p_{21}}W_2,$$

  where $\frac{1}{p_{21}}W_2$ gives the expected queueing time at location 2. This follows from the fact that the number of queueing delays ("trials") at location 2 a driver encounters before picking a ride back to location 1 ("success") follows a geometric distribution of "success" probability p_{21}. $W_l = Q_l/(\lambda_{l1} + \lambda_{l2})$ is the queueing time at location l due to Little's Law.

Substituting the above time functions in (25) we have:

$$\widetilde{\pi}(\eta_1; s, q) = \frac{(\bar{\gamma}p - c)\left[(1-\eta_1)s_{21}\frac{t_{12}}{t_{21}}(s_{11}+s_{12}) + (\eta_1 s_{11}\frac{t_{12}}{t_{11}} + s_{12})(s_{21}+s_{22})\right] - c\eta_1 s_{21}\frac{t_{12}}{t_{21}}(s_{11}\frac{t_{12}}{t_{11}} + s_{12})}{(1-\eta_1)s_{21}\frac{t_{12}}{t_{21}}(s_{11}+s_{12}+q_1) + (\eta_1 s_{11}\frac{t_{12}}{t_{11}} + s_{12})(s_{21}+s_{22}+q_2) + \eta_1 s_{21}\frac{t_{12}}{t_{21}}(s_{11}\frac{t_{12}}{t_{11}} + s_{12})}. \tag{50}$$

□

Proof of Proposition 2. With $\eta_2 = 0$, and hence $r_{21} = 0$ by (26), the repositioning decision reduces to picking η_1, and the driver-incentive compatible capacity allocation set is given by

$$\mathcal{D} = \left\{(s,r,q) \geq 0 : r_{21} = 0, \eta_1(s,r) \in \arg\max_{\eta_1}\ \widetilde{\pi}(\eta_1; s, q)\right\}. \tag{51}$$

Differentiating $\widetilde{\pi}(\eta_1; s, q)$ in (50) wrt η_1, we get

$$\frac{\partial \widetilde{\pi}}{\partial \eta_1} = \frac{s_{21}\frac{t_{12}}{t_{21}}(s_{11}\frac{t_{12}}{t_{11}} + s_{12})\left[(s_{21}+s_{22})\bar{\gamma}p - \left(s_{21}+s_{22}+s_{21}\frac{t_{12}}{t_{21}}\right)c\right]}{\left[(1-\eta_1)s_{21}\frac{t_{12}}{t_{21}}(s_{11}+s_{12}+q_1) + (\eta_1 s_{11}\frac{t_{12}}{t_{11}} + s_{12})(s_{21}+s_{22}+q_2) + \eta_1 s_{21}\frac{t_{12}}{t_{21}}(s_{11}\frac{t_{12}}{t_{11}} + s_{12})\right]^2}$$
$$\times\left[q_1 - (q_1^*(s) + k(s)q_2)\right],$$

where

$$q_1^*(s) = \frac{(s_{11}+s_{12})s_{21}\frac{t_{12}}{t_{21}} + (s_{21}+s_{22})s_{12}}{(s_{21}+s_{22}) - \left(s_{21}+s_{22}+s_{21}\frac{t_{12}}{t_{21}}\right)\frac{c}{\bar{\gamma}p}}, \quad k(s) = \frac{(s_{11}+s_{12}) - s_{11}\frac{c}{\bar{\gamma}p}}{(s_{21}+s_{22}) - \left(s_{21}+s_{22}+s_{21}\frac{t_{12}}{t_{21}}\right)\frac{c}{\bar{\gamma}p}}. \tag{52}$$

The sign of $\partial\widetilde{\pi}/\partial\eta_1$ only depends on the sign of $q_1 - (q_1^*(s) + k(s)q_2)$. Note that $\eta_1(s,r) = r_{12}/(s_{11}\frac{t_{12}}{t_{11}} + s_{12} + r_{12})$ from (26), by (51):

(i) When $r_{12} = 0$, $\eta_1(s,r) = 0$ requires $\partial\widetilde{\pi}/\partial\eta_1 \leq 0$, hence $q_1 \leq q_1^*(s) + k(s)q_2$;

(ii) When $r_{12} > 0$ and $s_{11}+s_{12} > 0$, $\eta_1(s,r) \in (0,1)$ requires $\partial\widetilde{\pi}/\partial\eta_1 = 0$, hence $q_1 = q_1^*(s)+k(s)q_2$ with $q_1^*(s), k(s) > 0$;

(iii) When $r_{12} > 0$ and $s_{11} + s_{12} = 0$, $\eta_1(s,r) = 1$ and $q_1^*(s) = k(s) = 0$, all drivers reposition at location 1 without waiting in a queue.

It follows that the driver-incentive compatible capacity allocation set given by (28). □

Proof of Proposition 3. We want to show there is a unique feasible capacity utilization satisfying (18)–(20), (29)–(31) and (28), i.e., $(s,r,q,n) \in \mathcal{M}$ given by (32). Note that $r_{21} = 0$ by (28). By

(18) and (29) we can express service capacities in terms of s_{12} and r_{12}:

$$s_{11} = s_{12}\frac{S_{11}}{S_{12}}, \quad s_{21} = (s_{12}+r_{12})\frac{t_{21}}{t_{12}}, \quad s_{22} = (s_{12}+r_{12})\frac{t_{21}}{t_{12}}\frac{S_{22}}{S_{21}}. \tag{53}$$

We will focus on s_{12}, r_{12}, q, and recover the remaining quantities using (53). The other constraints, (19), (28), (30), (31) and (20), are rewritten below.

$$\frac{n_1^M}{S_{12}}s_{12} + \left[\frac{t_{21}}{t_{12}}\left(1+\frac{S_{22}}{S_{21}}\right)+1\right]r_{12} + q_1 + q_2 = n, \tag{54}$$

$$q_1 \begin{cases} \leq q_1^*(s) + k(s)q_2 & \text{if } r_{12} = 0 \\ = q_1^*(s) + k(s)q_2 & \text{if } r_{12} > 0 \end{cases}, \tag{55}$$

$$(S_{12} - s_{12})r_{12} = 0, \tag{56}$$

$$(S_{12} - s_{12})q_1 = 0, \quad \left(S_{21} - (s_{12}+r_{12})\frac{t_{21}}{t_{12}}\right)q_2 = 0, \tag{57}$$

$$0 \leq s_{12} \leq S_{12}, \quad 0 \leq \frac{t_{21}}{t_{12}}(s_{12}+r_{12}) \leq S_{21}, \quad r_{12} \geq 0, \quad q \geq 0. \tag{58}$$

We make the following observations about the feasible solution.

(i) *The allocation where all capacity serves rider demand, i.e.,* $r_{12} = 0, q = 0$, $s_{12} = \frac{n}{n_1^M}S_{12}$, *is feasible if and only if* $n \leq n_1^M$. (a) "$\Rightarrow$": this is immediate; this also implies that if $n > n_1^M$, then $r_{12} > 0$ or $q > 0$. (b) "$\Leftarrow$": given $n \leq n_1^M$, suppose first that $r_{12} > 0$, then (56) implies that $s_{12} = S_{12}$, thus $\bar{s} > n_1^M$ and $n > n_1^M$, a contradiction; second, that $q_1 > 0$, then (57) implies that $s_{12} = S_{12}$, thus $\bar{s} \geq n_1^M$ and $n > n_1^M$, a contradiction; or third, that $q_2 > 0$, then (57) implies that $s_{21} = S_{21}, r_{12} > 0$, thus $n > n_1^M$ is still a contradiction. Hence $n \leq n_1^M \Rightarrow r_{12} = 0, q = 0, s_{12} = \frac{n}{n_1^M}S_{12}$, which satisfies (54)–(58) and is thus feasible.

(ii) *When* $n > n_1^M$, *any feasible solution must serve all demand at location 1, i.e.,* $s_{12} = S_{12}$, *and moreover,* $q_1^*(s) \equiv q_1^*(S)$. By (i), $n > n_1^M$ implies that $r_{12} > 0$ and/or $q \neq 0$, and either of these assertions implies that $s_{12} = S_{12}$ by (56) and (57). To show that $q_1^*(s) \equiv q_1^*(S)$, we substitute s_{11}, s_{21}, s_{22} in $q_1^*(s)$ in (52) by (53),

$$q_1^*(s) = \frac{(\frac{S_{11}}{S_{12}}+1)\frac{t_{12}}{t_{21}} + (1+\frac{S_{22}}{S_{21}})}{(1+\frac{S_{22}}{S_{21}}) - (1+\frac{S_{22}}{S_{21}}+\frac{t_{12}}{t_{21}})\frac{c}{\bar{\gamma}p}}s_{12},$$

which is equal to a constant multiplying s_{12}. For $s_{12} = S_{12}$, we have that $q_1^*(s) \equiv q_1^*(S)$.

With these two observations, we derive the feasible solution given by the Proposition.

Zone (1): the result follows directly from observation (i).

Zone (2): for $n_1^M < n \le n_2^M := n_1^M + q_1^*(S)$, observation (ii) gives $s_{12} = S_{12}$ and $q_1^*(s) = q_1^*(S)$, so (54), (55) and (58) immediately imply that $r_{12} = 0$. Putting $s_{12} = S_{12}$ and $r_{12} = 0$ into (57) we get $q_2 = 0$. It also follows from (54) that $q_1 = n - n_1^M$. In this zone q_1 increases with n while $r_{12} = q_2 = 0$. The feasible solution is summarized as

$$r = 0, \quad s = \left(S_{11}, S_{12}, S_{21}\frac{\Lambda_{12}}{\Lambda_{21}}, S_{22}\frac{\Lambda_{12}}{\Lambda_{21}}\right), \quad q = (n - n_1^M, 0).$$

Zone (3): for $n_2^M < n \le n_3^M := n_2^C + q_1^*(S)$, if $s_{12} = S_{12}$ and $r_{12} = 0$, then $q_2 = 0$ by (57), which implies that $q_1 \le q_1^*(S)$ by (55). By (54), this contradicts the fact that $n > n_2^M$. It follows that $r_{12} > 0$, and (55) yields $q_1 = q_1^*(S) + k(s)q_2$. Together with $s_{12} = S_{12}$, it is then easy to verify that (54), (57) and $n \le n_3^M$ imply that $q_2 = 0$. In this zone r_{12} increases with n while $q_1 = q_1^*(S), q_2 = 0$ and the feasible solution is as follows:

$$r_{12} > 0, \quad s = \left(S_{11}, S_{12}, (S_{12} + r_{12})\frac{t_{21}}{t_{12}}, (S_{12} + r_{12})\frac{t_{21}}{t_{12}}\frac{S_{22}}{S_{21}}\right), \quad q = (q_1^*(S), 0).$$

Zone (4): for $n > n_3^M$, the above argument still implies that $r_{12} > 0$ and $q_1 = q_1^*(S) + k(s)q_2$. And, by (54) and (58) we get that $q_2 > 0$. It then follows from (57) that $r_{12} = S_{21}\frac{t_{12}}{t_{21}} - S_{12} = n_2^C - \overline{S}$ and $s = S$. In this zone q_1 and q_2 increase with n while s and r stay constant. The feasible solution is given by

$$r = (n_2^C - \overline{S}, 0), \quad s = S, \quad q = (q_1^*(S) + k(S)q_2, q_2).$$

This completes the proof. □

Proof of Corollary 2. Substituting $\bar{s}$ and $\bar{r}$ from Proposition 3 into (21) yields

$$\pi_M(n) = \frac{(\bar{\gamma}p - c)\bar{s} - c\bar{r}}{n} = \begin{cases} \bar{\gamma}p - c & \text{zone (1) } (n \le n_1^M), \\ \frac{n_1^M}{n}(\bar{\gamma}p - c) & \text{zone (2) } (n_1^M < n \le n_2^M), \\ \frac{S_{21}+S_{22}}{S_{21}+S_{22}+S_{21}\frac{t_{12}}{t_{21}}}\bar{\gamma}p - c & \text{zone (3) } (n_2^M < n \le n_3^M), \\ \frac{1}{n}(\bar{\gamma}p\overline{S} - cn_2^C) & \text{zone (4) } (n > n_3^M). \end{cases} \tag{59}$$

It is easy to see that $\pi_M(n)$ is continuously decreasing in n and that $\lim_{n\to\infty} \pi_M(n) = 0$. Therefore, the participation equilibrium condition (22), $n = NF(\pi_M(n))$, has a unique solution n_M^*. □

Proof of Proposition 4. We have the following observations about the optimal solution.

(i) *Allocating all capacity n towards serving riders (i.e., $r_{12} = 0, q = 0$) is feasible (hence optimal) if and only if $n \leq n_1^A := n_1^C$.* This is the same as observation (i) in the Proof of Proposition 1. Also notice that constraint (28) is satisfied.

(ii) *If for some capacity level n_1 the service capacity $\overline{s} > n_1^A$, then for all capacity levels $n_2 \geq n_1$ the optimal solution involves repositioning.* First, note that by the definition of n_1^A, $\overline{s} > n_1^A$, which implies that $r_{12} > 0$ (which holds at n_1), and hence we only need to show that the optimal service capacity at n_2 is higher than n_1^A. It suffices to find one feasible solution at n_2 that has the same service capacity as at n_1, which is higher than n_1^A. To achieve this, let the service capacity vector s and the repositioning capacity $r_{12} > 0$ at n_2 be the same as those at n_1, respectively, and put the extra capacity $n_2 - n_1$ into q satisfying $q_1 = q_1^*(s) + k(s)q_2$. In this way all constraints are still satisfied while the service capacity $\overline{s} > n_1^A$ remains unchanged.

(iii) *$r_{12} > 0$ for all $n > n_1^A + q_1^*(S_1^A)$, where $S_1^A = S_1^C$ defined in (49) such that $S_1^A \cdot 1 = n_1^A$.* It suffices to show that at capacity levels in the right neighborhood of $n_1^A + q_1^*(S_1^A)$, the optimal service capacity is higher than n_1^A, which then, by observation (ii), will prove the result. To show this, for an arbitrarily small $\epsilon > 0$, let n_ϵ be the minimum feasible total capacity to provide service vector $S_1^A + (0, 0, \epsilon, 0)$, and hence service capacity $n_1^A + \epsilon > n_1^A$. Following constraints (18)–(20) and (28), we have

$$n_\epsilon = n_1^A + \epsilon + \frac{t_{12}}{t_{21}}\epsilon + q_1^*(S_1^A + (0, 0, \epsilon, 0)),$$

and $n_0 = n_1^A + q_1^*(S_1^A)$ for $\epsilon = 0$. It is easy to see that n_ϵ increases in ϵ, since by definition (52),

$$\frac{\partial q_1^*(s)}{\partial s_{21}} = \frac{((s_{11} + s_{12})\bar{\gamma}p - s_{11}c)s_{22}\frac{t_{12}}{t_{21}}}{\left[(s_{21} + s_{22})\bar{\gamma}p - (s_{21} + s_{22} + s_{21}\frac{t_{12}}{t_{21}})c\right]^2} > 0, \quad \forall s_{22} > 0, s_{11} + s_{12} > 0, \tag{60}$$

i.e., $q_1^*(s)$ increases wrt s_{21} when $s_{22}, s_{11} + s_{12} > 0$. Therefore, the optimal service capacity must be higher than n_1^A at capacity levels in the right neighborhood of $n_1^A + q_1^*(S_1^A)$.

(iv) *$r_{12} = 0$ for $n \in [n_1^A, n_1^A + \delta]$ for a small $\delta > 0$.* We first prove this for the optimal solution at $n = n_1^A + \delta$ by establishing that for any feasible q_1 it must be that $q_1 < q_1^*(s)$, from which we deduce $r_{12} = 0$ from (28). Then, observation (ii) yields the same result for $n \in [n_1^A, n_1^A + \delta]$. Pick

$$\delta = \min\left\{q_1^*(S_1^A), \frac{(S_{12})^2}{S_{12}\left(1 + \frac{t_{21}}{t_{12}}\right) + S_{21} + S_{22}}\right\},$$

so that $n_1^A + \delta \leq n_1^A + q_1^*(S_1^A)$, $\delta < S_{12}$ and

$$\delta < \frac{(S_{12})^2}{S_{12}\left(1+\frac{t_{21}}{t_{12}}\right) + S_{21} + S_{22}} \left(1 + \frac{t_{21}}{t_{12}}\right) \Rightarrow \frac{S_{12}(S_{12}-\delta)}{S_{21}+S_{22}} \left(1+\frac{t_{21}}{t_{12}}\right) > \delta. \tag{61}$$

First note that the optimal solution at $n_1^A + \delta$ must have $\overline{s} \geq n_1^A$, since a feasible solution $s = S_1^A, r_{12} = 0, q_1 = 0, q_2 = \delta$ yields $\overline{s} = n_1^A$. Therefore

$$r_{12}, q_1 \leq \delta, \quad s_{21} \geq S_{12}\frac{t_{21}}{t_{12}}, \quad s_{12} \geq S_{12} - \delta > 0, \tag{62}$$

where the first inequality follows from $\overline{s} \geq n_1^A$ and capacity constraint (19), the second is by $\overline{s} \geq n_1^A$, and the third follows from the second and (18) in that $s_{12} = s_{21}\frac{t_{12}}{t_{21}} - r_{12} \geq S_{12} - \delta$. Then,

$$q_1^*(s) = \frac{(s_{11}+s_{12})s_{21}\frac{t_{12}}{t_{21}} + (s_{21}+s_{22})s_{12}}{(s_{21}+s_{22}) - \left(s_{21}+s_{22}+s_{21}\frac{t_{12}}{t_{21}}\right)\frac{c}{\gamma p}} \geq \frac{s_{12}s_{21}\frac{t_{12}}{t_{21}} + s_{21}s_{12}}{S_{21}+S_{22}} \geq \frac{S_{12}(S_{12}-\delta)\frac{t_{21}}{t_{12}}\left(1+\frac{t_{12}}{t_{21}}\right)}{S_{21}+S_{22}} > \delta,$$

where the the second inequality follows from the second and third inequalities in (62), and the last inequality is by (61). Finally the first inequality in (62) leads to $q_1 \leq \delta < q_1^*(s)$, which implies that $r_{12} = 0$ by constraint (28). Hence we have shown $r_{12} = 0$ at $n = n_1^A + \delta$. By observation (ii), the optimal solution at $n \in [n_1^A, n_1^A + \delta]$ has service capacity $\overline{s} = n_1^A$ and no repositioning.

(v) *An optimal solution can serve all demand ($s = S$) if and only if $n \geq n_3^A := n_2^C + q_1^*(S)$.* "$\Leftarrow$": given $n \geq n_2^C + q_1^*(S)$, it is easy to verify that the capacity allocation $s = S, r = (n_2^C - \overline{S}, 0)$ and $q_1 = q_1^*(S) + k(S)q_2$ with $q_1 + q_2 = n - n_2^C$ is feasible and serves all demand (hence optimal). "$\Rightarrow$": an optimal (hence feasible) solution that serves all demand must have $s = S, r = (n_2^C - \overline{S}, 0)$ and $q_1 = q_1^*(S) + k(S)q_2$. By (19) this yields $n = n_2^C + q_1^*(S) + k(S)q_2 + q_2 \geq n_2^C + q_1^*(S) = n_3^A$.

With these five observations, we can derive the optimal solution given by the Proposition. Zone (1) follows directly from observation (i) and zone (4) follows from observation (v). In $(n_1^A, n_3^A]$, not all drivers are serving riders and not all riders are served. There exists a threshold n_2^A such that $n_1^A < n_2^A < n_3^A$ which separates zone (2) and (3) apart: in zone (2), $(n_1^A, n_2^A]$, optimal solution has service capacity $\overline{s} = n_1^A$, no repositioning ($r = 0$), and extra capacity queues at location 1 with $q = (n - n_1^A, 0)$; whereas in zone (3), $(n_2^A, n_3^A]$, optimal solution involves repositioning

($r_{12} > 0$), serves $\overline{s} > n_1^A$, and extra capacity queues at location 1 with $q = (q_1^*(s), 0)$.[9] Note that $n_2^A > n_1^A$ by observation (iv). $n_2^A < n_3^A$ follows from $n_2^A \le n_1^A + q_1^*(S_1^A)$ by observation (iii) and $n_1^A + q_1^*(S_1^A) < n_2^C + q_1^*(S) = n_3^A$ by property (60). Furthermore, the fact that optimal solution involves repositioning at any capacity level in zone (3) follows from observation (ii). □

The following proofs of Corollary 3 and Propositions 6–7 refer to two technical lemmas, Lemmas 2 and 3. The statements and proofs of these lemmas as well as the proof of Proposition 5 are relegated to the Supplemental Materials.

Proof of Corollary 3. Substituting $\overline{s}$ and $\overline{r}$ from Proposition 4 into (21) yields

$$\pi_A(n) = \frac{(\bar{\gamma}p - c)\overline{s} - c\overline{r}}{n} = \begin{cases} \bar{\gamma}p - c & \text{zone (1) } (n \le n_1^A), \\ \frac{n_1^A}{n}(\bar{\gamma}p - c) & \text{zone (2) } (n_1^A < n \le n_2^A), \\ \frac{1}{n}[(\bar{\gamma}p - c)\overline{s}^* - c\overline{r}^*] & \text{zone (3) } (n_2^A < n \le n_3^A), \\ \frac{1}{n}(\bar{\gamma}p\overline{S} - cn_2^C) & \text{zone (4) } (n > n_3^A). \end{cases} \tag{63}$$

Note that there is no (simple) explicit expression in zone (3), in which the platform may reject riders at the low-demand location if condition (37) holds. Nevertheless, we show in Lemma 3 (see Supplemental Materials) that $\pi_A(n)$ is decreasing in n and $\lim_{n\to\infty} \pi_A(n) = 0$. Hence there is a unique equilibrium participation capacity n_A^* to inequalities

$$NF(\pi_A(n_A^{*+})) \le n_A^* \le NF(\pi_A(n_A^{*-})). \tag{64}$$

Moreover, when (37) is not satisfied, there is no demand rejection at the low-demand location. It follows from pattern (1) in Lemma 2 (see Supplemental Materials) that only s_{21} is increasing in zone (3) and therefore $\pi_A(n)$ is continuous. In this case (64) simplifies to $NF(\pi_A(n_A^*)) = n_A^*$. □

Proof of Proposition 6. (1) Given a level of participating capacity n, the platform revenue rate

[9] The steady state system flow equations do not differentiate between queueing in locations 1 and 2 in this capacity regime. A more detailed transient analysis would show that when the platform makes admission control decisions, it would choose to clear the queue in the high-demand location given that the demand exceeds the available capacity, and drivers would only queue in the low-demand location.

under the three control regimes $\{M, A, C\}$ is computed as follows:

$$\begin{aligned}
\mathit{\Pi}_M(n) &= \underset{s,r,q}{\operatorname{argmax}}\{\mathit{\Pi}(s) : (18) - (20), (29) - (31), (28)\}, \\
\mathit{\Pi}_A(n) &= \underset{s,r,q}{\operatorname{argmax}}\{\mathit{\Pi}(s) : (18) - (20), (28)\}, \\
\mathit{\Pi}_C(n) &= \underset{s,r,q}{\operatorname{argmax}}\{\mathit{\Pi}(s) : (18) - (20)\}.
\end{aligned}$$

These formulations share the same objective function, but have a decreasing set of constraints from $M \to A \to C$, from where (38) holds.

(2) and (3): Based on (24), (59) and (63), it is straightforward to verify that $\pi_M(n) \leq \pi_C(n)$ for any n, and to verify that $\pi_M(n) \leq \pi_A(n) \leq \pi_C(n)$ for $n \leq n_2^A$ and $n > n_3^A$ (i.e., zone (1), (2) and (4) in Proposition 4 for regime A). For $n_2^A < n \leq n_3^A$ (zone (3) of regime A), using Lemma 2 one can verify that $\pi_A(n) \leq \pi_C(n)$ always holds, and $\pi_M(n) \leq \pi_A(n)$ holds if (37) is not satisfied. This proves (39) and (40). $\square$

Proof of Proposition 7. Note the following properties about $\mathit{\Pi}_X(\cdot), \pi_X(\cdot)$ and n_X^*:

(i) Proposition 6 (1), $\mathit{\Pi}_M(n) \leq \mathit{\Pi}_A(n) \leq \mathit{\Pi}_C(n), \forall n$. It is also easy to verify that $\mathit{\Pi}_X(n)$ is continuously increasing in n for $X \in \{M, A, C\}$. This property follows immediately for the centralized control regime $X = C$. For regimes $X \in \{M, A\}$, one can verify that increasing capacity can be allocated into IC queues as in (28) without any reduction in the capacity that serves rider demand.

(ii) $\pi_X(n)$ is decreasing in n with $\lim_{n\to\infty} \pi_X(n) = 0$ for $X \in \{M, A, C\}$. Moreover, $\pi_X(\cdot)$ is continuous for $X \in \{M, C\}$, and is continuous for $X = A$ if (37) is not satisfied. This follows from (24), (59), (63) and Lemma 3.

(iii) Let $\pi_1(n), \pi_2(n) : \mathbb{R}_+ \to \mathbb{R}_+$ be candidate per-driver profit functions that are decreasing in n (but may be *discontinuous*) with $\lim_{n\to\infty} \pi_i(n) = 0$, $i = 1, 2$. Let n_i^* be defined as $NF(\pi_i(n_i^{*+})) \leq n_i^* \leq NF(\pi_i(n_i^{*-}))$ and $\pi_i(n_i^*)$ satisfies $NF(\pi_i(n_i^*)) = n_i^*$ if n_i^* is a *discontinuity* of $\pi_i(\cdot)$. If $\pi_1(n) \leq \pi_2(n), \forall n$, then $n_1^* \leq n_2^*$ and $\pi_1(n_1^*) \leq \pi_2(n_2^*)$. To see $n_1^* \leq n_2^*$, by definition of n_1^* and $\pi_1(n) \leq \pi_2(n), \forall n$, we have

$$n_1^* \leq NF(\pi_1(n_1^{*-})) \leq NF(\pi_2(n_1^{*-})). \tag{65}$$

Suppose $n_2^* < n_1^*$, then $n_2^{*+} < n_1^{*-}$, by definition of n_2^* and (65) we have

$$n_2^* \geq NF(\pi_2(n_2^{*+})) \geq NF(\pi_2(n_1^{*-})) \geq n_1^*,$$

contradicting $n_2^* < n_1^*$. Hence, it must be that $n_1^* \leq n_2^*$. It then follows from $NF(\pi_i(n_i^*)) = n_i^*,\ i = 1, 2$ that $\pi_1(n_1^*) \leq \pi_2(n_2^*)$.

With these three properties and Proposition 6, we are ready to prove the three parts in this Proposition. Parts (2) and (3) follow from Property (ii), (iii) and Proposition 6 (2), (3). The ranking of n_X^* given by (42) and Property (i) imply that $\mathit{\Pi}_M^* \leq \mathit{\Pi}_C^*$ and $\mathit{\Pi}_A^* \leq \mathit{\Pi}_C^*$ in Part (1). The only thing left to show is $\mathit{\Pi}_M^* \leq \mathit{\Pi}_A^*$ and we complete this considering whether (37) holds.

If (37) is not satisfied, then Part (3) implies that $n_M^* \leq n_A^*$ and hence $\mathit{\Pi}_M^* \leq \mathit{\Pi}_A^*$ by Property (i).

If (37) is satisfied, then $\pi_A(\cdot)$ may be discontinuous in zone (3) of regime A. Consider the value of n_M^* and notice (59) and (63): In zone (1) or (4) ($n_M^* \leq n_1^M < n_1^A$ or $n_M^* \geq n_3^M = n_3^A$), $\pi_M(n_M^*) = \pi_A(n_M^*)$ and hence $n_A^* = n_M^*$, $\mathit{\Pi}_A^* = \mathit{\Pi}_M^*$; in zone (2) ($n_1^M < n_M^* \leq n_2^M$), it is obvious that $n_1^M < n_A^*$ since $\pi_A(n) = \pi_M(n)$ for $n \leq n_1^M$, therefore $\mathit{\Pi}_M^* = \mathit{\Pi}_M(n_1^M) < \mathit{\Pi}_A(n_A^*) = \mathit{\Pi}_A^*$; in zone (3) ($n_2^M < n_M^* < n_3^M = n_3^A$), since $\pi_A(n) > \pi_M(n_M^*)$ for $n < n_3^M = n_3^A$, there must be $n_M^* < n_A^*$ and hence $\mathit{\Pi}_M^* < \mathit{\Pi}_A^*$. Therefore $\mathit{\Pi}_M^* \leq \mathit{\Pi}_A^*$ for any n_M^*, and this completes the proof of Part (1). $\square$

Proof of Corollary 4. Using the expressions of $\pi_X(n)$ for $X \in \{M, A, C\}$ in (24), (59), (63), respectively, and Proposition 6 (3), we have

$$\pi_A(n) \begin{cases} > \pi_M(n) & n \in (n_1^M, n_3^M),\ (37) \text{ not satisfied} \\ = \pi_M(n) & n \leq n_1^M \text{ or } n \geq n_3^M \end{cases}, \quad \pi_C(n) \begin{cases} > \pi_A(n) & n \in (n_1^A, n_3^A) \\ = \pi_A(n) & \text{o.w.} \end{cases}. \tag{66}$$

(1) For performance gains from admission control (control regime A over M), if $n_M^* \leq n_1^M$ or $n_M^* \geq n_3^M$, clearly $n_A^* = n_M^*$ since $\pi_A(n) = \pi_M(n)$ in these ranges; hence $\mathit{\Pi}_M^* = \mathit{\Pi}_A^*, \pi_M^* = \pi_A^*$ and there is no gain. If $n_1^M < n_M^* < n_3^M$ and (37) is not satisfied, then it must be $n_M^* < n_A^*$. To see this, (66) implies that $\pi_A(n_M^*) > \pi_M(n_M^*)$, therefore $n_M^* = NF(\pi_M(n_M^*)) < NF(\pi_A(n_M^*))$. Since (37) is not satisfied, $n_A^* = NF(\pi_A(n_A^*))$, and from the monotonicity of $\pi_A(\cdot)$ we deduce that $n_M^* < n_A^*$. As a result, $\mathit{\Pi}_M^* < \mathit{\Pi}_A^*$ and $\pi_M^* < \pi_A^*$ following the proof of Proposition 7.

(2) For performance gains from centralized repositioning (control regime C over A) we have a similar argument as the above proof of Part (1) and the details are omitted. $\square$

Proof of Proposition 8. (1) According to Corollary 4, if (37) is not satisfied, the platform revenue rate gain from admission control is positive only when $n_M^* \in (n_1^M, n_3^M)$. This yields

$$\Pi_M^* \geq \Pi_M(n_1^M) = \gamma p n_1^M, \tag{67}$$

where the equality holds for $n_M^* \in (n_1^M, n_2^M]$. By Corollary 4, $n_M^* \in (n_1^M, n_3^M)$ also implies that $\pi_M^* < \pi_A^*$, thus $n_A^* \in (n_1^M, n_3^M)$ since otherwise $n_M^* = n_A^*$ and $\pi_M^* = \pi_A^*$. This yields

$$\Pi_A^* \leq \Pi_A(n_3^M) = \Pi_A(n_3^A) = \gamma p \overline{S}, \tag{68}$$

where the equality is approached by $n_A^* \to n_3^M = n_3^A$.

Given that $N \geq n_3^M = n_3^A$, $n_X^* = NF(\pi_X(n_X^*))$ can take on values in $[0, n_3^M]$ depending on the choice of $F(\cdot)$, for $X \in \{M, A, C\}$. Consequently, the bounds in (67) and (68) can be approached and therefore

$$\max_{F(\cdot)} \frac{\Pi_A^* - \Pi_M^*}{\Pi_M^*} \leq \frac{\gamma p \overline{S} - \gamma p n_1^M}{\gamma p n_1^M} = \left(\frac{\Lambda_{21}}{\Lambda_{12}} - 1 \right) \frac{1}{1 + \frac{1-\rho_2}{1-\rho_1} \frac{1}{\tau}}.$$

To approach this upper bound, we need $n_M^* \in (n_1^M, n_2^M]$ and $n_A^* \to n_3^M = n_3^A$ so that $\Pi_M^* = \gamma p n_1^M$ and $\Pi_A^* \to \gamma p \overline{S}$. (Refer to Figure 7 (a) and (c) for an illustration.) This holds for opportunity cost distributions $F(\cdot)$ satisfying

$$F^{-1}(n_2^M/N) = \pi_M(n_2^M) \quad \text{and} \quad F^{-1}(n_3^M/N) = \pi_M(n_2^M)^+,$$

i.e., the value of F at $\pi_M(n_2^M)$ is fixed at n_2^M/N ($\Rightarrow n_M^* = n_2^M$) and F grows sufficiently fast to $n_3^M/N = n_3^A/N$ at $\pi_M(n_2^M)^+ = \pi_A(n_3^A)^+$ ($\Rightarrow n_A^* \to n_3^A = n_3^M$); in words, there is a sufficiently large mass of potential drivers with opportunity cost around $\pi_M(n_2^M)^+$.

(2) According to Corollary 4, the platform revenue rate gain from centralized repositioning is positive only when $n_A^* \in (n_1^A, n_3^A)$. This yields

$$\Pi_A^* \geq \Pi_A(n_1^A) = \gamma p n_1^A, \tag{69}$$

where the equality holds for $n_A^* \in (n_1^A, n_2^A]$. By Corollary 4, $n_A^* \in (n_1^A, n_3^A)$ also implies that $\pi_A^* < \pi_C^*$. Thus, $n_C^* \in (n_1^A, n_3^A)$, since otherwise $n_A^* = n_C^*$ and $\pi_A^* = \pi_C^*$. This yields

$$\Pi_C^* \leq \Pi_C(n_3^A) = \gamma p \overline{S}, \tag{70}$$

where the equality holds for $n_C^* \in [n_2^C, n_3^A)$.

Given that $N \geq n_3^M = n_3^A$, $n_X^* = NF(\pi_X(n_X^*))$ can take on values in $[0, n_3^A]$ depending on the choice of $F(\cdot)$, for $X \in \{M, A, C\}$. Using the bounds in (69) and (70) we have

$$\max_{F(\cdot)} \frac{\Pi_C^* - \Pi_A^*}{\Pi_A^*} \leq \frac{\gamma p \overline{S} - \gamma p n_1^A}{\gamma p n_1^A} = \left(\frac{\Lambda_{21}}{\Lambda_{12}} - 1\right) \frac{1}{1 + \frac{1}{1-\rho_1}\frac{1}{\tau} + \frac{\rho_2}{1-\rho_2}\frac{\Lambda_{21}}{\Lambda_{12}}}.$$

To achieve this upper bound, we need $n_A^* \in (n_1^A, n_2^A]$ and $n_C^* \in [n_2^C, n_3^A)$ so that $\Pi_A^* = \gamma p n_1^A$ and $\Pi_C^* = \gamma p \overline{S}$. Noticing Proposition 6 (2) and Property (ii) in the proof of Proposition 7 about $\pi_A(\cdot), \pi_C(\cdot)$, this holds for opportunity cost distributions $F(\cdot)$ satisfying

$$F(\pi_A(n_2^A)) \leq n_2^A/N \quad \text{and} \quad F(\pi_C(n_2^C)) \geq n_2^C/N \tag{71}$$

when $\pi_A(n_2^A) < \pi_C(n_2^C)$. When $\pi_A(n_2^A) \geq \pi_C(n_2^C)$, (71) cannot be satisfied by any F and hence the upper bound is not tight. □

Proof of Proposition 9. Since $\pi_M(n) = \pi_A(n) = \pi_C(n)$ for $n \leq n_1^M$ and $n \geq n_3^M$, the per-driver profit rate gain from admission control only (regime A over M) and from admission control plus centralized repositioning (regime C over M) can be positive only for $n_M^* \in (n_1^M, n_3^M)$, and $n_A^*, n_C^* \in (n_1^M, n_3^M)$ simultaneously. It follows from the (decreasing) monotonicity of $\pi_X(\cdot)$, $X \in \{M, A, C\}$ that

$$\pi_M^* \geq \pi_M(n_3^M), \quad \pi_A^* \leq \pi_A(n_1^M) = \bar{\gamma}p - c, \quad \pi_C^* \leq \pi_C(n_1^M) = \bar{\gamma}p - c. \tag{72}$$

Therefore,

$$\max_{F(\cdot)} \frac{\pi_A^* - \pi_M^*}{\pi_M^*} = \max_{F(\cdot)} \frac{\pi_C^* - \pi_M^*}{\pi_M^*} \leq \frac{\bar{\gamma}p - c}{\pi_M(n_3^M)} - 1 = \frac{1 - \rho_2}{\tau - (1 - \rho_2 + \tau)\kappa}.$$

To achieve this upper bound, we need $n_M^* \in [n_2^M, n_3^M)$ and $n_A^*, n_C^* \in (n_1^M, n_1^A]$ so that the equalities in (72) are satisfied. If $n_2^M \leq n_1^A = n_1^C$, these conditions hold for $F(\cdot)$ satisfying

$$F(\pi_M(n_2^M)) \geq n_2^M/N \quad \text{and} \quad F(\pi_A(n_1^A)) \leq n_1^A/N. \tag{73}$$

If $n_2^M > n_1^A = n_1^C$, then (73) cannot be satisfied by any F and the upper bound is not tight. □

Supplemental Materials

S1 Supplemental Lemmas and Proofs for Control Regime A

Under regime A, the lower capacity threshold n_2^A of zone (3)—moderate capacity *with repositioning*—and the optimal capacity allocation within this zone, do not have explicit expressions (see Proposition 4). Lemmas 2 and 3 fill in the remaining details.

Given a level of participating capacity n, Proposition 4 shows that the optimal capacity allocation in zone (3) has $r_{12} > 0, r_{21} = 0$ and $q = (q_1^*(s), 0)$. Therefore, the constraints of Problem A at a given capacity n, (18)–(20) and (28), simplify to

$$\overline{s} + \left(\frac{t_{12}}{t_{21}} s_{21} - s_{12}\right) + q_1^*(s) = n \tag{74}$$

and $0 \leq s \leq S$, $\frac{t_{12}}{t_{21}} s_{21} > s_{12}$. Note that s determines r_{12} by the second term and q by $q_1^*(s)$.

Lemma 2 shows that there are three possible optimal capacity allocation *patterns* at any level of participating capacity in zone (3). These patterns differ in terms of whether demand is rejected at the low-demand location, and if so, for which route(s).

Lemma 2. *Under control regime A, the optimal capacity allocation of any fixed participating capacity $n \in (n_2^A, n_3^A]$ (moderate capacity zone with repositioning) is determined by s that takes one of the following 3 patterns with the largest service capacity,* $\max_{i\in\{1,2,3\}} \overline{s_i}(n)$*:*

(1) <u>No demand rejection at the low-demand location</u>: only s_{21} is increasing in this zone.

$$s_1(n) = (S_{11}, S_{12}, s_{21}, S_{22}) \text{ subject to } (74), \quad n \in (\overline{s_1}^{-1}(n_1^A), n_3^A].$$

(2) <u>Rejecting cross-traffic demand at the low-demand location</u>: for small n, s_{21} is increasing while $s_{12} = 0$; for large n, $s_{21} = S_{21}$ and s_{12} is increasing.

$$s_2(n) = (S_{11}, s_{12}, s_{21}, S_{22}) \text{ subject to } (S_{21} - s_{21})s_{12} = 0 \text{ and } (74), \quad n \in (\overline{s_2}^{-1}(n_1^A), n_3^A].$$

(3) <u>Rejecting local and cross-traffic demand at the low-demand location</u>: for small n, s_{21} is increasing while $s_{11} = s_{12} = 0$; for medium n, $s_{21} = S_{21}$, s_{11} is increasing and $s_{12} = 0$; for large n, $s_{21} = S_{21}, s_{11} = S_{11}$ and s_{12} is increasing.

$$s_3(n) = (s_{11}, s_{12}, s_{21}, S_{22}) \text{ subject to } (S_{21}-s_{21})s_{11} = (S_{11}-s_{11})s_{12} = 0 \text{ and } (74), \quad n \in (\overline{s_3}^{-1}(n_1^A), n_3^A].$$

Proof. By Proposition 4, the optimal capacity allocation of given participating capacity $n \in (n_2^A, n_3^A]$ has $r_{12} > 0, r_{21} = 0$ and $q = (q_1^*(s), 0)$. Therefore, for fixed n, Problem A reduces to $\max_{s,r,q}\{\mathit{\Pi}(s) : (18) - (20), (28)\}$, and it can be reformulated as maximizing the total service capacity over the service capacity vector s:

$$\max_{s} \quad \overline{s} \tag{75}$$

$$\text{s.t.} \quad g(s) := \overline{s} + \left(\frac{t_{12}}{t_{21}} s_{21} - s_{12}\right) + q_1^*(s) \leq n \tag{76}$$

$$0 \leq s_{ij} \leq S_{ij}, \ \forall i, j. \tag{77}$$

Note that $g(s)$ is the total capacity expressed with respect to s. Relaxing the equality constraint (19) to the inequality constraint (76) does not matter since positive q_2 and $q_1 = q_1^*(s) + k(s)q_2$ are feasible by (28) (but not optimal by Proposition 4). Constraint $r_{12} = \frac{t_{12}}{t_{21}} s_{21} - s_{12} > 0$ is omitted since a violation results in $s_{21} \leq \frac{t_{21}}{t_{12}} S_{12} \Rightarrow \overline{s} \leq n_1^A$, clearly suboptimal in zone (3).

To prove the lemma, we first establish that any optimal solution to problem (75)–(77) for $n \in (n_2^A, n_3^A]$ must satisfy the following four necessary conditions:

(a) All capacity is used within this zone and s_{21} has a lower bound:

$$g(s) = n, \tag{78}$$

$$0 < S_{12} \frac{t_{21}}{t_{12}} \leq s_{21} \leq S_{21}. \tag{79}$$

(b) Rejecting local demand at the high-demand location (s_{22}) is suboptimal:

$$s_{22} = S_{22}. \tag{80}$$

(c) Rejecting cross-traffic demand (s_{12}) is more profitable than rejecting local demand (s_{11}) at the low-demand location:

$$(S_{11} - s_{11}) s_{12} = 0. \tag{81}$$

(d) Neither demand stream at the low-demand location is partially served unless s_{21} is fully served:

$$s_{12}(S_{12} - s_{12})(S_{21} - s_{21}) = 0, \tag{82}$$

$$s_{11}(S_{11} - s_{11})(S_{21} - s_{21}) = 0. \tag{83}$$

We establish conditions (a)–(d) using the KKT conditions for the reformulated maximization problem (75)–(77). Before writing the KKT conditions, we first derive the first and second partial derivatives of $g(s)$ which will be used later. The first partial derivatives are

$$
\begin{aligned}
\frac{\partial g(s)}{\partial s_{11}} &= \frac{s_{21}(1+\frac{t_{12}}{t_{21}})+s_{22}}{(s_{21}+s_{22})\bar{\gamma}p-\left(s_{21}+s_{22}+s_{21}\frac{t_{12}}{t_{21}}\right)c}(\bar{\gamma}p-c) > 1,\\
\frac{\partial g(s)}{\partial s_{12}} &= \frac{s_{21}(1+\frac{t_{12}}{t_{21}})+s_{22}}{(s_{21}+s_{22})\bar{\gamma}p-\left(s_{21}+s_{22}+s_{21}\frac{t_{12}}{t_{21}}\right)c}\bar{\gamma}p > \frac{\partial g(s)}{\partial s_{11}} > 1,\\
\frac{\partial g(s)}{\partial s_{21}} &= 1+\frac{t_{12}}{t_{21}}+\frac{(s_{11}(\bar{\gamma}p-c)+s_{12}\bar{\gamma}p)s_{22}\frac{t_{12}}{t_{21}}}{\left[(s_{21}+s_{22})\bar{\gamma}p-\left(s_{21}+s_{22}+s_{21}\frac{t_{12}}{t_{21}}\right)c\right]^2}\bar{\gamma}p > 1,\\
\frac{\partial g(s)}{\partial s_{22}} &= 1-\frac{(s_{11}(\bar{\gamma}p-c)+s_{12}\bar{\gamma}p)s_{21}\frac{t_{12}}{t_{21}}}{\left[(s_{21}+s_{22})\bar{\gamma}p-\left(s_{21}+s_{22}+s_{21}\frac{t_{12}}{t_{21}}\right)c\right]^2}\bar{\gamma}p < 1.
\end{aligned}
$$

For second partial derivatives, we do not need the ones involving s_{22}. Fixing s_{22} and letting $\overline{g}(s_{11},s_{12},s_{21})=g(s)$, the Hessian of $\overline{g}(s_{11},s_{12},s_{21})$ is given by

$$
\mathbf{H}(\overline{g}) = \frac{s_{22}\frac{t_{12}}{t_{21}}\bar{\gamma}p}{\left[(s_{21}+s_{22})\bar{\gamma}p-\left(s_{21}+s_{22}+s_{21}\frac{t_{12}}{t_{21}}\right)c\right]^2}\begin{bmatrix} 0 & 0 & \bar{\gamma}p-c \\ 0 & 0 & \bar{\gamma}p \\ \bar{\gamma}p-c & \bar{\gamma}p & -\frac{2(s_{11}(\bar{\gamma}p-c)+s_{12}\bar{\gamma}p)\left(\bar{\gamma}p-\left(1+\frac{t_{12}}{t_{21}}\right)c\right)}{(s_{21}+s_{22})\bar{\gamma}p-\left(s_{21}+s_{22}+s_{21}\frac{t_{12}}{t_{21}}\right)c} \end{bmatrix}.
$$

Hence we have

$$
\frac{\partial^2 g(s)}{\partial s_{11}^2}=\frac{\partial^2 g(s)}{\partial s_{12}^2}=\frac{\partial^2 g(s)}{\partial s_{11}\partial s_{12}}=0,\quad \frac{\partial^2 g(s)}{\partial s_{11}\partial s_{21}},\frac{\partial^2 g(s)}{\partial s_{12}\partial s_{21}}>0,\quad \frac{\partial^2 g(s)}{\partial s_{21}^2}\le 0, \tag{84}
$$

where the last inequality is strict when $s_{11}+s_{12}>0$.

Let $\alpha,\overline{\beta}_{ij},\underline{\beta}_{ij}$ be the dual variables associated with the capacity constraint (76), the upper and lower bound constraints (77), respectively. The KKT conditions are

$$\text{(stationarity)} \quad \alpha \frac{\partial g(s)}{\partial s_{ij}} + \overline{\beta}_{ij} - \underline{\beta}_{ij} = 1, \ \forall i, j, \tag{85}$$

$$\text{(complementary slackness)} \quad \alpha(n - g(s)) = \overline{\beta}_{ij}(S_{ij} - s_{ij}) = \underline{\beta}_{ij} s_{ij} = 0, \ \forall i, j, \tag{86}$$

$$\text{(dual feasibility)} \quad \alpha, \overline{\beta}_{ij}, \underline{\beta}_{ij} \geq 0, \ \forall i, j, \tag{87}$$

$$\text{(primal feasibility)} \quad g(s) \leq n, \tag{88}$$

$$\text{(primal feasibility)} \quad 0 \leq s_{ij} \leq S_{ij}, \forall i, j. \tag{89}$$

The complementary slackness constraints (86) and dual feasibility constraints (87) establish the relationship between primal and dual variables: $\overline{\beta}_{ij} = 0$ ($\underline{\beta}_{ij} = 0$) when s_{ij} is not at its upper (lower) bound; s_{ij} must be at its upper (lower) bound when $\overline{\beta}_{ij} > 0$ ($\underline{\beta}_{ij} > 0$); $\alpha = 0$ when $g(s) < n$ and $g(s) = n$ when $\alpha > 0$. Moreover, $\overline{\beta}_{ij} \cdot \underline{\beta}_{ij} = 0$. We omit explicit references to the primal and dual feasibility constraints (87)–(89) in the following proof.

Now we are ready to prove the four conditions in this lemma.

(a) When $s \neq S$, pick any $s_{ij} < S_{ij}$, then $\overline{\beta}_{ij} = 0$ by (86). By (85) this implies that $\alpha \neq 0$ and hence $g(s) = n$ by (86). When $s = S$, $g(s) = n = n_3^A$. These prove (78). For (79), $s_{21} \geq S_{12} \frac{t_{21}}{t_{12}} > 0$ follows directly from $\bar{s} > n_1^A$ in zone (3). By (86), $s_{21} > 0$ also implies that $\underline{\beta}_{21} = 0$.

(b) Using $\underline{\beta}_{21} = 0$ from part (a) and $\frac{\partial g(s)}{\partial s_{21}} > 1$, stationarity constraints (85) imply that $\alpha < 1$. Putting this and $\frac{\partial g(s)}{\partial s_{22}} < 1$ back to (85), we obtain $\overline{\beta}_{22} > 0$. Therefore it follows from (86) that $s_{22} = S_{22}$ and $\underline{\beta}_{22} = 0$.

(c) We prove this by contradiction using (85) and (86). Suppose on the contrary $(S_{11} - s_{11})s_{12} > 0$ for some $s_{11} < S_{11}$ and $s_{12} > 0$, then (86) require $\overline{\beta}_{11} = \underline{\beta}_{12} = 0$ and hence (85) yield

$$\alpha \frac{\partial g(s)}{\partial s_{11}} - \underline{\beta}_{11} = \alpha \frac{\partial g(s)}{\partial s_{12}} + \overline{\beta}_{12} = 1.$$

This cannot happen due to $\frac{\partial g(s)}{\partial s_{11}} < \frac{\partial g(s)}{\partial s_{12}}$ and (87). Therefore we must have $(S_{11} - s_{11})s_{12} = 0$.

(d) We prove the two equations in similar ways by showing that any violation will lead to suboptimality. For (82), suppose on the contrary $s_{12}(S_{12} - s_{12})(S_{21} - s_{21}) > 0$ for some $0 < s_{12} < S_{12}$ and $s_{21} < S_{21}$, then (86) and $\underline{\beta}_{21} = 0$ from part (a) require $\overline{\beta}_{12} = \underline{\beta}_{12} = \overline{\beta}_{21} = \underline{\beta}_{21} = 0$. It hence follows from (85) that

$$\alpha \frac{\partial g(s)}{\partial s_{12}} = \alpha \frac{\partial g(s)}{\partial s_{21}} = 1,$$

thus $\alpha > 0$ and $1 < \frac{\partial g(s)}{\partial s_{11}} < \frac{\partial g(s)}{\partial s_{12}} = \frac{\partial g(s)}{\partial s_{21}}$. By the second derivatives in (84), increasing s_{12} and decreasing s_{21} will always maintain the inequality

$$1 < \frac{\partial g(s)}{\partial s_{11}} < \frac{\partial g(s)}{\partial s_{12}} < \frac{\partial g(s)}{\partial s_{21}}. \tag{90}$$

Therefore we can keep increasing s_{12} ($\Delta s_{12} > 0$) and decreasing s_{21} ($\Delta s_{21} < 0$) simultaneously such that the following equality holds at any subsequent s_{12} and s_{21}:

$$\Delta s_{12} \frac{\partial g(s)}{\partial s_{12}} + \Delta s_{21} \frac{\partial g(s)}{\partial s_{21}} = 0. \tag{91}$$

In this way we can maintain

$$\Delta g(s) = \sum_{i,j} \frac{\partial g(s)}{\partial s_{ij}} \Delta s_{ij} = \Delta s_{12} \frac{\partial g(s)}{\partial s_{12}} + \Delta s_{21} \frac{\partial g(s)}{\partial s_{21}} = 0,$$

i.e., keep $g(s)$ constant, while improving the objective function (service capacity) by

$$\Delta \bar{s} = \Delta s_{12} + \Delta s_{21} = \Delta s_{12} \left(1 - \frac{\partial g(s)/\partial s_{12}}{\partial g(s)/\partial s_{21}}\right) > 0,$$

which follows from (90) and (91), until $s_{12}(S_{12} - s_{12})(S_{21} - s_{21}) = 0$ is satisfied.

Similarly, for (83), suppose on the contrary $s_{11}(S_{11} - s_{11})(S_{21} - s_{21}) > 0$ for some $0 < s_{11} < S_{11}$ and $s_{21} < S_{21}$, then (86) and $\underline{\beta}_{21} = 0$ from part (a) require $\overline{\beta}_{11} = \underline{\beta}_{11} = \overline{\beta}_{21} = \underline{\beta}_{21} = 0$. It hence follows from (85) that

$$\alpha \frac{\partial g(s)}{\partial s_{11}} = \alpha \frac{\partial g(s)}{\partial s_{21}} = 1,$$

thus $\alpha > 0$ and $1 < \frac{\partial g(s)}{\partial s_{21}} = \frac{\partial g(s)}{\partial s_{11}} < \frac{\partial g(s)}{\partial s_{12}}$. By the second derivatives in (84), increasing s_{21} and decreasing s_{11} will always maintain the inequality

$$1 < \frac{\partial g(s)}{\partial s_{21}} < \frac{\partial g(s)}{\partial s_{11}} < \frac{\partial g(s)}{\partial s_{12}}. \tag{92}$$

Therefore we can keep increasing s_{21} ($\Delta s_{21} > 0$) and decreasing s_{11} ($\Delta s_{11} < 0$) simultaneously such that the following equality holds at any subsequent s_{11} and s_{21}:

$$\varDelta s_{11}\frac{\partial g(s)}{\partial s_{11}}+\varDelta s_{21}\frac{\partial g(s)}{\partial s_{21}}=0. \tag{93}$$

In this way we can maintain

$$\varDelta g(s)=\sum_{i,j}\frac{\partial g(s)}{\partial s_{ij}}\varDelta s_{ij}=\varDelta s_{11}\frac{\partial g(s)}{\partial s_{11}}+\varDelta s_{21}\frac{\partial g(s)}{\partial s_{21}}=0,$$

i.e., keep $g(s)$ constant, while improving the objective function (service capacity) by

$$\varDelta\overline{s}=\varDelta s_{11}+\varDelta s_{21}=\varDelta s_{21}\left(1-\frac{\partial g(s)/\partial s_{21}}{\partial g(s)/\partial s_{11}}\right)>0,$$

which follows from (92) and (93), until $s_{11}(S_{11}-s_{11})(S_{21}-s_{21})=0$ is satisfied.

It is then easy to verify that the necessary conditions (a)–(d) directly imply the three patterns stated in Parts (1)–(3) of Lemma 2. Note that for each pattern i, the service capacity $\overline{s_i}(n)$ increases with n from $n=\overline{s_i}^{-1}(n_1^A)$, where $\overline{s}=n_1^A$ is equal to the constant service capacity in zone (2), and up to $n=n_3^A$, the right end of zone (3). This also implies that $n_2^A=\min_i\{\overline{s_i}^{-1}(n_1^A)\}$. □

Next we prove the monotonicity of the per-driver profit rate with respect to n in zone (3).

Lemma 3. *Per-driver profit rate under control regime A, $\pi_A(n)$, is decreasing in n for $n\in(n_2^A,n_3^A]$.*

Proof. Lemma 2 shows that for participating capacity $n\in(n_2^A,n_3^A]$, the optimal capacity allocation *may alternate* among three patterns characterized by $s_i(n)$, with service capacity $\overline{s_i}(n)$ for $i=1,2,3$. To prove this lemma, we show that the per-driver profit rate given by (21) is decreasing for n varying within each of the 3 patterns or at *feasible* transitions between patterns. First, note by (78) in the proof of Lemma 2 that $n=g(s)$ in zone (3) and hence $\partial n/\partial s_{ij}=\partial g(s)/\partial s_{ij}$. Then:

(i) Within pattern (1): only s_{21} is increasing,

$$\begin{aligned}\pi'(n)&=\frac{\left[(\bar{\gamma}p-c)-c\frac{t_{12}}{t_{21}}\right]\left(\frac{\partial g(s)}{\partial s_{21}}\right)^{-1}n-[(\bar{\gamma}p-c)\overline{s}-cr_{12}]}{n^2}\\&=-\frac{S_{22}\bar{\gamma}p}{n^2}\frac{t_{12}}{t_{21}}\left(\frac{\partial g(s)}{\partial s_{21}}\right)^{-1}\left(1+\frac{S_{11}(\bar{\gamma}p-c)+S_{12}\bar{\gamma}p}{(s_{21}+S_{22})\bar{\gamma}p-\left(s_{21}+S_{22}+s_{21}\frac{t_{12}}{t_{21}}\right)c}\right)^2<0.\end{aligned}$$

(ii) Within pattern (2): for small n, s_{21} is increasing while $s_{12} = 0$,

$$\pi'(n) = \frac{\left[(\bar{\gamma}p - c) - c\frac{t_{12}}{t_{21}}\right]\left(\frac{\partial g(s)}{\partial s_{21}}\right)^{-1} n - [(\bar{\gamma}p - c)\bar{s} - cr_{12}]}{n^2}$$
$$= -\frac{S_{22}\bar{\gamma}p}{n^2}\frac{t_{12}}{t_{21}}\left(\frac{\partial g(s)}{\partial s_{21}}\right)^{-1}\left(1 + \frac{S_{11}(\bar{\gamma}p - c)}{(s_{21} + S_{22})\bar{\gamma}p - \left(s_{21} + S_{22} + s_{21}\frac{t_{12}}{t_{21}}\right)c}\right)^2 < 0.$$

For large n, $s_{21} = S_{21}$ and s_{12} is increasing,

$$\pi'(n) = \frac{\bar{\gamma}p\left(\frac{\partial g(s)}{\partial s_{12}}\right)^{-1} n - [(\bar{\gamma}p - c)\bar{s} - cr_{12}]}{n^2} = 0. \tag{94}$$

Note that $\pi(n)$ is continuous at the turning (non-differentiable) point where $s = (S_{11}, 0, S_{21}, S_{22})$.

(iii) Within pattern (3): for small n, s_{21} is increasing while $s_{11} = s_{12} = 0$,

$$\pi'(n) = \frac{\left[(\bar{\gamma}p - c) - c\frac{t_{12}}{t_{21}}\right]\left(\frac{\partial g(s)}{\partial s_{21}}\right)^{-1} n - [(\bar{\gamma}p - c)\bar{s} - cr_{12}]}{n^2} = -\frac{S_{22}\bar{\gamma}p}{n^2}\frac{t_{12}}{t_{21}}\left(\frac{\partial g(s)}{\partial s_{21}}\right)^{-1} < 0.$$

For medium n, $s_{21} = S_{21}$, s_{11} is increasing and $s_{12} = 0$,

$$\pi'(n) = \frac{(\bar{\gamma}p - c)\left(\frac{\partial g(s)}{\partial s_{11}}\right)^{-1} n - [(\bar{\gamma}p - c)\bar{s} - cr_{12}]}{n^2} = 0.$$

For large n, $s_{21} = S_{21}, s_{11} = S_{11}$ and s_{12} is increasing, we have the same (94). Note that $\pi(n)$ is continuous at the two turning (non-differentiable) points where $s = (0, 0, S_{21}, S_{22})$ and $s = (S_{11}, 0, S_{21}, S_{22})$.

As n increases, a *feasible transition* from pattern i to j at n must satisfy

$$\overline{s_i}(n) = \overline{s_j}(n) \quad \text{and} \quad \overline{s_i}'(n^-) < \overline{s_j}'(n^+). \tag{95}$$

Namely, pattern i and j have the same service capacity at transition n, and the service capacity increases faster after the transition. We then discuss all three possible transitions.

(i) Between pattern (1) and (2). If for pattern (2) s_{12} is increasing (or just reaches 0) at the transition, there must be $\overline{s_1}'(n^-) < \overline{s_2}'(n^+)$ since otherwise

$$\overline{s_1}(n_3^A) = \overline{s_1}(n) + \int_n^{n_3^A} \overline{s_1}'(n)dn > \overline{s_2}(n) + \int_n^{n_3^A} \overline{s_2}'(n)dn = \overline{s_2}(n_3^A),$$

hence it must be a transition from pattern (1) to (2): $(S_{11}, S_{12}, s_{21}, S_{22}) \to (S_{11}, s_{12}, S_{21}, S_{22})$. Obviously r_{12} jumps up and $\pi(n)$ jumps down at the transition. If for pattern (2) s_{21} is increasing at the transition, we have $s_1 = (S_{11}, S_{12}, s_{21}^{(1)}, S_{22})$, $s_2 = (S_{11}, 0, s_{21}^{(2)}, S_{22})$. By (95) there must be $s_{21}^{(1)} < s_{21}^{(2)}$ and hence

$$\overline{s_1}'(n) = \left(\frac{\partial g(s)}{\partial s_{21}^{(1)}}\right)^{-1} < \left(\frac{\partial g(s)}{\partial s_{21}^{(2)}}\right)^{-1} = \overline{s_2}'(n),$$

i.e., a transition from pattern (1) to (2): $(S_{11}, S_{12}, s_{21}^{(1)}, S_{22}) \to (S_{11}, 0, s_{21}^{(2)}, S_{22})$. Similarly we have $\pi(n)$ jumps down at the transition.

(ii) Between pattern (1) and (3). Due to similarities between pattern (2) and (3), the cases where s_{12} or s_{21} is increasing under pattern (3) have been similarly shown above. For the case where s_{11} is increasing (or just reaches 0) at the transition under pattern (3), there must be $\overline{s_1}'(n^-) < \overline{s_3}'(n^+)$ since otherwise

$$\overline{s_1}(n_3^A) = \overline{s_1}(n) + \int_n^{n_3^A} \overline{s_1}'(n)dn > \overline{s_3}(n) + \int_n^{n_3^A} \overline{s_3}'(n)dn = \overline{s_3}(n_3^A),$$

where $\int_n^{n_3^A} \overline{s_3}'(n)dn$ is an integration over n from n to $s_3^{-1}((S_{11}, 0, S_{21}, S_{22}))$ and from $s_3^{-1}((S_{11}, 0, S_{21}, S_{22}))$ to n_3^A. Therefore this must be a transition from pattern (1) to (3): $(S_{11}, S_{12}, s_{21}, S_{22}) \to (s_{11}, 0, S_{21}, S_{22})$, obviously r_{12} jumps up and $\pi(n)$ jumps down at the transition.

(iii) Between pattern (2) and (3). Similar to the analysis of the transitions between pattern (1) and (2) where $s_{11} \equiv S_{11}$ and we focus on s_{12} and s_{21}, here $s_{12} \equiv 0$ at any transition between pattern (2) and (3) so that we can adopt the same approach by focusing on s_{11} and s_{21}. The details are omitted. Note that the transitions, if any, are always from pattern (2) to (3).

□

Proof of Proposition 5. From Lemma 2, it is optimal to reject rider requests at the low-demand location for some n in zone (3) if and only if pattern (2) and/or (3) provide the *largest* service capacity at some $n \in (n_2^A, n_3^A)$, i.e., $\exists n \in (n_2^A, n_3^A)$ such that $\overline{s_1}(n) \neq \max_{i \in \{1,2,3\}} \overline{s_i}(n)$. We need to compare the three patterns in terms of their service capacity $\overline{s_i}(n)$, $i = 1, 2, 3$. We have the following three observations.

(i) At the right end of zone (3), $\overline{s_1}(n_3^A) = \overline{s_2}(n_3^A) = \overline{s_3}(n_3^A) = \overline{S}$.

(ii) For n close to n_3^A $(n \to n_3^{A-})$, it follows from Lemma 2 that pattern (1) has $s_1(n) = (S_{11}, S_{12}, s_{21}, S_{22})$ with s_{21} varying, while pattern (2) and (3) both have $s_2(n) = s_3(n) = (S_{11}, s_{12}, S_{21}, S_{22})$ with s_{12} varying. Therefore we have

$$\begin{aligned} \overline{s_1}'_-(n_3^A) &= \left.\frac{\partial \overline{s}/\partial s_{21}}{\partial g(s)/\partial s_{21}}\right|_{s=S} = \left(\left.\frac{\partial g(s)}{\partial s_{21}}\right|_{s=S}\right)^{-1} > 0, \\ \overline{s_1}''_-(n_3^A) &= \left.\frac{\partial \overline{s_1}'_-(n_3^A)/\partial s_{21}}{\partial g(s)/\partial s_{21}}\right|_{s=S} = -\left.\frac{\partial^2 g(s)/\partial s_{21}^2}{(\partial g(s)/\partial s_{21})^3}\right|_{s=S} < 0, \end{aligned}$$

i.e., $\overline{s_1}(n)$ is strictly convex and increasing in n near n_3^A. And

$$\begin{aligned} \overline{s_2}'_-(n_3^A) = \overline{s_3}'_-(n_3^A) &= \left.\frac{\partial \overline{s}/\partial s_{12}}{\partial g(s)/\partial s_{12}}\right|_{s=S} = \left(\left.\frac{\partial g(s)}{\partial s_{12}}\right|_{s=S}\right)^{-1}, \\ \overline{s_2}''_-(n_3^A) = \overline{s_3}''_-(n_3^A) &= \left.\frac{\partial \overline{s_2}'_-(n_3^A)/\partial s_{12}}{\partial g(s)/\partial s_{12}}\right|_{s=S} = -\left.\frac{\partial^2 g(s)/\partial s_{12}^2}{(\partial g(s)/\partial s_{12})^3}\right|_{s=S} = 0, \end{aligned}$$

i.e., $\overline{s_2}(n)$ and $\overline{s_3}(n)$ both increase linearly in n near n_3^A.

(iii) The proof of Lemma 3 establishes that any feasible pattern of transitions as n *increases* in zone (3) must be from pattern (1) to (2), from pattern (1) to (3), or from pattern (2) to (3)—not vice versa.

Using the above observations, the sufficient and necessary condition that patterns (2) and (3) provide the *largest* service capacity at some $n \in (n_2^A, n_3^A)$ is

$$\overline{s_2}'_-(n_3^A) = \overline{s_3}'_-(n_3^A) < \overline{s_1}'_-(n_3^A). \tag{96}$$

To see this, if (96) holds, observation (i) and (ii) immediately imply that $\overline{s_2}(n_3^{A-}) = \overline{s_3}(n_3^{A-}) > \overline{s_1}(n_3^{A-})$, hence patterns (2) and (3) provide the largest service capacity near n_3^A. On the other hand, if (96) does not hold, observation (i) and (ii) imply that $\overline{s_2}(n_3^{A-}) = \overline{s_3}(n_3^{A-}) < \overline{s_1}(n_3^{A-})$, i.e., pattern (1) is optimal near n_3^A. It then follows from observation (iii) that there is no transition from pattern (2) or (3) to pattern (1) as n increases in zone (3), hence pattern (1) is optimal throughout zone (3). Therefore (96) is necessary for patterns (2) and (3) to be optimal somewhere in zone (3). Putting in the derivative expressions from the proof of Lemma 2 and some algebraic manipulation will transform (96) to inequality (37) in the Proposition. □

S2 Driver Supply and Actual Gains in Platform Revenue and Per-Driver Profit

In this section we illustrate the impact of the driver supply characteristics, specifically, the outside opportunity cost distribution F, on the *actual* platform revenue and per-driver profit gains, compared to the upper bounds in Proposition 8 and 9, and on the tension between the drivers' and the platform's gains. For simplicity we focus on the gains from admission control, i.e., regime A over M. (Similar effects determine the actual gains from repositioning.)

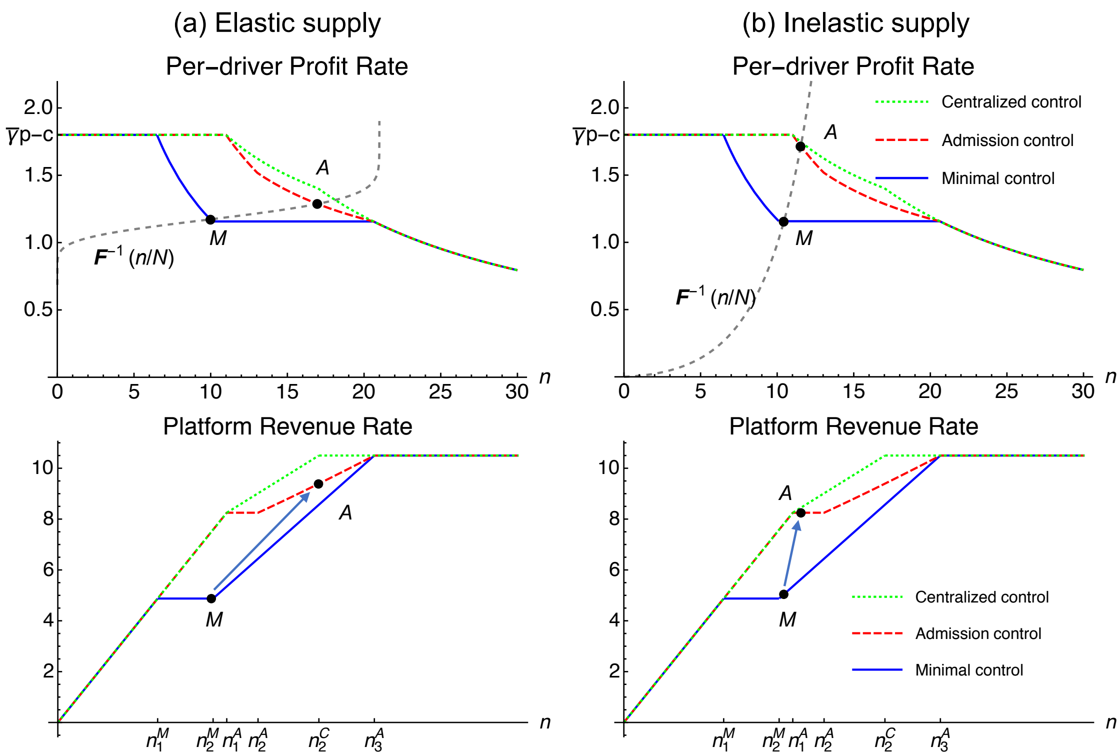


Figure 7: Impact of admission control on the equilibrium capacity, per-driver profit, and platform revenue $(S = (3, 1, 4, 6), \overline{S} = 14, N = 21, t = 1, \gamma = 0.25, p = 3, c = 0.45)$

Figure 7 illustrates these gains for two opportunity cost distributions. Panel (a) presents a case where admission control yields large benefits for the platform as a result of a large increase in driver participation, and consequently only small benefits for individual drivers. Specifically, the top chart in panel (a) shows for the three control regimes the per-driver profits that are non-increasing functions of the capacity, and the increasing marginal opportunity cost function $F^{-1}(n/N)$. Achieving the upper bound on platform revenue gains from admission control requires two conditions, namely, $n_M^* = n_2^M$ or equivalently, $F^{-1}(n_2^M/N) = \pi_M(n_2^M)$, and $n_A^* = n_3^A$. The first condition holds in the example, the second condition requires infinitely elastic supply around the profit level $\pi_M(n_2^M)$, i.e., that F grows sufficiently fast around this point such that $n_3^A - n_2^M$ additional drivers join if the per-driver profit is slightly larger, so that $F^{-1}(n_3^A/N) = \pi_M(n_3^A)$. The example depicted in Figure 7 (a) shows how the upper bound can be approached if the supply increases substantially for a moderate change in per-driver profit rate.

Panel (b), in contrast, presents a case where admission control (under regime A or C) yields the maximum achievable per-driver profit gains as a result of a small increase in driver participation, and consequently only modest platform revenue gains. As shown in the top chart of panel (b), in this case the marginal opportunity cost function yields the same equilibrium capacity under minimal control as in panel (a), i.e., $F^{-1}(n_2^M/N) = \pi_M(n_2^M)$; however, the driver supply is so inelastic that the number of drivers willing to participate at the maximum profit rate ($\overline{\gamma}p - c$) is smaller than the minimum number required to serve all riders without repositioning, that is, $n_A^* \leq n_1^A$ where $F^{-1}(n_A^*/N) = \overline{\gamma}p - c$. The platform's commission is too high to entice more drivers to participate.