Coordinate Package Delivery and On-Demand Rides: A Zoning Policy and Analysis

Junyu Cao
McCombs School of Business, The University of Texas at Austin. junyu.cao@mccombs.utexas.edu
Sheng Liu
Rotman School of Management, University of Toronto. Sheng.Liu@rotman.utoronto.ca

Cities are facing mounting pressure from the surging demand for package delivery and on-demand ride services. Meanwhile, more drivers are empowered by platforms to provide flexible transportation and logistics services with their own vehicles. Integrating freight and passenger transportation services, termed co-modality, has been promoted as a promising means to improve vehicle utilization, driver productivity, and logistics efficiency. Motivated by these issues, we investigate the value of co-modality to the driver in the context of joint package delivery and ride-hailing from a revenue-maximizing perspective. The driver must decide when and how on-demand ride jobs should be coordinated with package deliveries. Because the general optimal coordination problem is intractable, we develop a zoning-based coordination policy that allows us to capture the synergies between these two types of jobs and characterize its revenue performance analytically. We derive an asymptotically tight lower bound on the revenue rate with approximations and identify three optimal solutions for this bound: the pure package delivery policy, the four-zone policy (limited ride opportunities), and the switching policy (flexible ride opportunities). Our analysis reveals that co-modality is most beneficial when package volume is relatively low, while limited flexibility can be optimal when ride profit and package volume are moderate. In addition, we analytically quantify the optimality gap between these three policies and the true optimal zoning policy; numerically, they achieve over 99.7% of the optimal revenue. We also develop a full-information benchmark that helps us characterize the potential optimality loss of the optimal zoning policy. Lastly, we complement our analysis using numerical studies based on taxi data from New York City and evaluate the value of coordination in practical scenarios.

Key words: co-modality, ride-hailing, logistics, smart city operations.

1. Introduction

As urbanization accelerates and e-commerce prospers, the demand for delivery and ride services continues growing steadily. Many cities worldwide are overwhelmed by vehicles moving passengers and goods, raising heightened concerns for congestion and sustainability. NYC Department of Transportation (2018) report that the travel speeds in the Manhattan Central Business District (CBD) of New York City saw a decline of 21% from 2010 to 2016, reflecting the growing demand for the city's road network. In particular, the rise of on-demand transportation platforms and crowdsourcing delivery services (such as parcel delivery or grocery delivery) complicate the policymaker's efforts in

confronting these issues. For example, ride-hailing services may substitute more sustainable transit modes and result in worse congestion, as suggested by empirical evidence from three major Indian cities (Agarwal et al., 2023).

In a typical urban environment, vehicles delivering packages and serving passenger rides are operated separately. It is estimated that urban passenger and delivery traffic account for 74% and 26% of the total urban transportation-related CO_2 emissions (Fehn et al., 2023). These two types of traffic cause competing demand for road space, forcing some municipalities to limit specific types of traffic to their core urban areas, see, e.g., circulation restriction zones in the city of São Paulo, Brazil (Bontempo et al., 2014). Whereas investing in transportation infrastructure may alleviate this tension, it is not always financially or physically feasible.

With the emergence of on-demand and e-commerce platforms, it has become viable for self-scheduling drivers to participate in package delivery and passenger rides operations using their own vehicles (see a relevant discussion on comparing these two jobs on UberPeople[1]). For example, crowdsourcing drivers can sign up for package delivery for Amazon according to their own schedules (Amazon Flex, 2023). They may deliver up to 50 packages per trip in a time slot of 3-5 hours. The integration and coordination of package delivery and mobility services, coined as co-modality, is considered a promising solution for improving the efficiency and sustainability of urban logistics and mobility systems (Ronald et al., 2016). As automation technology advances, vehicles also become smarter and can meet multifaceted demands in a more seamless fashion. For example, Toyota has unveiled a new multi-purpose vehicle concept, called e-Palette, to combine mobility, freight, and shopping utilities with a modular design approach (Toyota, 2018).

Conceptually, co-modality can improve vehicle utilization and courier productivity at the expense of increased system complexity. From an operations perspective, it is challenging to characterize the optimal coordination policy and evaluate the value of coordination between package delivery and on-demand passenger rides analytically. Specifically, the spatial distribution of package locations and ride origins/destinations gives rise to a high-dimensional state and decision space. The optimal routing of vehicles in visiting these locations in the face of ride request uncertainty is notoriously difficult and often calls for easy-to-implement heuristics (Arslan et al., 2019). Moreover, when the drivers are self-scheduled, it is critical to understand the impact of coordination on their revenues and make sure the incentives are compatible.

In this paper, we study a co-modal delivery scenario in which a driver can deliver packages and serve on-demand passenger rides at the same time (packages and passengers take the same vehicle). Specifically, the driver preloads packages from a warehouse and delivers them to customers

[1] https://www.uberpeople.net/threads/delivery-vs-passengers.392703/

in a predetermined service region, following the common package delivery practice in e-commerce applications. When the driver delivers packages (e.g., based on the solution to a traveling salesman problem, TSP), he/she can also accept real-time ride requests from an on-demand ride platform. Unlike package delivery to known destinations, passenger ride requests are uncertain in both temporal and spatial dimensions. Intuitively, the driver would prefer to take a ride with a pickup location near one of the package drop-off locations. Moreover, ideally, the drop-off location of the ride may not fall in the areas that the driver has visited during previous delivery trips (otherwise, it could lead to an "inefficient" route to deliver the remaining packages). We operationalize these rules in a new coordination policy based on zoning – the use of zoning is prevalent in practical delivery systems and often leads to interpretable routes preferred by drivers (Arslan and Abay, 2021), and our policy leverages zoning to trade off the flexibility of serving on-demand rides and package delivery efficiency.

We build our analytical model to shed light on the potential of coordinating package delivery and on-demand rides *from the driver's perspective* by understanding the revenue/productivity implications. The key part of our analysis is to characterize the synergy between package delivery and serving on-demand passenger rides. We note that the considered setting is not limited to the coordination of package delivery and passenger rides but scheduled jobs and on-demand jobs in a spatial context, where the on-demand jobs include any mobility/logistics services that are associated with an origin-destination pair.

1.1. Our Contributions

The main contributions of our paper can be summarized below.

1.1.1. Modeling of A Co-Modal System. We introduce a parsimonious co-modal logistics setting that differentiates package delivery and on-demand rides in two dimensions: (i) spatial uncertainty and (ii) routing. On the one hand, package delivery locations are revealed beforehand, whereas on-demand ride pickup and drop-off locations are only known upon arrival. On the other hand, packages are preloaded and can be delivered through a TSP route, while on-demand rides are point-to-point trips. Our model captures various trip components that connect the two types of jobs, including the deviation from the TSP route due to pickups and drop-offs. To the best of our knowledge, we are the first to study this co-modal system analytically with a focus on the driver's revenue performance.

1.1.2. A Zoning-Based Coordination Policy and Its Revenue Performance. We develop a new coordination policy that coordinates package delivery and on-demand ride requests by partitioning the service region into zones. The key idea is to use zones to filter and sequence package delivery and on-demand ride jobs based on their locations. Specifically, the decision to accept the next ride

request depends on the proximity of its origin and whether its destination zone has already been delivered. This coordination policy enjoys the desirable theoretical and practical properties of zoning in optimal route planning, easing the need for real-time re-optimization. Furthermore, it goes beyond the existing partitioning research in logistics to present a unique way to balance delivery efficiency and service flexibility of ride-hailing. Notably, varying the number of zones changes the level of flexibility of the driver in responding to on-demand rides. By analyzing the stochastic driver trajectory, we characterize the expected revenue rate of the zoning policy.

1.1.3. Optimal Zoning Policy and Its Performance Guarantee. We first examine a special case where the service region is divided into four symmetric zones and derive exact conditions under which the zoning policy outperforms the pure package delivery policy. For the general case where the number of partitioned zones needs to be optimized, we develop approximation schemes to estimate the trip distances between zones at different system states by exploiting a connection to the coverage problems in geometry that can be of independent interest. This gives rise to a lower bound of the expected revenue rate function that is asymptotically tight for a large number of zones. We prove that the revenue maximizer of the lower bound only takes three forms: the pure delivery policy, the four-zone policy, and the flexible switching policy (the driver can switch between package delivery and on-demand rides constantly). We then provide theoretical and numerical evidence to show that focusing on the above three policies can get performance close to that of the optimal zoning policy. Furthermore, we develop a full-information benchmark that gives an upper bound of the revenue rate of any coordination policy. We characterize the performance ratio of the optimal zoning policy to this benchmark and show that the zoning policy is close to optimal for a large number of packages.

1.1.4. Value of Co-Modality with Zoning. Our analysis offers a set of insights for operating co-modal transportation systems and understanding the value of coordinating package delivery and on-demand rides. First, we show that the zoning policy outperforms the pure package delivery policy when the package density is not too high and the profit from on-demand rides is large enough. Crucially, a higher package volume does not favor coordination; instead, the value of coordination diminishes relatively. In addition, we characterize the minimum revenue improvement of the optimal zoning policy against the pure delivery policy and delineate its dependence on the system parameters. We complement our analysis with a numerical study with parameters calibrated using the New York City taxi data. The numerical results validate our analysis and provide hints on how the revenue rate evolves with package volume and ride arrival rate in practical scenarios. In particular, we find that the revenue improvement from coordination can be up to 60% when the delivery capacity is very limited, which tends to be the case for crowdsourcing vehicles. Last but not least, the optimal zoning

policy's revenue rate achieves at least the same scaling pattern as the pure package delivery policy with respect to package density.

The remainder of this paper is organized as follows. Section 1.2 discusses related literature. Section 2 introduces the model setup for coordinating package delivery and on-demand rides, followed by the proposed zoning-based coordination policy. Section 3 analyzes the four-zone policy and its optimality conditions. Section 4 studies the general zoning policy and characterizes its structure under different market conditions, supported by a numerical study. Section 5 concludes the paper.

1.2. Literature Review

Our work on coordinating package delivery and on-demand rides is related to the three streams of literature: last-mile delivery (with shared mobility), ride-hailing, and co-modality in transportation.

1.2.1. Last-mile Delivery. Last-mile delivery has received growing attention due to the rapid evolution of e-commerce. The most relevant to our work is the line of research on shared mobility for last-mile delivery, where shared passenger vehicles are dispatched to deliver packages (see practical examples such as Amazon Flex (2023)). Qi et al. (2018) consider a shared-mobility logistics setting that features a central warehouse and multiple subregions. They decide the optimal size (number) of the delivery subregions to minimize the total delivery cost under a continuous approximation scheme. Arslan et al. (2019) formulate a dynamic pickup and delivery problem that matches delivery jobs to both dedicated and ad hoc drivers. They apply a rolling-horizon approach that handles both the dynamically arriving jobs and ad hoc drivers. Cao et al. (2020) develop a framework that combines the use of ride-sharing platforms with traditional in-house van delivery systems for optimizing last-mile delivery. They formulate a discrete sequential packing problem, in which the drivers select bundles of packages and collect them from the warehouse in a decentralized manner. Fatehi and Wagner (2022) propose a robust crowdsourcing optimization model to study labor planning and pricing for crowdsourced last-mile delivery systems. Behrendt et al. (2023) consider a scheduling problem for the crowdsourced delivery platform, in which a mix of scheduled and ad hoc couriers serves dynamically arriving pickup and delivery orders. They present a prescriptive machine learning method that combines simulation optimization for offline training and a neural network for online solution prescription. These papers mainly treat crowdsourcing couriers as an additional supply with self-scheduling behaviors and do not consider the possibility of coordinating package delivery and on-demand rides.

In addition, our work contributes to the emerging literature on same-day and food delivery operations management (Mao et al., 2019; Yildiz and Savelsbergh, 2019; Liu et al., 2021; Chen and Hu, 2023; Feldman et al., 2023; Besbes and Cachon, 2023; Carlsson et al., 2024). From a methodological

perspective, our work builds on the idea of region partitioning in logistics planning. In particular, Voronoi diagrams and ham sandwich cuts are widely applied for partitioning the service region such that each partitioned zone is served by one vehicle (Ouyang, 2007; Carlsson, 2012; Banerjee et al., 2022a). The main goal of the existing partitioning studies is to balance the workload among delivery zones. In contrast, we leverage partitioning to filter and match ride requests with the delivery driver contingent on the location of the vehicle.

1.2.2. Ride-hailing. On-demand ride-hailing platforms such as Uber and Lyft have become a pivotal part of modern transportation systems. We refer the reader to Furuhata et al. (2013) and Wang and Yang (2019) for comprehensive reviews of ride-hailing and -sharing systems. Specifically, the study of optimal pricing, matching, and relocation decisions on the two-sided ride-hailing market has received particular interest from operations management (Banerjee et al., 2015; Cachon et al., 2017; Braverman et al., 2019; Yan et al., 2020; Ata et al., 2020; Afèche et al., 2023; Banerjee et al., 2022b; Garg and Nazerzadeh, 2022). For example, Taylor (2018) and Bai et al. (2019) investigate the optimal price and wage payment schemes for on-demand service platforms considering delay-sensitive customers and independent service providers. Guda and Subramanian (2019) further analyze the interactions between surge pricing and drivers' relocation decisions on ride-hailing platforms. More generally, the optimal spatial (and temporal pattern) of surge pricing has been closely examined under a market equilibrium (Bimpikis et al., 2019; Besbes et al., 2021; Hu et al., 2022; Ma et al., 2022; Chen et al., 2023b). These papers highlight the technical difficulty stemming from the spatial dimension of the ride-hailing problem. On the matching side, different matching mechanisms have been proposed to better tradeoff passenger waiting time, driver earning rate, and overall system efficiency (Castro et al., 2021; Feng et al., 2021). With a similar rationale, our coordination policy seeks to match passenger rides with drivers performing package delivery trips, but our goal is to understand the underlying logistics efficiency instead of the market equilibrium. Relatedly, Martin et al. (2024) analyze the value of coordinating passengers and drivers on ride-hailing platforms through a static walking policy. Unlike most ride-hailing settings that are concerned about point-to-point trips, (preloaded) package delivery jobs are pooled together through a TSP tour that requires a different analysis.

1.2.3. Co-modality in Transportation. Co-modal systems have been traditionally used in medium- and long-haul transportation and have recently gained traction for short-haul or intra-city logistics (Zhu et al., 2023). Existing studies have explored both public transportation infrastructure and private transportation networks for the integration of passenger and goods flows (for an excellent review of co-modality, we refer the reader to Cleophas et al. 2019). For example, Dampier and Marinov (2015) examine the feasibility of transporting freight with the urban railway network (a metro system) through a case study in Newcastle upon Tyne, UK. Li et al. (2021) optimize the urban rail transit

schedule to facilitate passenger and freight movement. Masson et al. (2017) and Azcuy et al. (2021) study a two-tier urban delivery system that leverages existing transit routes (bus routes or tramways) to move goods to intermediate transfer locations before being delivered by dedicated freight vehicles. However, the efficiency of such integrated systems hinges on the connectivity of transit networks, which may lack the capacity and investment to support its self-sufficiency (Cochrane et al., 2017).

Closest to our setting is the stream of literature on co-modality using on-demand transportation vehicles. In Li et al. (2014), the authors propose a Share-a-Ride problem that dispatches taxis to transport people and deliver parcels at the same time. They formulate the problem as a mixed integer linear program (MILP) and develop several heuristics to solve the problem. This framework has been extended to account for practical constraints and uncertain travel times (Nguyen et al., 2015; Li et al., 2016). Due to the complexity of the problem, more recent studies resort to reinforcement learning and agent-based simulations to prescribe solutions in real-time, see, e.g., Li et al. (2022); Chen et al. (2023a); Fehn et al. (2023). Although these studies provide promising computational methods (or heuristics) and case studies, there lacks a coherent and generalizable understanding of the potential of co-modality in integrating package delivery jobs and on-demand passenger ride requests. Moreover, unlike the existing literature that assumes the delivery requests pop up along with passenger ride requests (so their main differences only lie in the service time constraint), we focus on the setting where the driver has preloaded packages with known destinations, which is common for e-commerce applications.

2. Model Setup and Coordination Policy

In this section, we introduce the basic setting for modeling package delivery and on-demand rides, followed by the proposed coordination policy.

2.1. Model Setup

We consider a driver who needs to deliver n packages (on average) in a service region $\mathcal{A}$. The packages are preloaded from a warehouse/delivery station. Without loss of generality, we assume the service region is of a diamond shape with edge length R, in which the travel distance is measured in the ℓ_1-distance (Manhattan distance). As a result, the points on the periphery of the service region share the same distance of $R/\sqrt{2}$ from the centroid (see Figure 1), and the area of the service region is $A = R^2$. Package locations are assumed to follow a homogeneous two-dimensional point process with rate γ such that $n = \gamma R^2$. Note that this assumption is for analytical tractability, and our coordination policy can work for generally distributed package locations, which will be discussed in Section 2.2.

The driver travels with a fixed speed of v. Each delivered package contributes a unit revenue of r_p to the driver. The package-based compensation scheme is adopted by the majority of crowdsourcing

delivery platforms, including Postmates, Deliv, and DHL MyWays (Alnaggar et al., 2021). In addition to package delivery, the driver can take on-demand passenger ride requests from a platform. Accordingly, the driver is paid r_t per unit distance of passenger ride.[2] However, unlike package delivery to known customer locations, passenger ride requests follow a stochastic process with uncertain pickup and drop-off locations (the locations are only realized upon request arrival). Specifically, we assume that the ride arrival process follows a Poisson distribution with rate λ, and both the origins and destinations are distributed uniformly over the service region. Here, λ is taken to be exogenous, allowing us to focus on the logistics tradeoff between package delivery and on-demand rides. Understanding the platform-driver interactions is important to study the general management problems of co-modality systems, which we leave as a future research question. To prioritize the passenger ride experience, we assume the driver can not make deliveries when the passenger is already on board.

The driver desires a policy π to coordinate package delivery and on-demand rides. Specifically, a coordination policy decides whether to accept an on-demand ride based on its origin and destination and the remaining packages to be delivered at any given time point. The goal is to maximize the expected revenue rate, which is defined by

$$R^{\pi}(n,\lambda) = \frac{\mathbb{E}[\mathsf{Revenue}]}{\mathbb{E}[\mathsf{Total\ delivery\ time}]}, \tag{1}$$

where the total delivery time includes travel time for delivery, passenger ride time, and connecting time between the two types of jobs (from a package delivery location to a passenger pickup location or from a passenger drop-off location to a package delivery location). The revenue function (1) stands for the long-run average revenue rate for the driver. We assume the delivery processes on different days are independent of each other, so one can treat the delivery process on each day as a renewal process. Then, according to the renewal reward theorem (Ross, 2014, Proposition 7.3), the empirical average of reward converges, almost surely and in mean, to (1). Note that the expected revenue rate can be interpreted as the driver productivity that measures the operational efficiency of co-modality.

As a benchmark, the expected revenue rate of package delivery (referred to as the pure delivery policy, PD), when the driver does not take any passenger ride requests, can be estimated as

$$R^{PD}(n,\lambda) = \frac{vnr_p}{\beta\sqrt{nA}} = \frac{v\sqrt{n}r_p}{\beta R},$$

where $\beta\sqrt{nA}$ approximates the expected TSP travel distance incurred to visit n locations in the service region of area A, following the Beardwood-Halton Hammersley (BHH) Theorem (Beardwood et al., 1959).

[2] Whereas the ride-hailing fare may include a fixed component, the main differentiating revenue structure between scheduled delivery jobs and ride-hailing jobs lies in the per-mile trip fare. Our result will hold if we introduce a fixed revenue part to ride-hailing trips.

The uncertainty of passenger ride requests poses a major challenge to the efficient coordination of the two types of jobs. While the driver can follow a predetermined route for package delivery (e.g., using a TSP solution), the presence of random pickup and drop-off locations causes unavoidable detours and requires trading off the route efficiency and the uplift of revenue carefully. This problem may be modeled as a dynamic program, where the state consists of the locations of all undelivered packages as well as the passenger's pick-up and drop-off points. This general formulation does not impose restrictions on the driver's visit sequence through the package locations (e.g., no need to adhere to the original TSP tour), and tracking each package's exact location is essential because the spatial distribution of packages affects the future coordination policy. The resulting formulation is difficult to analyze because (i) the state space is high given the number of package locations; (ii) the acceptance policy (of on-demand rides) needs to be state-dependent. This motivates us to seek a tractable coordination policy with performance guarantees.

In the next subsection, we propose a novel zoning-based coordination policy that facilitates the assessment of the value of joint package delivery and on-demand rides.

2.2. A Zoning-Based Coordination Policy

Under this policy, the entire service region is partitioned into N zones (subregions). We consider $N = k^2$ for $k \in \{m \in \mathbb{N}^+ | 1 \le m \le \sqrt{n}\}$, so the zones are of the same size, and they are evenly distributed across the service region. We note that this is mainly for ease of analysis because the partition for a general number of zones does not yield a clear characterization. To simplify the analysis but without loss of generality, we further assume that each zone contains the same number of packages, i.e., n/N. We note that this is mainly a technical assumption, and when the package distribution exhibits more variability, our result can serve as a first-order approximation for analyzing the policy performance. Moving beyond the uniform distribution, the design of the partition can adapt to spatially heterogeneous package demand and ride arrivals, using techniques from Ouyang (2007); Carlsson et al. (2024). We provide examples of such partitions in Appendix A. Within each zone, the driver follows the optimal route through the package delivery locations; the driver also incurs additional connecting trips to travel between different zones.

Without on-demand rides, the above zoning algorithm has claimed both theoretical and practical significance for solving TSPs. The seminal work of Karp (1977) shows that zoning algorithms enjoy a relative error of $O((n/N)^{-1/2})$ in approximating the optimal TSP length. Moreover, they yield near-optimal performance when combined with other TSP heuristics. In practice, zoning algorithms are often preferable because of their interpretability and ease of implementation. Based on an extensive analysis of Amazon's delivery routes, Arslan and Abay (2021) observe that the majority of Amazon

drivers visit customers in a sequence of zones and conclude that zoning is critical for improving delivery efficiency.

We now describe how zoning can be extended for co-modal logistics in the presence of on-demand rides. During package delivery within a zone, the driver starts screening passenger rides once there are t_0 minutes left until the completion of delivery jobs in that zone. A passenger ride is accepted if (i) the pickup location is within a distance of $r = \alpha R/\sqrt{2}$ from the final delivery location[3] and (ii) the drop-off location is in a zone that the driver has not delivered.[4] The distance r can be interpreted as the maximum allowable pickup distance, and the arrival rate of the admissible rides is $\alpha^2\lambda$. The driver will serve the accepted ride, if any, after the package deliveries are completed in the zone. Therefore, the screening time t_0 together with pick-up distance r affect the passenger waiting time and can be adjusted to meet service quality constraints.[5] If no rides are accepted, the driver moves to the next unvisited zone for package delivery following a prespecified sequencing rule. In the remainder of the paper, we assume the "nearest-neighbor" rule is adopted unless otherwise stated, e.g., the driver travels to the nearest unvisited zone for package delivery if no qualified rides are received.

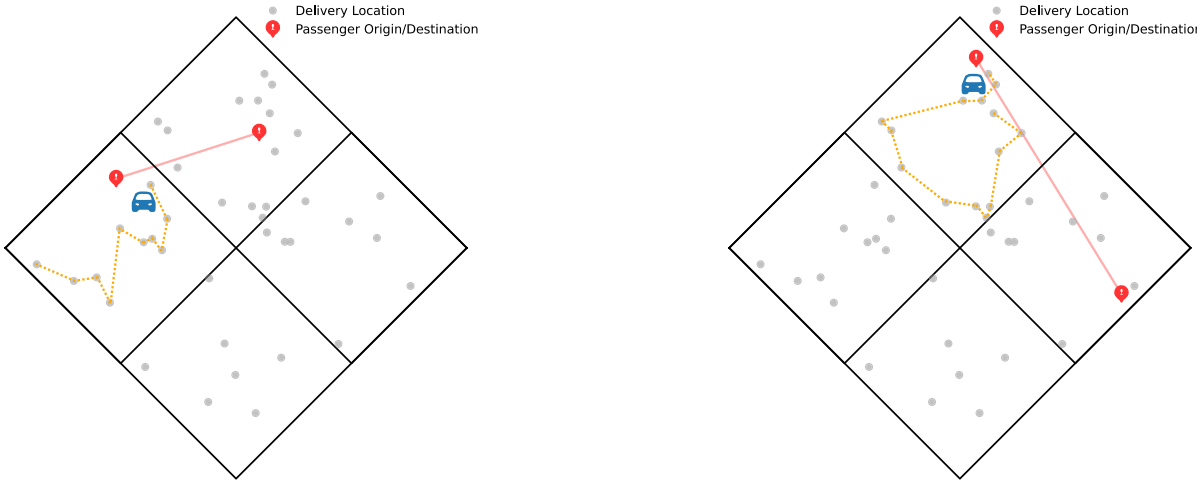
Under the zoning policy, the number of zones indicates the maximum number of on-demand rides a driver can accept during the package delivery process: the driver can take up to $N-1$ rides when N zones are in place. In other words, more zones imply more flexibility for the driver to serve on-demand rides, which manifests the unique role of zoning in the co-modal application. Accordingly, one can impose additional constraints on N to satisfy work hours or labor constraints so the total expected delivery time is not too long. In extreme cases, if $N = 1$, the driver does not take any rides, and no extra working hours are incurred.

Figure 1 depicts one realization of the zoning-based coordination policy with four zones (left, top, right, and bottom zones). In this example, the driver starts from the left zone and first delivers packages within its boundary. Upon finishing all delivery jobs in the left zone, the driver accepts a passenger ride that ends in the top zone; after completing this ride, the driver moves to deliver packages there before accepting another ride to the right zone. Then, the driver makes package deliveries within the right zone and travels to the bottom zone (the last unvisited zone) for the remaining package deliveries. In this case, there are no qualified passenger rides that the driver can accept when delivering in the right zone. Therefore, the last segment of the trip does not reap the benefit of joint package delivery and ride-hailing for this realized scenario.

[3] It is possible that the origin of a passenger ride request is outside the service region when the final delivery location is close to the region boundary. We include those requests to ease the need to treat boundary cases.

[4] For example, Uber allows drivers to filter requests by the order destination; see https://www.uber.com/us/en/drive/basics/driver-destinations/.

[5] Technically, t_0 should be shorter than the total package delivery time in a zone, which is likely to hold for practical scenarios.



(a) First passenger ride (b) Second passenger ride

Figure 1 The zoning policy with $N = 4$ is run with two passenger rides: the first coming from left to top and the second coming from top to right. The dashed (yellow) line indicates the delivery route within a zone, whereas the (red) solid line connects the pickup and drop-off locations of an on-demand ride.

REMARK 1. The proposed zoning policy can be viewed as a static decision rule, where the control parameters such as α and t_0 are fixed. Note that increasing α or t_0 results in more ride-hailing opportunities at the expense of higher pickup time for the passengers. In practice, these parameters are likely controlled by the platform and can be dynamically adjusted in response to ride arrival uncertainties. Moreover, if the market condition is time-varying, we may design time-dependent zoning policies, where the driver's response behavior is time-dependent, or re-solving policies that reoptimize the zoning based on the updated information.

REMARK 2. The rationale behind restricting the ride destination to unvisited zones is to maximize the synergy between package delivery and passenger rides: taking rides to unvisited zones reduces (profitless) inter-zone travels that otherwise will be incurred during package delivery. Relaxing this restriction can open up more ride opportunities at the expense of adding detours and lengthening the total delivery time. In Appendix B, we simulate the zoning policy that allows coming to previously visited zones and show it is less promising than the proposed zoning policy.

3. Four-Zone Policy

To shed light on the efficiency of the zoning-based coordination policy, we first study a special class of the zoning policy with $N = 4$ equal-size zones formed within the service region, as illustrated in Figure 1. In this case, the constructed zones are symmetric, in the sense that the distances between the centroids of any pair of zones are equal to $\frac{\sqrt{2}}{2}R$ in ℓ_1-distance. Following the uniform distribution

assumption, each zone observes the same passenger request arrival rate of $\lambda/4$. We call this policy the four-zone policy (FZ).

Because the driver only accepts ride requests going to a zone that is yet to be delivered, the probability of taking up a passenger ride upon finishing the package delivery in the i-th visited zone is

$$\begin{aligned}&\mathbb{P}(\text{accept a passenger ride request when delivering in the } i^{th} \text{ visited zone})\\ =&1-\mathbb{P}(\text{no requests with a drop-off location from the } (4-i) \text{ unvisited zones}) = 1-\exp(-(4-i)\alpha^2\lambda t_0/4),\end{aligned}$$

where $\alpha^2\lambda$ is the passenger ride arrival rate to the admissable area that is within αR from the final delivery location and $(4-i)/4$ is the probability that the drop-off location is from an unvisited zone.

We introduce the key components of the driver's travel distance under the four-zone policy:

(I) The first component, L_1, captures the expected passenger ride distance between a random pickup and a drop-off location from different zones, which can be computed as $c_1(\alpha)R$. The dependence of L_1 on α is due to the fact that the passenger ride origin location is affected by the maximum pickup distance r. For a given value of α, we show in Appendix C how to evaluate L_1 by integrating the density function of pickup and drop-off locations. To simplify the integral, we approximate the distribution of the drop-off location by a uniform distribution over the service zone; technically, the drop-off location is only uniformly distributed over the unvisited zones, but the randomness of the unvisited zones makes the drop-off location approximately uniform across the whole region (see Appendix C for numerical comparisons).

(II) The second component, L_2, characterizes the expected (additional) pickup and drop-off distance required to serve the passenger ride request. Specifically, the passenger pickup trip starts from the last package delivery location towards the origin of the passenger request, of which the distance can be computed as $\alpha_2 R$. The drop-off distance accounts for the trip from the passenger drop-off location (destination) to the nearest package delivery location in the same zone, which we can compute as $c_2 R/\sqrt{n}$. We formalize these relationships in Proposition 1 (the proofs of the technical results in the paper are provided in Appendix D).

PROPOSITION 1. *Suppose package locations follow a homogeneous two-dimensional point process with rate γ such that $n=\gamma R^2$. Under the proposed zoning policy, the expected pickup and drop-off distance for each passenger ride request is*

$$L_2 = \frac{\sqrt{2}}{3}\alpha R + \frac{\sqrt{\pi}}{2\sqrt{2}}\frac{R}{\sqrt{n}} = \alpha_2 R + c_2\frac{R}{\sqrt{n}},$$

where $\alpha_2 = \frac{\sqrt{2}}{3}\alpha$ and $c_2 = \frac{\sqrt{\pi}}{2\sqrt{2}}$.

The inverse square root function in the expected drop-off distance reveals the *complementarity* between package delivery and on-demand rides: a greater number of package delivery locations

makes it easier to connect the passenger drop-off location and the package delivery route. In other words, when the packages are more densely distributed, the driver can traverse ride locations and delivery locations more efficiently.

(III) Lastly, we use L_3 to denote the expected travel distance between a pair of zones when no passenger ride requests are accepted, which we interpret as the inter-zone travel distance for package delivery. We approximate L_3 by the expected distance between the centroids of a pair of zones, $c_3R = \sqrt{2}R/2$ when $N = 4$.[6] Here, L_3 is independent of the visiting sequence of zones because the four zones are symmetric, but this relationship may not hold for $N > 4$, which we will discuss in the next section.

Based on the presented distance components, we are ready to define the expected revenue rate function. The expected total revenue of the four-zone policy, including the revenue from package delivery and passenger rides, is given by

$$nr_p + \sum_{i=1}^{3}(1 - \exp(-(4-i)\alpha^2\lambda t_0/4))r_t L_1, \tag{2}$$

where the first term captures the expected package delivery revenue and the second term represents the expected ride-hailing revenue.

The total expected delivery time consists of three parts. First, the expected within-zone package delivery time is incurred because a delivery path is formed to visit all packages in each zone, and we need to estimate the shortest Hamiltonian path length (instead of a TSP tour length). To this end, we adopt the strip strategy developed by Daganzo (1984), which suggests that the path distance can be approximated by $\beta(k-1)/\sqrt{\delta}$ for k locations, where δ is the package location (point) density. It follows that the within-zone delivery distance for each zone is

$$\beta\frac{n/N - 1}{\sqrt{n/A}} = \beta\left(\frac{\sqrt{n}}{N} - \frac{1}{\sqrt{n}}\right)R,$$

and the total within-zone delivery time across all zones is (note $R = \sqrt{A}$)

$$N \times \beta\left(\frac{\sqrt{n}}{N} - \frac{1}{\sqrt{n}}\right)\frac{R}{v} = \beta\left(\sqrt{n} - \frac{N}{\sqrt{n}}\right)\frac{R}{v} = \beta\left(\sqrt{n} - \frac{4}{\sqrt{n}}\right)\frac{R}{v} \quad \text{when } N = 4.$$

Second, the expected travel time due to serving passenger rides can be computed by $\sum_{i=1}^{3}(1 - \exp(-(4-i)\alpha^2\lambda t_0/4))(L_1 + L_2)/v$. Third, the expected total inter-zone travel time without carrying

[6] Upon finishing all delivery jobs in a subregion, the driver may choose the nearest location from the neighboring zone to start the new delivery trip (given no passenger rides are accepted). In this case, L_3 can be shorter than c_3R. However, the resulting ending location may not be optimal for connecting with subsequent trips in other zones. Instead of finding the optimal "connection" policy, we assume the driver picks the starting location randomly so L_3 can be approximated by the expected distance between a random point and another random point from its neighboring zone.

passengers is $\sum_{i=1}^{3}\exp(-(4-i)\alpha^2\lambda t_0/4)L_3/v$. Combining these parts, we derive the expected revenue rate under the four-zone policy, denoted by $R^{FZ}(n,\lambda)$, as

$$v\frac{nr_p+\sum_{i=1}^{3}(1-\exp(-(4-i)\alpha^2\lambda t_0/4))r_tL_1}{\beta\sqrt{n}R-4\beta R/\sqrt{n}+\sum_{i=1}^{3}(1-\exp(-(4-i)\alpha^2\lambda t_0/4))(L_1+L_2)+\sum_{i=1}^{3}\exp(-(4-i)\alpha^2\lambda t_0/4)L_3}.$$

For ease of notation, we define $\psi_0(\lambda)=\sum_{i=1}^{3}(1-\exp(-(4-i)\alpha^2\lambda t_0/4))$, then the expected revenue rate can be simplified as

$$R^{FZ}(n,\lambda)=v\frac{nr_p+\psi_0(\lambda)r_tc_1(\alpha)R}{\beta\sqrt{n}R-4\beta R/\sqrt{n}+\psi_0(\lambda)((c_1(\alpha)+\alpha_2-c_3)R+c_2\frac{R}{\sqrt{n}})+3c_3R}. \tag{3}$$

We first analyze the properties of $R^{FZ}(n,\lambda)$ and delineate its relationship with n and λ.

PROPOSITION 2. *$R^{FZ}(n,\lambda)$ satisfies*

(I) *For any fixed package number $n>0$, $R^{FZ}(n,\lambda)$ is either increasing or decreasing in λ. Moreover, there exist two constants $\tilde{n}_1$ and $\tilde{n}_2$ such that $R^{FZ}(n,\lambda)$ is increasing in λ when $n\le\tilde{n}_1$ and decreasing in λ when $n\ge\tilde{n}_2$.*

(II) *For any fixed $\lambda>0$, there exists a threshold $\bar{n}(\lambda)$ such that $R^{FZ}(n,\lambda)$ is increasing in n when $n\ge\bar{n}(\lambda)$.*

When n is small, the expected revenue rate from package delivery is low, so a higher passenger arrival rate boosts the chance of earning from on-demand rides and the overall revenue rate. On the other hand, when n is large, package delivery is more profitable than on-demand rides, because the latter incurs additional pickup/drop-off distance that drags down the revenue performance. Observe that the expected revenue rate from on-demand rides is bounded by a constant, whereas package delivery achieves an expected revenue rate that scales with $\sqrt{n}$. Proposition 2 also implies that $R^{FZ}(n,\lambda)$ is not necessarily monotonic in n. This can be attributed to the term $\beta\sqrt{nA}+\psi_0(\lambda)c_2R/\sqrt{n}$ in the denominator of the revenue rate function, where $\psi_0(\lambda)c_2R/\sqrt{n}$ captures the synergy between package delivery and ride-hailing. When n is small and the package delivery revenue is negligible compared to on-demand rides, the synergy effect plays a major role in driving the revenue; when n is larger, the TSP travel distance $\beta\sqrt{nA}$ starts to overshadow the synergy effect, causing the revenue to decline. As n becomes large enough, the expected revenue rate is mainly driven by package delivery and scales with $\sqrt{n}$.

Next, we compare the expected revenue rate of the four-zone policy with the pure package delivery policy. We characterize the conditions under which the four-zone policy yields a higher revenue rate.

PROPOSITION 3. *The four-zone policy can outperform the pure package delivery policy for some (n,λ) if and only if $\psi_0(\lambda)(\beta r_tc_1(\alpha)R-r_pc_2)+4\beta r_pR>0$ and*

$$n<\left[\frac{\psi_0(\lambda)(\beta Rr_tc_1(\alpha)-r_pc_2)+4\beta r_pR}{r_p(\psi_0(\lambda)(c_1(\alpha)+\alpha_2-c_3)+3c_3)}\right]^2.$$

Proposition 3 implies that coordinating the two types of jobs (using the four-zone policy) is beneficial when the package volume is low, given that the passenger ride is relatively profitable compared to package delivery. The intuition is that package delivery enjoys economies of scale as the number of delivery locations increases. The additional revenue from passenger rides can not efficiently compensate for the pickup and drop-off distance – the insertion of passenger rides disrupts the delivery route and causes a downgrade in its efficiency. This downgrade becomes more substantial as n increases. In other words, the coordination policy should give way to pure package delivery when the package demand is sufficiently high to sustain its own operations. Moreover, as suggested by Proposition 3, the threshold function for n increases in the revenue rate ratio r_t/r_p and λ given a reasonably large r_t (recall $\psi_0(\lambda)$ is increasing in λ). We illustrate which policy (four-zone or pure delivery) yields a higher revenue rate for different combinations of n and λ in Figure 2, where α is fixed to 0.25. In Appendix E, we further examine the impact of α on the choice between the four-zone policy and the pure delivery policy and similar insights hold.

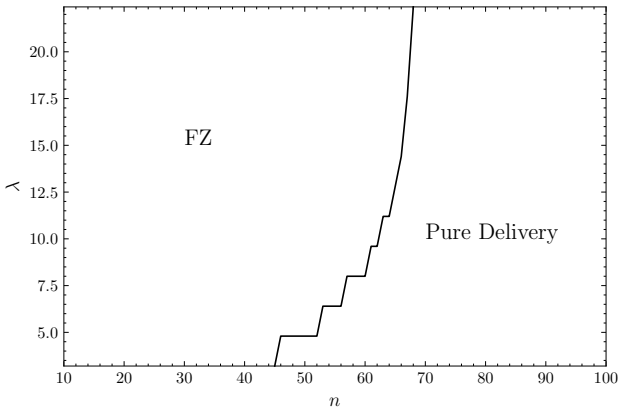


Figure 2 Four-zone policy versus pure delivery policy in different parameter regimes of (n, λ) ($R = 1$, $v = 1$, $\alpha = 0.25$, $r_t/r_p = 8$).

One major restriction of the four-zone policy is that the maximum number of on-demand passenger rides that the driver can take is three, which indicates limited coordination opportunities. To boost coordination, one may increase the number of zones so the driver can take up more passenger rides when delivering packages. Under the zoning policy, the driver can only serve on-demand rides after finishing all package deliveries from a zone, so more zones imply fewer packages in each zone, and the driver can open up more ride opportunities. In the extreme case, the driver can serve a ride immediately after a package delivery, which will be discussed in the next section.

4. General Zoning Policy: Approximation and Analysis

In this section, we study the general zoning-based coordination policy that determines the optimal number of zones to maximize revenue. A higher number of zones enables the driver to serve more on-demand rides and improve revenues, but the resulting increased inter-zone travel distance may downgrade the logistics efficiency of package delivery. In contrast, a smaller number of zones facilitates package consolidation and reduces routing deviation (from the TSP solution) at the expense of more limited on-demand ride opportunities. In other words, the number of zones can be viewed as a lever to balance TSP routing efficiency and ride-hailing flexibility in different operating regimes.

4.1. Expected Revenue Rate

For the general zoning policy with N zones, the probability of taking up a passenger ride request upon finishing package deliveries in the i-th visited zone is

$$\mathbb{P}(\text{accept a passenger ride request in the } i^{th} \text{ visited zone}) = 1 - \exp(-(N-i)\alpha^2\lambda t_0/N).$$

Similar to the four-zone policy, the expected total travel distance of the general zoning policy comprises three parts: (1) the expected passenger ride distance $L_1 \approx c_1(\alpha)R$; (2) the expected pickup and drop-off distance $L_2 = \alpha_2 R + c_2\frac{R}{\sqrt{n}}$; (3) the expected inter-zone delivery distance $L_3(i)$ is incurred when the driver travels from a zone to the nearest unvisited zone (without taking a ride) for package delivery. Unlike in the four-zone policy, the inter-zone delivery distance under the general zoning policy depends on the number of visited zones due to the lack of symmetry, which we will discuss in more detail later.

Summing up the above distances and the TSP delivery distance, the expected total travel distance under the general zoning policy, denoted by L^{GZ}_{total}, is

$$\beta\sqrt{n}R - \beta NR/\sqrt{n} + \sum_{i=1}^{N-1}(1-\exp(-\alpha^2\lambda t_0(N-i)/N))(L_1+L_2) + \sum_{i=1}^{N-1}\exp(-\alpha^2\lambda t_0(N-i)/N)L_3(i).$$

Similar to the four-zone policy, the expected total revenue from serving on-demand rides can be computed as

$$\sum_{i=1}^{N-1}(1-\exp(-\alpha^2\lambda t_0(N-i)/N))r_t L_1.$$

Therefore, the expected revenue rate of the general zoning policy, denoted by $R^{GZ}(N)$, is

$$v\frac{nr_p + \sum_{i=1}^{N-1}(1-\exp(-\alpha^2\lambda t_0(N-i)/N))r_t L_1}{(\beta\sqrt{n} - \beta N/\sqrt{n})R + \sum_{i=1}^{N-1}\left[(1-\exp(-\alpha^2\lambda t_0(N-i)/N))(L_1+L_2) + \exp(-\alpha^2\lambda t_0(N-i)/N)L_3(i)\right]}, \tag{4}$$

where we omit the dependence of $R^{GZ}(N)$ on system parameters n, λ to stress N as the main decision variable. In the numerator, the first term captures the expected package delivery revenue and the

second term is the expected ride-hailing revenue. Our goal is to find the optimal number of zones to maximize the expected revenue rate, i.e., $\max_{N=k^2,1\le k\le\sqrt{n}} R^{GZ}(N)$. Note that when $N=1$, the zoning policy reduces to the pure delivery policy.

4.2. Approximation Function for the Inter-Zone Delivery Distance $L_3(i)$

The key analytical difficulty of assessing the general zoning policy lies in the inter-zone delivery distance that is a function of the number of remaining zones to be visited. Formally, the expected inter-zone delivery distance under the zoning-based coordination policy with N zones $\{\mathcal{Z}_1,\dots,\mathcal{Z}_n\}$ is

$$L_3(i)=\mathbb{E}_{\mathcal{R}}\left[d(\mathcal{Z}_i,\mathcal{Z}_{-\mathcal{R},i})\right],\quad 1\le i\le N-1, \tag{5}$$

where $\mathcal{R}=\{\mathcal{Z}_1,\dots,\mathcal{Z}_i\}$ is the random visiting sequence of i zones following the coordination policy, $\mathcal{Z}_{-\mathcal{R},i}$ indicates the unvisited zone (i.e., $\mathcal{Z}_{-\mathcal{R},i}\notin\mathcal{R}$) that is closest to $\mathcal{Z}_i$ (when there is a tie in terms of the closest distance, we assume it is broken randomly), and $d(\mathcal{Z}_i,\mathcal{Z}_{-\mathcal{R},i})$ is the distance between a random package location in $\mathcal{Z}_i$ and a random package location in $\mathcal{Z}_{-\mathcal{R},i}$.

As the number of unvisited zones goes up (down), the probability that there is an unvisited zone close to the driver increases (decreases). Therefore, it is intuitive to conjecture that the expected inter-zone delivery distance $L_3(i)$ should be non-increasing in the number of unvisited zones $(N-i)$. For example, the inter-zone delivery distance is $\sqrt{2}R/\sqrt{N}$ (the distance between two neighboring zones) when there are $N-1$ zones to be visited, and it would go up as more zones are visited.

Moreover, this distance would pertain to the travel trajectory $\mathcal{R}$, which is stochastic due to random passenger ride requests. The visiting sequence can be viewed as a stochastic permutation of zones that depends on the delivery policy. Figure 3 presents two possible travel trajectories that cover 6 out of 25 zones (for ease of presentation, we omit the exact visiting sequence). Evaluating the expectation with respect to all the permutations is challenging, so we propose an approximation function for it by limiting the distributional uncertainty around the trajectory.

First, if we assume the unvisited zones are uniformly distributed in the service region, then $L_3(i)$ can be approximated by $c_2\frac{R}{\sqrt{N-i}}$ where $c_2=\frac{\sqrt{\pi}}{2\sqrt{2}}$, which we show in Proposition 7 in Appendix D. The rationale is the same as the estimation of the expected dropoff distance when the driver has to travel from a random starting location to the nearest location out of $N-i$ uniformly distributed destinations. We note that the unvisited zones are more likely to follow a uniform distribution when the number of zones is large, and the boundary effect is minimal.[7]

In addition, to further motivate our approximation function, we develop an upper bounding scheme for $L_3(i)$ by analyzing the worst-case distribution of unvisited zones. Specifically, we propose to bound

[7] The finite service region indicates that the zones are not spatially "identical" regarding their location in the service region.

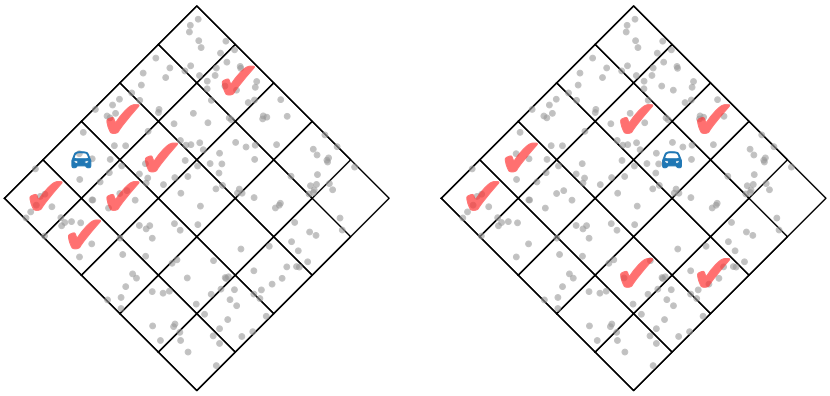


Figure 3 Two possible realizations of visited zones when $N = 25$ and $i = 6$: the visited zones are labeled by check marks. The driver's location is denoted by a label of car.

the expected inter-zone delivery distance by the solution to a max-min problem. To capture this travel distance, we consider placing $K = N - i + 1$ points $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_K\}$ within $\mathcal{A}$ (including the current location) such that $\sum_{k=1}^{K} d_k(\mathbf{x}_1, \ldots, \mathbf{x}_K)/K$ is maximal, where $d_k(\mathbf{x}_1, \ldots, \mathbf{x}_K)$ is the shortest distance between $\mathbf{x}_k$ and a point in $\mathcal{X} \setminus \mathbf{x}_k$. Equivalently, the bounding function $G_1(K)$ parameterized by K is the optimal objective function given by

$$G_1(K) = \max_{\mathcal{X}=\{\mathbf{x}_1,\ldots,\mathbf{x}_K\}\subset\mathcal{A}} \frac{1}{K} \sum_{k=1}^{K} \min_{\mathbf{x}\in\mathcal{X}\setminus\mathbf{x}_k} ||\mathbf{x} - \mathbf{x}_k||. \tag{6}$$

When $K = 2$, $G_1(K)$ is the longest distance between a pair of points in the service region, which equals $\sqrt{2}R$.

The idea behind the above formulation is to distribute the unvisited zones as sparsely as possible so the average distance between a zone and its nearest neighbor is maximal. Accordingly, the objective function of (6) serves as an upper bound for the expected inter-zone delivery distance, assuming every zone is equally likely to be the zone where the driver is currently located. Although the equal-probability distribution assumption may not be exactly satisfied under the proposed coordination policy due to the boundary effect, the objective function of (6) is still a good proxy for the upper bound, as we will show later in the numerical evaluation.

Although problem (6) can be formulated as a mixed integer linear program (MILP) by linearizing the inner minimization problem with binary variables, it can not be solved efficiently even for moderate values of K: the optimal solution can not be found for $K \geq 10$ on Gurobi with a time limit of six hours (see Appendix F.1 for the formulation and experiment setup). Next, we offer an alternative upper bounding problem by exploiting the connection between problem (6) and the region (square) covering problems (Alon and Kleitman, 1986).

Specifically, we map each point in $\mathcal{X}$ of problem (6) to the centroid of a square: $\mathbf{x}_i$ is associated with a square with inradius r_i (the inradius of a square equals half of its edge length), which is denoted by $\mathcal{S}(\mathbf{x}_i, r_i)$. Similar to problem (6), the goal is to identify K squares placed on $\mathcal{A}$ to maximize their average edge length while ensuring (i) their centroids lie within or on the boundary of $\mathcal{A}$ and (ii) they are nonoverlapping. Formally, the proposed region covering problem can be stated as follows

$$G_2(K) = \max_{\substack{\mathcal{X}=\{\mathbf{x}_1,\ldots,\mathbf{x}_K\}\subset\mathcal{A} \\ r_1,\ldots,r_K\geq 0}} \frac{2\sqrt{2}}{K}\sum_{k=1}^{K} r_k : \mathcal{S}(\mathbf{x}_i, r_i)\cap\mathcal{S}(\mathbf{x}_j, r_j) = \emptyset \quad \forall i\neq j, \tag{7}$$

where the objective function has a scaling coefficient of $2\sqrt{2}$ because the distance between the centroids of two neighboring squares is $\sqrt{2}$ times the sum of their inradiuses, and each square's inradius is counted twice. Note that this problem is not exactly a region covering problem because the squares do not necessarily cover $\mathcal{A}$: the non-overlapping constraint is softer than a coverage constraint (such relaxation is necessary because the general region covering problem is not always feasible). We show that $G_2(K)$ defined above is at least as large as $G_1(K)$ for $K \geq 2$.

Proposition 4. *$G_1(K) \leq G_2(K)$ for any $K \geq 2$.*

The key to establishing Proposition 4 is to show that the optimal solution to (6) constitutes a feasible solution to (7) by setting the radius r_k properly. Notably, $G_1(K)$ coincides with $G_2(K)$ for certain values of K, e.g., they are both equal to $\sqrt{2}R$ when $K = 2$. Due to Proposition 4, $G_2(K)$ can also serve as an upper bound of the expected inter-zone delivery distance under the equal-probability assumption. Compared to $G_1(K)$, $G_2(K)$ renders a more structured solution, which also facilitates computational evaluation and inspires an interpretable closed-form approximation. We show in Appendix F that an optimal solution $\mathcal{X}^*$ to problem (7) would have the outermost points of $\mathcal{X}^*$ to be on the boundary of $\mathcal{A}$ and include the vertices of $\mathcal{A}$ when $K \geq 4$. Based on this property, a relaxed problem (7) can be solved as a convex program efficiently, which yields an upper bound of $G_2(K)$. Interestingly, we find that the computed upper bounds follow very closely the function in the form of $\sqrt{2}R/(\sqrt{K}-1)$, as shown in Figure 4.

Observation 1. *$G_2(K)$ is approximately less than or equal to $\frac{\sqrt{2}R}{\sqrt{K}-1}$ for $K \geq 4$.*

The above observation is not surprising because the specified function corresponds to a simple rule of placing squares when $\sqrt{K} \in \mathbb{N}$. Specifically, we can design $\sqrt{K} \times \sqrt{K}$ identical squares and organize them into $\sqrt{K}$ rows, each with $\sqrt{K}$ squares. Each row is parallel to the edge of the service region, and the centroids of the first and last row are located on the two edges. Because the edge length is R, the inradius of the squares is $R/2(\sqrt{K}-1)$, which leads to the desired upper bounds. Figure 4 suggests that this functional form provides an excellent approximation for all values of K, incurring an average relative error of 2.6% for $K \in [6, 200]$.

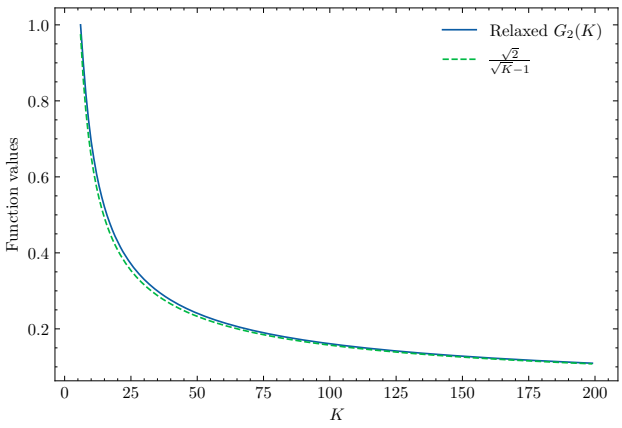


Figure 4 The optimal objective function value of relaxed problem (7) and the fitted function value.

Summarizing the above results, we have shown that $L_3(i)$ can be approximated by $c_2R/\sqrt{N-i}$ and upper bounded by $\sqrt{2}R/(\sqrt{N-i+1}-1)$ by imposing certain distributional assumptions about the unvisited zones. This motivates an approximation in the form of $c_3R/\sqrt{N-i+1}$, where c_3 can be set to satisfy $c_3 \leq \sqrt{2}$. This approximation is exact when $i=1$ (i.e., the vehicle has just visited one zone) because the expected inter-zone delivery distance between two neighboring zones is proportional to $R/\sqrt{N}$. Figure 5 compares the approximation proxy function, the upper bound function, and the simulated inter-zone travel distance for cases with small and large values of λ (the ride arrival rate affects the distribution of unvisited zones) when $N=100$. We observe that the proxy function preserves the convex increasing property of $L_3(i)$, and the average approximation error is 11.0% across the tested parameter regimes.

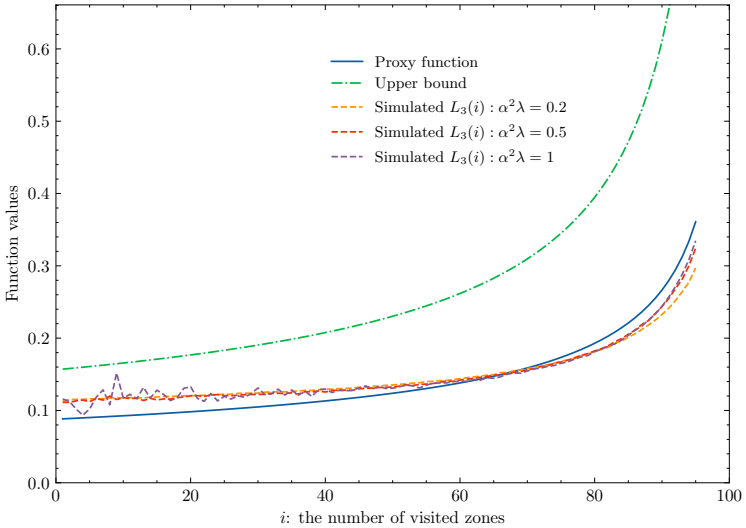


Figure 5 Simulated values of $L_3(i)$ versus the surrogate function $c_3R/\sqrt{N-i+1}$.

REMARK 3. Summing the inter-zone delivery distance based on the approximation, i.e., $\sum_{i=1}^{N-1} L_3(i) = \sqrt{2}R\sum_{i=1}^{N}(1/\sqrt{N-i+1}) = \Theta(\sqrt{N})$, implies that the maximum total additional travel distance (compared to TSP routing) incurred by zoning scales with $\sqrt{N}$. This is consistent with the result from Karp (1977) that shows a partitioning heuristic for solving TSP incurs a worst-case error of $O(\sqrt{N})$.

REMARK 4. By taking $c_3 = \sqrt{2}$, we observe that the approximation provides a lower bound of the revenue rate of the zoning policy, so it also serves as a lower bound to the optimal (and potentially complicated) coordination policy's performance.

Finally, to illustrate how R^{GZ} approximates the true revenue rate, we compare the derived revenue rate function (Equation (4)) and the simulated true revenue rate across all instances with $\lambda = \{0.5, 0.6, \ldots, 8\}, r_t = \{2, 2.1, \ldots, 10\}$, and $n = \{25, 36, \ldots, 100\}$ (36,936 instances in total). The average relative difference is 0.17% and the maximum relative difference is 2.4%; if we only consider cases with $\lambda > 1$, then the maximum relative difference is only 0.9%. These results indicate the robustness and reliability of our estimation of the revenue rate (using the analytical expression).

4.3. Zoning Optimization

In this section, we analyze the optimal zoning-based coordination policy, i.e., identify the optimal N to maximize the expected revenue rate. To this end, we first analyze two summations: $\sum_{i=1}^{N-1}(1-\exp(-\alpha^2\lambda t_0(N-i)/N)) = \sum_{i=1}^{N}(1-\exp(-\alpha^2\lambda t_0(N-i)/N))$ and $\sum_{i=1}^{N-1}\exp(-\alpha^2\lambda t_0(N-i)/N)/\sqrt{N+1-i}$, which appear in Equation (4). For ease of analysis, we use continuous integration to estimate the discrete summations[8]. Define

$$\psi_1(\lambda, N) = \int_0^N (1-\exp(-\hbar\lambda(N-x)/N))\, dx\,, \text{ and}$$
$$\psi_2(\lambda, N) = \int_1^N \exp(-\alpha^2\lambda t_0(N-x)/N)/\sqrt{N+1-x}\, dx\,. \tag{8}$$

Plugging in the distance functions $L_1 = c_1(\alpha)R$, $L_2 = \alpha_2 R + c_2 R/\sqrt{n}$, and $L_3 = c_3 R/\sqrt{N-i+1}$, the expected revenue rate under the zoning-based coordination policy with N zones can be written as (we divide R in both the numerator and denominator)

$$f_\lambda(N) := v\frac{nr_p/R + C_1 r_t \psi_1(\lambda, N)}{\beta\sqrt{n} - \beta N/\sqrt{n} + C_2\psi_1(\lambda, N) + C_3\psi_2(\lambda, N)}, \tag{9}$$

where we redefine the constants as

$$C_1 = c_1(\alpha),\ C_2 = c_1(\alpha) + \alpha_2 + c_2/\sqrt{n} = C_2^1(\alpha) + C_2^2/\sqrt{n},\ C_3 = c_3. \tag{10}$$

[8] For the discrete sum $\sum_{i=1}^{k-1} h(i)$, we can either use integration $\int_0^{k-1} h(x)dx$ or $\int_1^k h(x)dx$ to approximate. We take the former form of integration to approximate ψ_1 and the later one to approximate ψ_2 for ease of computation.

Here, $C_2^1(\alpha) = c_1(\alpha) + \alpha_2$ and $C_2^2 = c_2$. Accordingly, the optimal number of zones is given by

$$N^* = \underset{N=k^2, 1\le k\le\sqrt{n}}{\arg\max} f_\lambda(N). \tag{11}$$

Note that the restriction on N to be a square number is a technical assumption to simplify the geometrical partitioning process, and our analysis below applies to general values of N (partitioning algorithms are available for dividing the region into an arbitrary number of zones, see, e.g., Carlsson 2012). There are two special cases that can arise from this optimization problem: (i) $N^* = 1$: the optimal policy does not partition the service region, and pure package delivery is optimal; (ii) $N^* = n$: it is optimal to partition the service region into n zones, so each delivery location comprises of its own zone (on average); accordingly, the driver can switch between serving a package delivery and a passenger ride constantly. Specifically, when $N^* = n$, the driver can opt for passenger rides after finishing every package delivery, and the driver will travel to the nearest package location after completing a passenger ride. We refer to this policy as the *switching policy*. Under the switching policy, the driver enjoys higher flexibility in serving passenger rides, and the difference between the two types of jobs diminishes.

The behavior of $f_\lambda(N)$ hinges on $\psi_1(\lambda, N)$ and $\psi_2(\lambda, N)$, more specifically, their scaling pattern with respect to N. We first analyze the properties of these two functions and delineate their dependencies on N. We define $\hbar = t_0\alpha^2$ to simplify notations.

LEMMA 1. *The functions $\psi_1(\lambda, N)$ and $\psi_2(\lambda, N)$ satisfy*

$$\psi_1(\lambda, N) = \kappa N \text{ and } \psi_2(\lambda, N) = \varsigma_N\sqrt{N},$$

where $\kappa = 1 + (\exp(-\hbar\lambda) - 1)/(\hbar\lambda)$, and

$$\varsigma_N = \exp(\alpha^2\lambda t_0/N)\cdot\frac{\sqrt{\pi}}{\sqrt{\alpha^2\lambda t_0}}\left(erf(\sqrt{\alpha^2\lambda t_0}) - erf(\sqrt{\alpha^2\lambda t_0/N})\right),$$

and $erf(\cdot)$ is the error function defined by $erf(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}dt$.

Lemma 1 implies that $\psi_1(\lambda, N)$ scales with N linearly, for which the scaling coefficient depends on λ, t_0, and α. The scaling pattern of $\psi_2(\lambda, N)$ with respect to N is more nuanced, which makes it difficult to analyze the expected revenue rate function. Nevertheless, $\psi_2(\lambda, N)$ can be expressed as a function of $\sqrt{N}$, where the coefficient ς_N satisfies the following property.

LEMMA 2. *ς_N monotonically increases in N for $N \ge 4$, with values increasing from $\varsigma_4 = \exp(\alpha^2\lambda t_0/4)\cdot\frac{\sqrt{\pi}}{\sqrt{\alpha^2\lambda t_0}}(erf(\sqrt{\alpha^2\lambda t_0}) - erf(\frac{1}{2}\sqrt{\alpha^2\lambda t_0}))$ to $\varsigma_\infty := \frac{\sqrt{\pi}\cdot erf(\sqrt{\alpha^2\lambda t_0})}{\sqrt{\alpha^2\lambda t_0}}$.*

According to Lemma 2, we can show that $\psi_2(\lambda, N) \le \bar{\psi}_2(\lambda, N)$, where $\bar{\psi}_2(\lambda, N) := \varsigma_\infty\sqrt{N}$ and the relative gap between ς_N and ς_∞ can be bounded from the following corollary.

COROLLARY 1. *For all $N \geq N_0$ where $N_0 \geq 1$, it holds that*

$$\exp(\alpha^2 \lambda t_0/N_0)\left(1-\frac{erf(\sqrt{\alpha^2\lambda t_0/N_0})}{erf(\sqrt{\alpha^2\lambda t_0})}\right) \leq \frac{\varsigma_N}{\varsigma_\infty} \leq 1.$$

Corollary 1 implies that $\bar{\psi}_2(\lambda, N)$ can serve as a valid bound for $\psi_2(\lambda, N)$ and it is asymptotically tight for large N. This motivates a lower bound function for the expected revenue rate as

$$\underline{f}_\lambda(N) := v\frac{nr_p/R + C_1 r_t \psi_1(\lambda, N)}{\beta\sqrt{n} - \beta N/\sqrt{n} + C_2\psi_1(\lambda, N) + C_3\bar{\psi}_2(\lambda, N)} = \frac{v(nr_p/R + C_1 r_t \kappa N)}{\beta\sqrt{n} - \beta N/\sqrt{n} + C_2\kappa N + C_3\varsigma_\infty\sqrt{N}} \leq f_\lambda(N),$$

where we replace $\psi_2(\lambda, N)$ by $\bar{\psi}_2(\lambda, N)$. According to Corollary 1, $\underline{f}_\lambda(N)$ is an asymptotically tight bound for $f_\lambda(N)$ as N becomes large, i.e.,

$$\liminf_{N\to\infty} \frac{f_\lambda(N)}{\underline{f}_\lambda(N)} = 1.$$

We now analyze the structural properties of $\underline{f}_\lambda(N)$, which can shed light on the revenue performance of the optimal zoning policy.

PROPOSITION 5. *There exists a threshold N^Δ such that the expected revenue function $\underline{f}_\lambda(N)$ is decreasing for $N \leq N^\Delta$ and increasing for $N \geq N^\Delta$.*

Proposition 5 suggests that $\underline{f}_\lambda(N)$ is unimodal, and its maximum can only be taken on the boundary (while it has a unique minimum). The intuition is as follows. Note that $\underline{f}_\lambda(N)$ can be rewritten as

$$\underline{f}_\lambda(N) = \frac{r_1 + r_2 N}{d_1 + d_2 N + d_3\sqrt{N}}$$

where $r_2 N$ and $d_2 N + d_3\sqrt{N}$ denote the extra revenue and time cost associated with coordinating delivery jobs and on-demand ride requests, respectively. Both the additional revenue and the time cost increase with N because more zones imply more ride opportunities. Then, whether having more zones is beneficial hinges on the extra revenue rate from coordination, $r_2 N/(d_2 N + d_3\sqrt{N})$, relative to that from pure delivery (r_1/d_1). Note that $r_2 N/(d_2 N + d_3\sqrt{N})$ is also increasing in N due to the economics of scale in inter-zone traveling. It follows that when N is small, the additional revenue rate from coordination can be lower than that of pure delivery, and increasing N would drag down the total revenue rate. As N grows large enough, the extra revenue rate surpasses the revenue rate of pure delivery and pushes the overall revenue rate to edge higher. The unimodal behavior underscores the tension between delivery consolidation (that favors smaller N) and ride-hailing flexibility (that favors larger N).

Based on Proposition 5, we characterize the optimal $\underline{N}^* \in \{1, 4, \ldots, n\}$ that maximizes $\underline{f}_\lambda(N)$. For ease of notation, we let ω denote the ratio of ride-hailing and package delivery revenue rates, i.e., $\omega = r_t/r_p$.

THEOREM 1. *$\underline{N}^*$ satisfies (a) $\underline{N}^* \in \{1, 4, n\}$ and (b) there exist constants η_1, η_2, η_3, $n_1(\omega)$, $n_2(\omega)$, and $n_3(\omega)$ (that may also depend on λ) such that*

(I) *$\underline{N}^* = 1$ when both (I.a) either $\omega \le \eta_2$ or $n \ge n_2(\omega)$, and (I.b) either $\omega \le \eta_3$ or $n \ge n_3(\omega)$ are satisfied.*

(II) *$\underline{N}^* = 4$ when both (II.a) either $\omega \le \eta_1$ or $n \ge n_1(\omega)$, and (II.b) $\omega \ge \eta_3$ and $n \le n_3(\omega)$ are satisfied.*

(III) *$\underline{N}^* = n$ when (III.a) $\omega \ge \max\{\eta_1, \eta_2\}$ and (III.b) $n \le \min\{n_1(\omega), n_3(\omega)\}$.*

Theorem 1 indicates that the lower bound function $\underline{f}_\lambda(N)$ has a simple optimal solution structure, e.g., the zoning policy that attains the highest $\underline{f}_\lambda(N)$ is the pure delivery policy, four-zone policy, or the switching policy. Therefore, the optimal expected revenue rate under the zoning policy satisfies

$$R^*(n, \lambda) \ge \max\left\{\underline{f}_\lambda(1), \underline{f}_\lambda(4), \underline{f}_\lambda(n)\right\}.$$

Examining the form of the lower bound function further reveals that $R^*(n, \lambda) = \Omega(\sqrt{n})$ for any fixed λ. This implies that the optimal zoning policy would achieve (at least) asymptotically the same economies of scale as the pure delivery policy with respect to package density. The main reason lies in the fact that passenger rides are point-to-point trips that can not be pooled in a TSP tour. Nevertheless, the value of coordination can be manifested in the scaling constants that decide the marginal improvement of the revenue rate. Notably, the above result implies that the potential revenue improvement of co-modality against pure delivery is at least

$$\frac{R^*(n, \lambda)}{R^{PD}(n, \lambda)} \ge \frac{\underline{f}_\lambda(n)}{R^{PD}(n, \lambda)} = \frac{\beta(r_p + C_1 r_t \kappa R)}{r_p(C_2 \kappa \sqrt{n} + C_3 \varsigma_\infty)},$$

where the dependence on λ is implicitly captured in κ. We see that the relative revenue improvement tends to be higher for a smaller n, so the value of coordination can be very significant for crowdsourced drivers with a limited vehicle capacity.

Next, we investigate the properties of the optimal zoning policy as characterized by N^*. Because $\underline{f}_\lambda(N)$ is asymptotically tight, the expected revenue rate would follow $\underline{f}_\lambda(N)$ closely for large values of N, i.e., the increasing pattern in N. When N is relatively small, the gap between $f_\lambda(N)$ and $\underline{f}_\lambda(N)$ may render the behaviors of $f_\lambda(N)$ and $\underline{f}_\lambda(N)$ to diverge, e.g., N^* may be different than $\underline{N}^*$. Therefore, the optimal zoning policy may correspond to $N^* \notin \{1, 4, n\}$. Specifically, while the four-zone policy remains a promising choice, it can be optimal to partition the service region into slightly more zones. Nevertheless, such scenarios are relatively rare: we observe that only 1.18% of our tested cases yield $N^* \notin \{1, 4, n\}$ for all the combinations of $\lambda = \{0.5, 0.6, \ldots, 8\}, r_t = \{2, 2.1, \ldots, 10\}$, and $n = \{25, 36, \ldots, 100\}$ (36,936 instances in total). Notably, as shown in Corollary 2, adopting a zoning policy with $\underline{N}^*$ zones has a bounded performance ratio to the optimal zoning policy.

COROLLARY 2. *The expected revenue rate of the zoning policy with $\underline{N}^*$ zones satisfies*

$$f_\lambda(\underline{N}^*) \geq \frac{\varsigma_{N^*}}{\varsigma_\infty} f_\lambda(N^*).$$

Note that the above bound is not tight. Empirically, we observe that $f_\lambda(\underline{N}^*)$ achieves at least 92% of the optimal revenue rate. In fact, $\underline{N}^*$ coincides with N^* across 96% of the tested instances. Moreover, choosing the best zoning policy among the pure delivery policy, the four-zone policy, and the switching policy often reaps more than 99.7% of the optimal revenue rate. The summary statistics of the performance ratios on the test instances are presented in Table 1.

Table 1 Performance Ratio Evaluation of $f_\lambda(N^*)$ on the Tested Instances

$f_\lambda(\underline{N}^*)/f_\lambda(N^*)$		$\max\{f_\lambda(1), f_\lambda(4), f_\lambda(n)\}/f_\lambda(N^*)$	
Mean	Min	Mean	Min
98.911%	92.014%	99.999%	99.737%

The above results imply that although the properties from Theorem 1 may not hold for the optimal zoning policy, it is often good enough to restrict to three zoning policies defined by $N = \{1, 4, n\}$ or simply follow the zoning policy with $\underline{N}^*$ zones. Moreover, we can classify the optimal zoning policy into three classes: in addition to the pure delivery and switching policies, the zoning policy with a moderate number of zones ($N \notin \{1, n\}$) corresponds to the limited flexibility (LF) of coordination because the driver needs to wait for a number of package deliveries to finish before taking on-demand rides. Note that if there are practical constraints (e.g., work hours constraint) that restrict N to be below a certain threshold $\bar{N}$, then the switching policy could be replaced by a zoning policy with $\bar{N}$ zones.

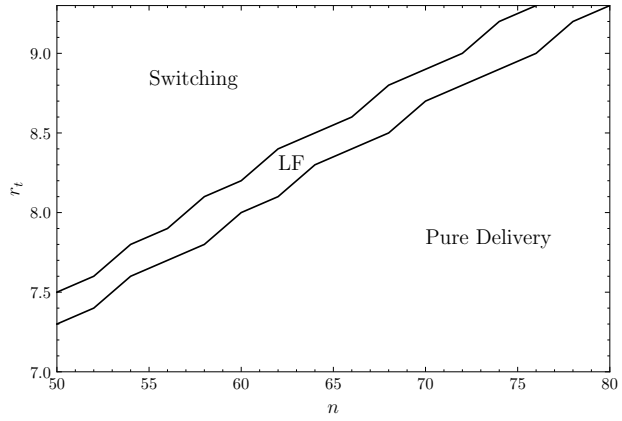


Figure 6 Optimal zoning policy in different parameter regimes of (n, r_t) ($R = 1$, $v = 1$, $\lambda = 2$, $\alpha = 0.25$, $r_p = 1$).

Figure 6 depicts the parameter regimes of (n, r_t) under which the optimal policy takes the form of pure delivery, limited flexibility, or switching. In this case, the optimal number of zones can be 4 or 9 in the limited-flexibility regime. We make several observations from Figure 6. First, when the profit from ride-hailing is low or the package density is high, it is optimal for the driver to focus on package delivery only because delivery routing is increasingly more efficient as n goes high, which echoes the result from Proposition 3. On the other hand, if ride-hailing is relatively lucrative and the package density is low, it is optimal to employ the flexible switching policy. In general, the added pickup and drop-off locations from ride requests cause relatively smaller disruptions to delivery routing as the package locations become sparser. When the package locations are sparse enough, the economies of scale from delivery routing are small and should give way to flexibility in responding to passenger ride requests. In fact, under the switching policy, the driver travels between package locations and passenger pickup/drop-off locations following a nearest-neighbor heuristic, which scales similarly to the optimal routing policy as the accepted ride requests grow (the tour length from the nearest-neighbor heuristic has a bounded performance ratio to the TSP tour length that scales with $\log(n)$, see Rosenkrantz et al. 1977). Lastly, when the ride profit and package density are moderate, offering limited flexibility is optimal. In such cases, the tension between delivery consolidation and ride flexibility is high, and we observe the resulting revenue performance is not very sensitive to N, so resorting to the zoning policy from $\{1, 4, n\}$ can be a good solution. In sum, the optimal number of zones lies in the tradeoff between routing efficiency gains from aggregated zones and the extra revenue rate from coordination that accounts for the routing deviation.

4.4. Performance Analysis

In this section, we consider a full-information coordination policy as a benchmark and upper-bounding scheme for the zoning policy. Specifically, under the full-information policy, the driver observes the origins and destinations of m passenger rides in the region and can decide how and when to transport these passengers; m is a parameter that is to be optimized. For this policy, the pickup time constraints on passenger rides are relaxed, allowing the driver to ignore the arrival sequence of passenger rides. Nevertheless, we respect the restriction that the driver cannot deliver packages during a passenger ride. Due to these relaxations, this full-information benchmark could establish an upper bound of the expected revenue rate of any coordination policy. Below, we characterize the performance ratio of the optimal zoning policy to the full-information benchmark: $R^*(n, \lambda)/R^\diamond(n, \lambda)$, where $R^\diamond(n, \lambda)$ is the expected revenue rate of the full-information coordination policy. This performance ratio quantifies the extent to which the zoning policy captures the potential of coordination.

THEOREM 2. *The performance ratio* $R^*(n,\lambda)/R^\diamond(n,\lambda)$ *satisfies that*

$$\frac{R^*(n,\lambda)}{R^\diamond(n,\lambda)} \geq \min\left\{\tau\sqrt{n}, 1-\frac{\xi r_t}{r_p\sqrt{n}+\xi r_t}\right\},$$

where $\tau = \frac{r_p}{\beta R r_t}$ *and* $\xi = (\beta - c_2/\sqrt{2})R$. *Furthermore, it holds that*

$$\lim_{\lambda\to\infty} \frac{R^*(n,\lambda)}{R^\diamond(n,\lambda)} \geq \min\left\{\max\left\{\tau\sqrt{n}, \frac{r_p + C_1 r_t R}{C_2 r_t R}\right\}, 1-\frac{\xi r_t}{r_p\sqrt{n}+\xi r_t}\right\},$$

where C_1 *and* C_2 *are defined in* (10).

Theorem 2 has several implications. First, the optimal zoning policy can reap the most value of coordination when the number of packages goes large (the package distribution becomes dense), due to the improvement in routing efficiency that benefits both the zoning policy and the full-information benchmark at the same scale. In the limit, it is close to optimal. Second, the positive constant ξ originates from the maximum additional revenue rate from coordination (in the full-information benchmark), which we have shown to take the form of $(\beta - c_2/\sqrt{2})vr_t/\beta \approx 0.63vr_t < vr_t$ (see Appendix D for more details). Note that vr_t is the revenue rate from serving on-demand rides; unless the vehicle always has a passenger on board, the driver can not earn the extra revenue rate of vr_t in full. Thus, the difference $(1-0.63)vr_t = 0.37vr_t$ indicates the minimum revenue loss from coordination. It is also important to note that the presented performance ratio is not tight, and we will examine the exact performance ratio numerically in the next section.

4.5. Numerical Study

In this subsection, we numerically evaluate the zoning-based coordination policy based on parameters calibrated using the yellow taxi data from New York City (NYC Taxi and Limousine Commission, 2023). Specifically, we consider a service region with area $A = 7.68 \times 7.68 = 59$ km^2, which is the size of the Manhattan borough. We set $v = 7.1$ mph $= 0.19$ km/minute (NYC Department of Transportation, 2018) and $\alpha = 0.25$. Based on the trip record, we set $\lambda \cdot t_0 = 100$ requests (Ata et al., 2019). According to the observed fare structure and statistics, the revenue rates from passenger rides and package delivery are $r_t = \$5.5$/km of ride time and $r_p = \$0.8$ per package ($20/hour for 25 packages per hour), respectively (Schneider, 2023; A Martínez, 2023). We also examine cases with a higher ride fare $r_t = \$6$/km, considering surge prices. Because the number of packages to be delivered varies according to the vehicle capacity, we test different values of n that can be as large as 60. More details about the parameter calibration are presented in Appendix G.

4.5.1. Scaling with λ. Figure 7 depicts how the expected revenue rates of the optimal zoning policy and the pure delivery policy scale with the passenger ride request rate, assuming $n = \{30, 50\}$. We observe that the pure delivery policy performs optimally only under very low arrival rates. Once the arrival rate surpasses a certain threshold, integrating on-demand rides and package delivery is more efficient. The optimal zoning policy can take the form of LF before transitioning to the switching policy, while the latter is more prevalent in the tested cases. Notably, the optimal zoning policy can lead to 20-30% higher revenue rate than the pure delivery policy when $n = 30$. We also observe that the threshold for λ increases with n: as the package volume grows, a higher arrival rate becomes necessary for the zoning policy to outperform the pure package delivery policy. Additionally, it is worth noting that most revenue rate improvement can be gained with a small ride arrival rate. As the arrival rate increases, the revenue rate of the switching policy grows at a slower pace as it reaches the limit of co-modality (the maximum revenue rate of the zoning policy is limited by the revenue rate of ride-hailing).

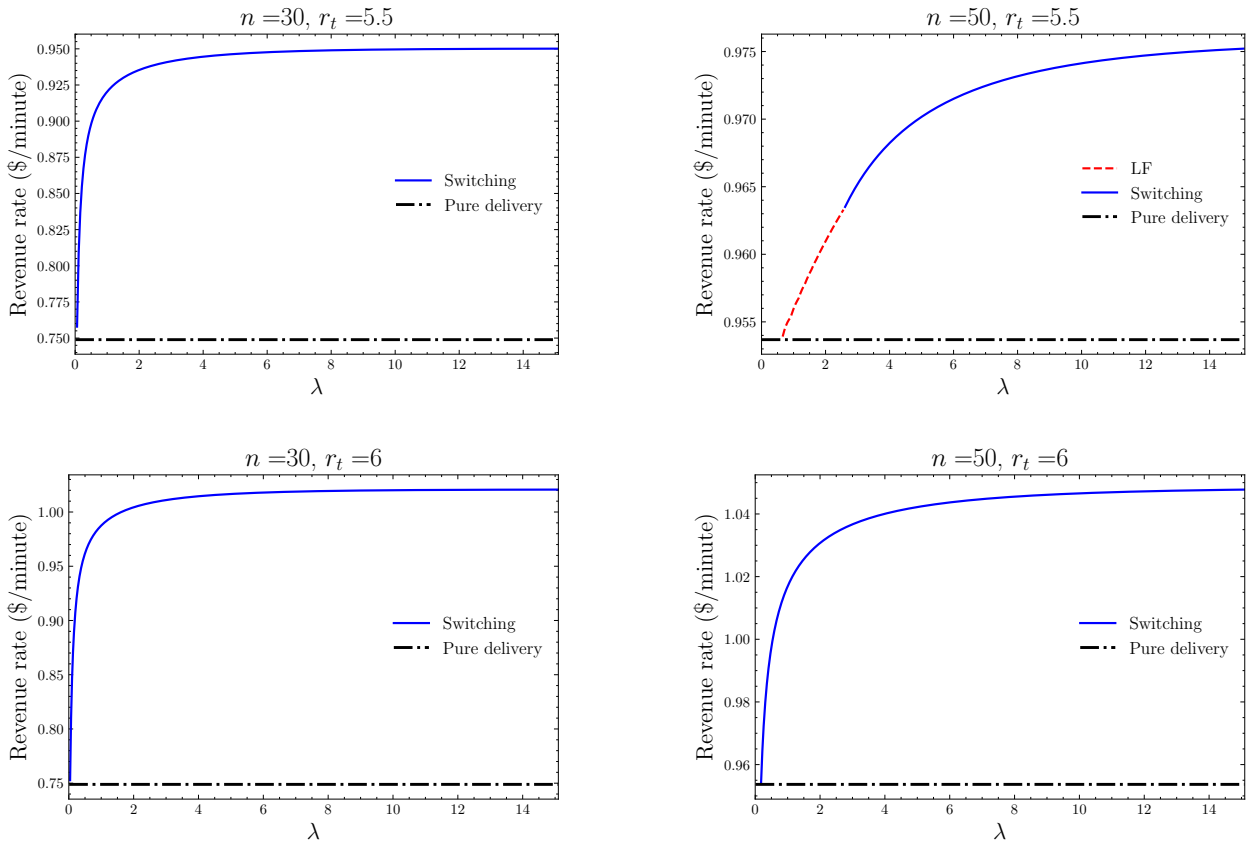


Figure 7 Expected revenue rate versus ride arrival rate for $n = \{30, 60\}$ and $r_t = \{5.5, 6\}$.

Similar observations hold when assessing the impact of α and t_0, both of which are positively related to the likelihood of matching with an eligible ride request. For example, as α increases, the

driver is open to ride requests from a larger area around the delivery location, so the effective ride request rate (that the driver can potentially take) increases. However, because increasing α and t_0 leads to longer pickup distance and lengthier matching process, setting proper values of α and t_0 requires trading off the passenger service constraint in terms of waiting time carefully.

4.5.2. Scaling with n. We evaluate the optimal revenue rate of the zoning policy with respect to the package volume in Figure 8. We observe that the switching policy yields the highest revenue rate in low-volume regimes, which echoes the result from Theorem 1. As the number of packages increases, the revenue rate of the zoning policy converges to that of pure delivery. This underscores that the gain from coordination is more pronounced when the package delivery volume is low or moderate, which is the case for crowdsourced private vehicles due to capacity limits. Specifically, the optimal zoning policy can achieve more than 60% improvement of revenue rate over the pure delivery policy when $n \leq 20$. Although more packages and ride requests both facilitate the coordination, *a higher package volume benefits delivery routing more than serving the ride requests*. Furthermore, we observe that the revenue rate increases faster with respect to the package volume under the pure delivery policy than under the switching policy. This can be attributed to the lower routing efficiency (due to deviation from the TSP route) of the switching policy. Nevertheless, their revenue all increases at the rate of $\sqrt{n}$ as n goes large, which is bounded by the TSP's economies of scale.

4.5.3. Performance Ratio. Finally, we analyze the performance ratio of the optimal zoning policy against the full-information policy described in Section 4.4. Figure 9 presents the performance ratio with respect to the ride arrival rate and the number of packages. We observe that the performance ratio is increasing in λ, edging over 75% when the ride arrival rate is greater than one. Interestingly, the performance ratio first decreases and then increases with n. Specifically, the optimal zoning policy can achieve more than 70% of the revenue rate from the full information policy when the package volume is less than 40, but this ratio could drop to 67% when the package volume varies between 50 and 70. We conjecture this is because the information advantage of knowing ride locations is greater for a moderate number of packages; when the packages are relatively sparse or dense, the disruption from taking a random ride is small. Consistent with Theorem 2, the performance ratio would increase and approach one for sufficiently large values of n.

5. Concluding Remarks

In this paper, we make an attempt to evaluate the potential of a co-modal transportation system that coordinates package delivery and on-demand rides, which is often deemed a promising solution for mitigating urban traffic congestion and emissions. We propose an analytical modeling framework and a zoning-based coordination policy, elucidating the synergies between these two types of jobs.

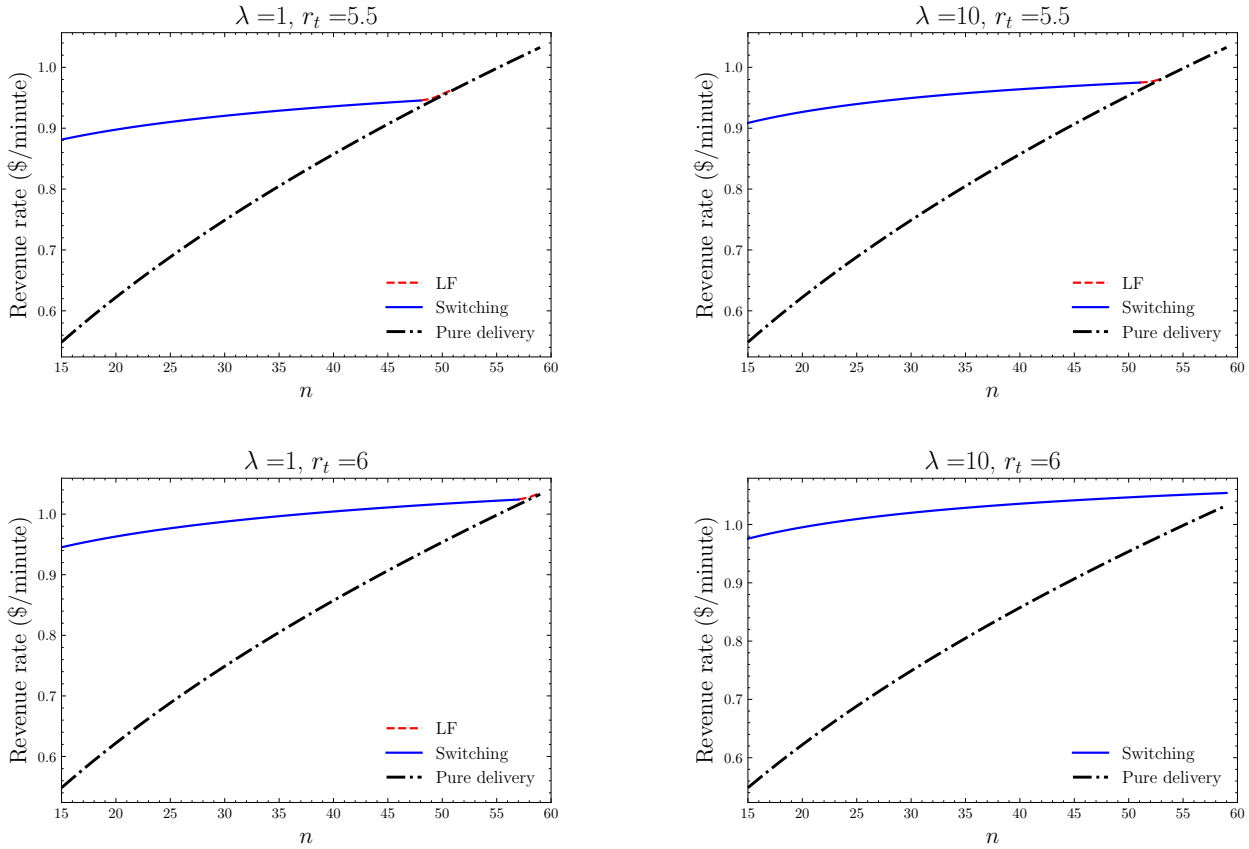


Figure 8 Expected revenue rate versus the number of packages for $\lambda = \{1, 10\}$ and $r_t = \{5.5, 6\}$.

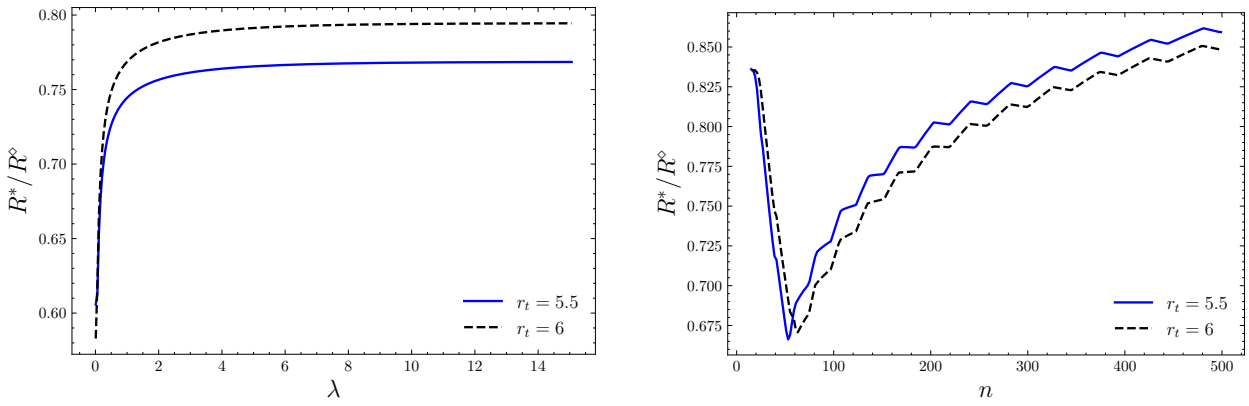


Figure 9 Performance ratio as a function of λ and n ($n = 30$ in the left panel and $\lambda = 10$ in the right panel)

Our results suggest that the revenue-maximizing zoning policy has a simple structure with three different levels of flexibility. Specifically, in the regimes with low package volumes, coordinating package delivery and on-demand rides is preferred over pure package delivery. When the package volume is lower than a threshold, the optimal coordination policy reduces to a flexible switching policy.

As the package volume grows, pure package delivery will emerge as the best policy due to its scaling efficiency. Our result supports the conjecture that co-modal systems may be more cost-efficiently implemented in low-demand areas with the intuition that building a standalone freight or mobility system would be too costly without economies of scale (Cleophas et al., 2019). However, the optimal way of coordination is more nuanced, as the optimal number of zones can vary across different regimes of package volumes and passenger ride arrival rates.

More work should be done to investigate the co-modal operations. One may generalize our single-driver setting to the case with multiple drivers. If the drivers are assigned to non-overlapping areas, we may separately analyze each driver following the proposed framework in this paper. However, in certain scenarios, assigning drivers to serve overlapping regions (Ledvina et al., 2022) may be more efficient, under which the optimal coordination policy is intriguing. One can also incorporate fixed pickup and delivery time into the calculation, which may weaken the economics of scale of package delivery and favor the coordination policy. Another important extension is to model the role of the platform and its interaction with drivers in co-modal transportation systems. Similar to ride-hailing platforms, co-modal operators can utilize multiple levers, such as surge price and dynamic wage payment, to influence crowdsourcing driver behaviors and the system throughput. Notably, the package assignment decisions have to account for package distributions in addition to imbalances between ride demand and driver supply, which may result in new optimization problems. We believe some variants of the zoning policy can be useful for coordinating drivers and assigning packages.

References

A Martínez, Danielle Kaye. 2023. Court in Manhattan to hear minimum wage case involving NYC, food delivery workers. https://www.npr.org/2023/08/03/1191792959/court-in-manhattan-to-hear-minimum-wage-case-involving-nyc-food-delivery-drivers. Accessed on August 4, 2023.

Afèche, Philipp, Zhe Liu, Costis Maglaras. 2023. Ride-hailing networks with strategic drivers: the impact of platform control capabilities on performance. *Manufacturing & Service Operations Management* **25**(5) 1890–1908.

Agarwal, Saharsh, Deepa Mani, Rahul Telang. 2023. The impact of ride-hailing services on congestion: Evidence from indian cities. *Manufacturing & Service Operations Management* **25**(3) 862–883.

Alnaggar, Aliaa, Fatma Gzara, James H Bookbinder. 2021. Crowdsourced delivery: A review of platforms and academic literature. *Omega* **98** 102139.

Alon, Noga, Daniel J. Kleitman. 1986. Covering a square by small perimeter rectangles. *Discrete & computational geometry* **1** 1–7.

Amazon Flex. 2023. Adjust your work, not your life. https://flex.amazon.ca/. Accessed on August 10, 2023.

Arslan, Alp M, Niels Agatz, Leo Kroon, Rob Zuidwijk. 2019. Crowdsourced delivery—a dynamic pickup and delivery problem with ad hoc drivers. *Transportation Science* **53**(1) 222–235.

Arslan, Okan, Rasit Abay. 2021. Data-driven vehicle routing in last mile delivery. *Technical Proceedings of the Amazon Last Mile Routing Research Challenge*. MIT Center for Transportation & Logistics.

Ata, Baris, Nasser Barjesteh, Sunil Kumar. 2019. Spatial pricing: An empirical analysis of taxi rides in new york city. *The University of Chicago Booth School of Busines Chicago, IL Working paper*.

Ata, Baris, Nasser Barjesteh, Sunil Kumar. 2020. Dynamic dispatch and centralized relocation of cars in ride-hailing platforms. Available at SSRN 3675888.

Azcuy, Irecis, Niels Agatz, Ricardo Giesen. 2021. Designing integrated urban delivery systems using public transport. *Transportation Research Part E: Logistics and Transportation Review* **156** 102525.

Bai, Jiaru, Kut C So, Christopher S Tang, Xiqun Chen, Hai Wang. 2019. Coordinating supply and demand on an on-demand service platform with impatient customers. *Manufacturing & Service Operations Management* **21**(3) 556–570.

Banerjee, Dipayan, Alan L Erera, Alejandro Toriello. 2022a. Fleet sizing and service region partitioning for same-day delivery systems. *Transportation Science* **56**(5) 1327–1347.

Banerjee, Siddhartha, Daniel Freund, Thodoris Lykouris. 2022b. Pricing and optimization in shared vehicle systems: An approximation framework. *Operations Research* **70**(3) 1783–1805.

Banerjee, Siddhartha, Carlos Riquelme, Ramesh Johari. 2015. Pricing in ride-share platforms: A queueing-theoretic approach. Available at SSRN 2568258.

Beardwood, Jillian, John H Halton, John Michael Hammersley. 1959. The shortest path through many points. *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 55. Cambridge University Press, 299–327.

Behrendt, Adam, Martin Savelsbergh, He Wang. 2023. A prescriptive machine learning method for courier scheduling on crowdsourced delivery platforms. *Transportation Science* **57**(4) 889–907.

Besbes, Omar, Gerard P Cachon. 2023. The fast and affordable delivery problem. Working paper.

Besbes, Omar, Francisco Castro, Ilan Lobel. 2021. Surge pricing and its spatial supply response. *Management Science* **67**(3) 1350–1367.

Bimpikis, Kostas, Ozan Candogan, Daniela Saban. 2019. Spatial pricing in ride-sharing networks. *Operations Research* **67**(3) 744–769.

Bontempo, Amanda P, Claudio B Cunha, Denise A Botter, Hugo TY Yoshizaki. 2014. Evaluating restrictions on the circulation of freight vehicles in brazilian cities. *Procedia-Social and Behavioral Sciences* **125** 275–283.

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

Cachon, Gerard P, Kaitlin M Daniels, Ruben Lobel. 2017. The role of surge pricing on a service platform with self-scheduling capacity. *Manufacturing & Service Operations Management* **19**(3) 368–384.

Cao, Junyu, Mariana Olvera-Cravioto, Zuo-Jun Shen. 2020. Last-mile shared delivery: A discrete sequential packing approach. *Mathematics of Operations Research* **45**(4) 1466–1497.

Carlsson, John Gunnar. 2012. Dividing a territory among several vehicles. *INFORMS Journal on Computing* **24**(4) 565–577.

Carlsson, John Gunnar, Sheng Liu, Nooshin Salari, Han Yu. 2024. Provably good region partitioning for on-time last-mile delivery. *Operations Research* **72**(1) 91–109.

Castro, Francisco, Hongyao Ma, Hamid Nazerzadeh, Chiwei Yan. 2021. Randomized FIFO mechanisms. arXiv preprint arXiv:2111.10706.

Chen, Jinwei, Zefang Zong, Yunlin Zhuang, Huan Yan, Depeng Jin, Yong Li. 2023a. Reinforcement learning for practical express systems with mixed deliveries and pickups. *ACM Transactions on Knowledge Discovery from Data* **17**(3) 1–19.

Chen, Mingliu, Ming Hu. 2023. Courier dispatch in on-demand delivery. *Management Science* (Article in Advance).

Chen, Q, Y Lei, Stefanus Jasin. 2023b. Real-time spatial-intertemporal dynamic pricing for balancing supply and demand in a ride-hailing network: near-optimal policies and the value of dynamic pricing. *Operations Research* (Article in Advance).

Cleophas, Catherine, Caitlin Cottrill, Jan Fabian Ehmke, Kevin Tierney. 2019. Collaborative urban transportation: Recent advances in theory and practice. *European Journal of Operational Research* **273**(3) 801–816.

Cochrane, Keith, Shoshanna Saxe, Matthew J Roorda, Amer Shalaby. 2017. Moving freight on public transit: Best practices, challenges, and opportunities. *International Journal of Sustainable Transportation* **11**(2) 120–132.

Daganzo, Carlos F. 1984. The length of tours in zones of different shapes. *Transportation Research Part B: Methodological* **18**(2) 135–145.

Dampier, Alex, Marin Marinov. 2015. A study of the feasibility and potential implementation of metro-based freight transportation in newcastle upon tyne. *Urban Rail Transit* **1** 164–182.

Fatehi, Soraya, Michael R Wagner. 2022. Crowdsourcing last-mile deliveries. *Manufacturing & Service Operations Management* **24**(2) 791–809.

Fehn, Fabian, Roman Engelhardt, Florian Dandl, Klaus Bogenberger, Fritz Busch. 2023. Integrating parcel deliveries into a ride-pooling service—an agent-based simulation study. *Transportation Research Part A: Policy and Practice* **169** 103580.

Feldman, Pnina, Andrew E Frazelle, Robert Swinney. 2023. Managing relationships between restaurants and food delivery platforms: Conflict, contracts, and coordination. *Management Science* **69**(2) 812–823.

Feng, Guiyun, Guangwen Kong, Zizhuo Wang. 2021. We are on the way: Analysis of on-demand ride-hailing systems. *Manufacturing & Service Operations Management* **23**(5) 1237–1256.

Furuhata, Masabumi, Maged Dessouky, Fernando Ordóñez, Marc-Etienne Brunet, Xiaoqing Wang, Sven Koenig. 2013. Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological* **57** 28–46.

Garg, Nikhil, Hamid Nazerzadeh. 2022. Driver surge pricing. *Management Science* **68**(5) 3219–3235.

Guda, Harish, Upender Subramanian. 2019. Your uber is arriving: Managing on-demand workers through surge pricing, forecast communication, and worker incentives. *Management Science* **65**(5) 1995–2014.

Hu, Bin, Ming Hu, Han Zhu. 2022. Surge pricing and two-sided temporal responses in ride hailing. *Manufacturing & Service Operations Management* **24**(1) 91–109.

Karp, Richard M. 1977. Probabilistic analysis of partitioning algorithms for the traveling-salesman problem in the plane. *Mathematics of operations research* **2**(3) 209–224.

Ledvina, Kirby, Hanzhang Qin, David Simchi-Levi, Yehua Wei. 2022. A new approach for vehicle routing with stochastic demand: Combining route assignment with process flexibility. *Operations Research* **70**(5) 2655–2673.

Li, Baoxiang, Dmitry Krushinsky, Hajo A Reijers, Tom Van Woensel. 2014. The share-a-ride problem: People and parcels sharing taxis. *European Journal of Operational Research* **238**(1) 31–40.

Li, Baoxiang, Dmitry Krushinsky, Tom Van Woensel, Hajo A Reijers. 2016. The share-a-ride problem with stochastic travel times and stochastic delivery locations. *Transportation Research Part C: Emerging Technologies* **67** 95–108.

Li, Feng, Xin Guo, Li Zhou, Jianjun Wu, Tongfei Li. 2022. A capacity matching model in a collaborative urban public transport system: integrating passenger and freight transportation. *International Journal of Production Research* **60**(20) 6303–6328.

Li, Zhujun, Amer Shalaby, Matthew J Roorda, Baohua Mao. 2021. Urban rail service design for collaborative passenger and freight transport. *Transportation Research Part E: Logistics and Transportation Review* **147** 102205.

Liu, Sheng, Long He, Zuo-Jun Max Shen. 2021. On-time last-mile delivery: Order assignment with travel-time predictors. *Management Science* **67**(7) 4095–4119.

Ma, Hongyao, Fei Fang, David C Parkes. 2022. Spatio-temporal pricing for ridesharing platforms. *Operations Research* **70**(2) 1025–1041.

Mao, Wenzheng, Liu Ming, Ying Rong, Christopher S Tang, Huan Zheng. 2019. Faster deliveries and smarter order assignments for an on-demand meal delivery platform. Available at SSRN 3469015.

Martin, Sebastien, Sean J Taylor, Julia Yan. 2024. Trading flexibility for adoption: Dynamic versus static walking in ridesharing. *Management Science* **Forthcoming**.

Masson, Renaud, Anna Trentini, Fabien Lehuédé, Nicolas Malhéné, Olivier Péton, Houda Tlahig. 2017. Optimization of a city logistics transportation system with mixed passengers and goods. *EURO Journal on Transportation and Logistics* **6**(1) 81–109.

Nguyen, Ngoc-Quang, Nguyen-Viet-Dung Nghiem, Phan-Thuan Do, Khac-Tuan LE, Minh-Son Nguyen, Naoto Mukai. 2015. People and parcels sharing a taxi for tokyo city. *Proceedings of the 6th International Symposium on Information and Communication Technology*. 90–97.

NYC Department of Transportation. 2018. New York City Mobility Report. https://www.nyc.gov/html/dot/downloads/pdf/mobility-report-2018-screen-optimized.pdf. Accessed on August 4, 2023.

NYC Taxi and Limousine Commission. 2023. TLC Trip Record Data. https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page. Accessed on August 4, 2023.

Ouyang, Yanfeng. 2007. Design of vehicle routing zones for large-scale distribution systems. *Transportation Research Part B: Methodological* **41**(10) 1079–1093.

Qi, Wei, Lefei Li, Sheng Liu, Zuo-Jun Max Shen. 2018. Shared mobility for last-mile delivery: Design, operational prescriptions, and environmental impact. *Manufacturing & Service Operations Management* **20**(4) 737–751.

Ronald, Nicole, Jie Yang, Russell G Thompson. 2016. Exploring co-modality using on-demand transport systems. *Transportation Research Procedia* **12** 203–212.

Rosenkrantz, Daniel J, Richard E Stearns, Philip M Lewis, II. 1977. An analysis of several heuristics for the traveling salesman problem. *SIAM journal on computing* **6**(3) 563–581.

Ross, Sheldon M. 2014. *Introduction to probability models*. Academic press.

Schneider, Todd. 2023. Taxi and Ridehailing Usage in New York City. https://toddwschneider.com/dashboards/nyc-taxi-ridehailing-uber-lyft-data/#notes. Accessed on August 4, 2023.

Taylor, Terry A. 2018. On-demand service platforms. *Manufacturing & Service Operations Management* **20**(4) 704–720.

Toyota. 2018. Toyota unveils e-Palette concept at CES 2018. https://www.toyota.ca/toyota/en/connect/2000/toyota-e-palette-concept-vehicle-ces-2018. Accessed on August 10, 2023.

Wang, Hai, Hai Yang. 2019. Ridesourcing systems: A framework and review. *Transportation Research Part B: Methodological* **129** 122–155.

Yan, Chiwei, Helin Zhu, Nikita Korolko, Dawn Woodard. 2020. Dynamic pricing and matching in ride-hailing platforms. *Naval Research Logistics (NRL)* **67**(8) 705–724.

Yildiz, Baris, Martin Savelsbergh. 2019. Provably high-quality solutions for the meal delivery routing problem. *Transportation Science* **53**(5) 1372–1388.

Zhu, Shengda, Michael GH Bell, Veronica Schulz, Michael Stokoe. 2023. Co-modality in city logistics: Sounds good, but how? *Transportation Research Part A: Policy and Practice* **168** 103578.

Appendix A: Example Partitions for Heterogeneous Package Demand and Ride Arrivals

The proposed zoning policy leverages simple equal-area partitions to balance package demand in the uniform case with $N = k^2$. For general cases with heterogeneous package demand (distribution) and any integer value of N, there are partitioning techniques to maintain the same package demand or delivery distance across zones. For example, let the package demand density function be $\sqrt{x_1^2 + x_2^2}$, where x_1 and x_2 are the x and y coordinates, respectively. Figure 10 presents an equal-demand partition for 7 zones using the workload-balancing zoning algorithm developed by Ouyang (2007). Furthermore, it is possible to balance the delivery workload and ride-request demand at the same time. To do so, we can employ the technique recently developed in Carlsson et al. (2024), where a partitioning algorithm that combines ham-sandwich cuts and fan cuts is developed to balance two measurable functions in 2D (such a partition is proved to always exist). As an example, Figure 11 illustrates the resulting partition that balances $\sqrt{x_1^2 + x_2^2}$ and $x_1^2 x_2^2$ at the same time.

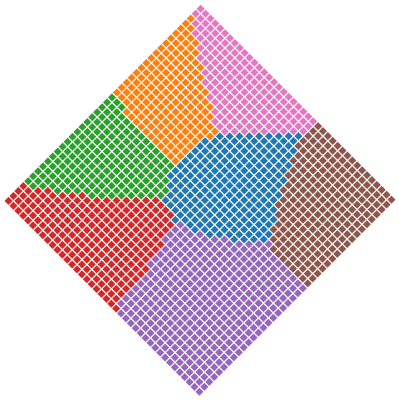


Figure 10 The partition that balances $\sqrt{x_1^2 + x_2^2}$ with $N = 7$ zones using the balancing algorithm from Ouyang (2007).

Appendix B: Relaxing the Restriction of Ride Destinations

In this section, we simulate the performance of the proposed zoning policy and the zoning policy that allows coming to previously visited zones (termed *flexible zoning policy*), of which the revenue performance and total delivery time are summarized in Table 2. Specifically, we set $n = 50$ and choose a relatively high passenger revenue rate $r_p = 5$ (the package revenue is normalized to be 1). Table 2 presents the revenue (delivery time) ratio between the flexible zoning policy and the proposed zoning policy, as denoted by R'/R (T'/T). We observe that the flexible zoning policy results in a lower revenue rate than the proposed zoning policy ($R'/R < 1$) when $\alpha^2 \lambda t_0 \in \{1, 2\}$ or $N \in \{20, 30\}$. The flexible zoning policy only outperforms the proposed zoning policy when both the arrival rate and the number of zones are large, and the improvement is less than 3%.

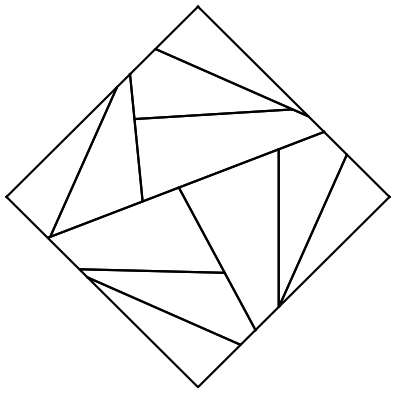


Figure 11 The partition that balances $\sqrt{x_1^2+x_2^2}$ and $x_1^2x_2^2$ at the same time with $N=9$ zones using the partitioning algorithm from Carlsson et al. (2024)

Table 2 Comparison between the flexible zoning policy and the proposed zoning policy

Effective arrival rate ($\alpha^2\lambda t_0$)	N	R'/R	T'/T
1	20	0.85	1.79
	30	0.87	1.92
	50	0.88	2.15
2	20	0.86	2.24
	30	0.92	2.42
	50	0.98	2.60
5	20	0.83	3.15
	30	0.92	3.45
	50	1.03	3.83
10	20	0.82	3.13
	30	0.91	3.55
	50	1.02	3.83
50	20	0.80	3.12
	30	0.89	3.38
	50	1.01	3.67

Besides, one practical limitation of the flexible zoning policy is that the resulting total delivery time (including both package delivery and passenger ride time) tends to be much longer than that of the proposed zoning policy. Table 2 shows that the flexible zoning policy often incurs two times longer total delivery time than the proposed zoning policy. This translates to 6 or 8 more hours, which may not be feasible for crowdsourced drivers. Moreover, taking rides to any destination makes the performance more variable as the zoning policy is subject to more randomness from the on-demand ride arrivals. Specifically, we observe the total delivery time of the flexible zoning policy can jump to 10 times higher than that of the proposed zoning policy in certain runs of the simulation. This happens when the driver continuously comes to visited zones before delivering the leftover packages (note that the delivery process only ends when all packages/zones are visited). While it can be interesting to

develop an adaptive (or dynamic) flexible zoning policy that can manage the total delivery time well, the revenue potential may be more limited.

Appendix C: Additional Results for Computing L_1

First, we present the probability density function of the passenger pickup location in Lemma 3. Note that because the pickup location is affected by the package location and the pickup radius r, it turns out that the ride pickup locations are not uniformly distributed over the service region, as reflected by the expression of $\rho(x,y)$. Specifically, locations closer to the centroid have a greater likelihood of being chosen as the pickup point, in contrast to locations on the periphery of the service region.

LEMMA 3. *Assume the package delivery location is uniformly distributed over the service region. In the rotated coordinate plane (by 45 degrees clockwise) with the origin located at the centroid, the probability density function of the passenger ride pickup location is*

$$\rho(x,y)=\begin{cases}1/R^2, \ \textit{if } |x|,|y|\le(1-\alpha)R/2;\\ ((1+\alpha)R/2-|x|)/(\alpha R^3), \ \textit{if } |x|\le(1-\alpha)R/2 \ \textit{and } |y|>(1-\alpha)R/2;\\ ((1+\alpha)R/2-|y|)/(\alpha R^3), \ \textit{if } |x|>(1-\alpha)R/2 \ \textit{and } |y|\le(1-\alpha)R/2;\\ ((1+\alpha)R/2-|x|)\left((1+\alpha)R/2-|y|\right)/(\alpha^2R^4), \ \textit{if } |x|,|y|>(1-\alpha)R/2.\end{cases}$$

Based on Lemma 3, L_1 can be computed by integrating $\rho(x,y)$ and the uniform density function of the drop-off location. There is a caveat: the drop-off location is uniformly distributed over the unvisited zones, which depend on the pickup location. Therefore, the integral can become complicated to evaluate when the number of possible combinations of unvisited zones gets large. To obtain a more tractable expression, we estimate L_1 by dropping the zoning restriction, so the drop-off location is uniformly distributed over the service region. Figure 12 depicts the relationship between the estimated L_1 and α, suggesting a convex increasing curve with a small increment. The values of L_1 range from 0.66 to 0.70 as α increases from 0 to 0.5.

Next, we compare the approximation $L_1 \approx c_1(\alpha)R$ and the simulated average ride distance under the zoning policy. We set $R=1$, $t_0=10$ and design a discrete-event simulation where (i) the starting zone is randomly chosen and (ii) the on-demand ride requests follow a Poisson arrival process with uniformly distributed origins and destinations. We vary $N=\{4,9,16,36,64,100\}$, $\lambda=\{5,10,50\}$, and $\alpha=\{0.1,0.2,0.3,0.4,0.5\}$. For each scenario, we repeat the simulation for 20,000 runs and report the average value of the ride distance. Table 3 summarizes the average relative error of the approximation for L_1. Note that for $N=4$, we can evaluate L_1 exactly using Proposition 6 presented below, where we utilize the fact that the four zones are symmetric. Therefore, we report the comparison result with the theoretical value for $N=4$ in Table 3. We observe that the used approximation provides good estimates of the average ride distance under different conditions (the error is below 5% most of the time). Although the error is larger when $N=4$, it only induces a small difference in the revenue rate (relative difference under 0.5%) for the considered parameter regimes because the revenue contribution from on-demand rides is relatively low.

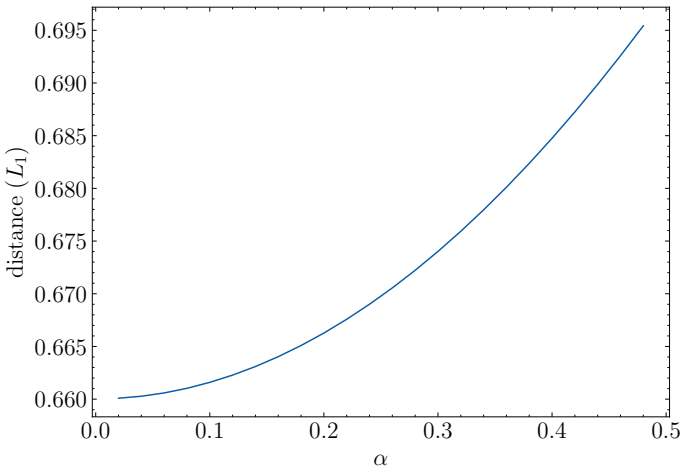


Figure 12 The relationship between L_1 and α when $R = 1$.

Table 3 Average Relative Error of Approximation L_1

λ	N	$\alpha = 0.1$	$\alpha = 0.2$	$\alpha = 0.3$	$\alpha = 0.4$	$\alpha = 0.5$
5	4	14.2%	14.0%	13.6%	13.0%	12.4%
	9	6.6%	5.7%	4.8%	3.6%	2.0%
	16	5.0%	2.9%	2.3%	1.0%	0.6%
	36	2.1%	1.2%	0.6%	1.2%	2.0%
	64	3.8%	1.4%	0.9%	1.5%	2.5%
	100	2.5%	1.1%	1.0%	1.7%	2.8%
10	4	14.2%	14.0%	13.6%	13.0%	12.4%
	9	6.5%	5.6%	4.7%	3.5%	1.8%
	16	4.1%	3.0%	2.4%	1.1%	0.5%
	36	1.4%	1.3%	0.4%	1.0%	2.1%
	64	2.5%	0.8%	0.7%	1.4%	2.4%
	100	1.7%	0.6%	0.9%	1.6%	2.8%
50	4	14.2%	14.0%	13.6%	13.0%	12.4%
	9	5.7%	5.3%	4.6%	3.5%	2.0%
	16	3.4%	3.2%	2.3%	1.0%	0.7%
	36	1.5%	1.1%	0.3%	1.1%	2.0%
	64	1.2%	0.4%	0.6%	1.4%	2.4%
	100	0.8%	0.3%	0.7%	1.6%	2.8%

PROPOSITION 6. *Suppose the package delivery location is uniformly distributed over the unvisited zones. Under the four-zone policy, the expected passenger ride distance is*

$$L_1^{FZ} = \iint_{(x_1,y_1)\in\mathcal{S}_O,(x_2,y_2)\in\mathcal{S}'} \frac{4}{3R^2}\rho_{R/2}(x_1,y_1)D((x_1,y_1),(x_2,y_2))d(x_1,y_1)d(x_2,y_2),$$

where $\rho_{R/2}(\cdot,\cdot)$ is defined in Lemma 3,

$$D((x_1,y_1),(x_2,y_2)) = \frac{\sqrt{2}}{2}|x_1 - x_2 - y_1 + y_2| + \frac{\sqrt{2}}{2}|y_1 - y_2 + x_1 - x_2|,$$

$\mathcal{S}_O = square((-(1+\alpha)R/4, -(1+\alpha)R/4), ((1+\alpha)R/4, (1+\alpha)R/4))$, *and* $\mathcal{S}' = square((-3R/4, -3R/4), (R/4, R/4)) \backslash square((-R/4, -R/4), (R/4, R/4))$.

C.1. Proofs for Appendix C

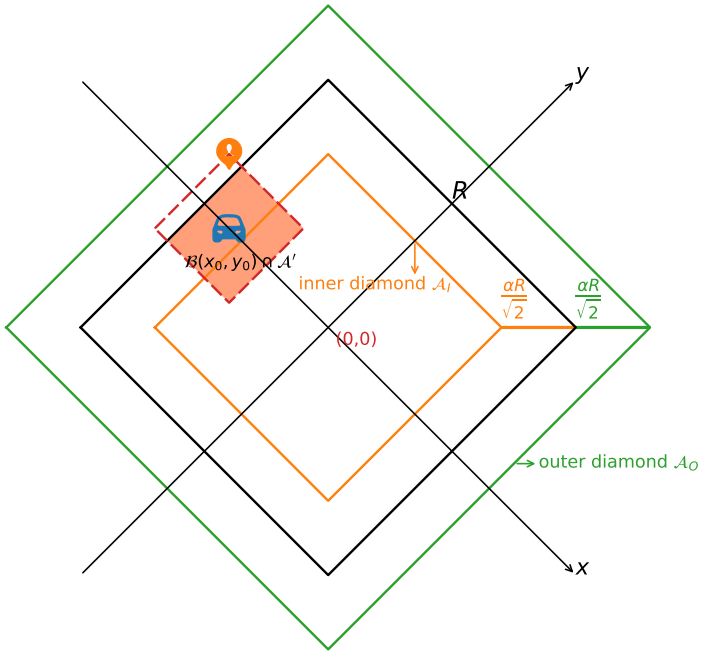


Figure 13 Pickup areas.

Proof of Lemma 3. We first characterize the probability distribution of the passenger pick-up location. Assume the service region centers at $(0,0)$ with edge length R, as shown in Figure 13. We construct the coordinates along the edge of diamonds (rotated by 45 degree clockwise). Since the driver will only accept the order from at most $r = \frac{\alpha R}{\sqrt{2}}$-distance away, the range of the entire pick-up area is larger than the original service region. It is a diamond with edge length $(1+\alpha)R$, which we call the outer diamond (green diamond in Figure 13), denoted as $\mathcal{A}_O = \text{square}((-(1+\alpha)R/2, -(1+\alpha)R/2), ((1+\alpha)R/2, (1+\alpha)R/2))$ where $\text{square}((x_1, y_1), (x_2, y_2)$ denotes the square defined by the diagonal $(x_1, y_1) - (x_2, y_2)$. Define the inner diamond $\mathcal{A}_I$ as the diamond with edge length $(1-\alpha)R$ (the orange diamond in Figure 13), i.e., $\mathcal{A}_I = \text{square}((-(1-\alpha)R/2, -(1-\alpha)R/2), ((1-\alpha)R/2, (1-\alpha)R/2))$. Also define $\mathcal{A}' = \text{square}((-R/2, -R/2), (R/2, R/2))$.

Under the assumption, the destination of the last delivery is uniformly distributed over the service region. For each point (x, y) in the outer diamond as the passenger pick-up location, the origin can be anywhere in the diamond with edge length αR and centroid (x, y), denoted as $\mathcal{B}(x, y)$. Therefore,

the probability density scales with the area $\mathcal{B}(x,y)\cap\mathcal{A}$. Next we compute the probability density function over the outer diamond. There are four scenarios:

(I) For each point (x,y) in $\mathcal{A}_I$, i.e., if $|x|,|y|\le(1-\alpha)R/2$, we have $|\mathcal{B}(x,y)\cap\mathcal{A}'|=\alpha^2R^2/2$;

(II) If $|x|\le(1-\alpha)R/2$ and $|y|>(1-\alpha)R/2$, we have $|\mathcal{B}(x,y)\cap\mathcal{A}'|=\alpha R((1+\alpha)R-y)$;

(III) If $|x|>(1-\alpha)R/2$ and $|y|\le(1-\alpha)R/2$, we have $|\mathcal{B}(x,y)\cap\mathcal{A}'|=\alpha R*((1+\alpha)R-x)$;

(IV) If $|x|>(1-\alpha)R/2$ and $|y|>(1-\alpha)R/2$, we have $|\mathcal{B}(x,y)\cap\mathcal{A}'|=((1+\alpha)R-x)((1+\alpha)R-y)$.

Suppose the density in the inner diamond is u. Then we integrate over the first quadrant, and the sum of the probability density should be 1/4:

$$\left((1-\alpha)\frac{R}{2}\right)^2u+2u\int_0^{(1-\alpha)R/2}\int_{(1-\alpha)R/2}^{(1+\alpha)R/2}\alpha R\cdot((1+\alpha)\frac{R}{2}-x)dxdy/(\alpha^2R^2)$$
$$+u\int_{(1-\alpha)R/2}^{(1+\alpha)R/2}\int_{(1-\alpha)R/2}^{(1+\alpha)R/2}\left((1+\alpha)\frac{R}{2}-x\right)\left((1+\alpha)\frac{R}{2}-y\right)dxdy/(\alpha^2R^2)=\frac{1}{4}.$$

It implies that

$$\left((1-\alpha)\frac{R}{2}\right)^2+\frac{(1-\alpha)\alpha R^2}{2}+\frac{1}{4}\alpha^2R^2=\frac{1}{4u}.$$

Therefore, we have

$$u=1/R^2.$$

Substituting u, we reach our conclusion that

$$\rho(x,y)=\begin{cases}1/R^2,\ \text{if } |x|,|y|\le(1-\alpha)R/2\\ ((1+\alpha)R/2-|x|)/(\alpha R^3),\ \text{if } |x|\le(1-\alpha)R/2 \text{ and } |y|>(1-\alpha)R/2\\ ((1+\alpha)R/2-|y|)/(\alpha R^3),\ \text{if } |x|>(1-\alpha)R/2 \text{ and } |y|\le(1-\alpha)R/2\\ ((1+\alpha)R/2-|x|)\left((1+\alpha)R/2-|y|\right)/(\alpha^2R^4),\ \text{if } |x|,|y|>(1-\alpha)R/2.\end{cases}$$

For any two points (x_1,y_1) and (x_2,y_2) in the constructed coordinate plane, the ℓ_1 distance in the original coordinate plane is

$$D((x_1,y_1),(x_2,y_2))=\frac{\sqrt{2}}{2}|x_1-x_2-y_1+y_2|+\frac{\sqrt{2}}{2}|y_1-y_2+x_1-x_2|.$$

Therefore, the expected passenger ride distance L_1 is

$$L_1=\iint_{(x_1,y_1)\in\mathcal{A}_O,(x_2,y_2)\in\mathcal{A}'}\frac{1}{R^2}\rho(x_1,y_1)D((x_1,y_1),(x_2,y_2))d(x_1,y_1)d(x_2,y_2).$$

□

Proof of Proposition 6. Pick a zone as the current zone (colored as orange) and assume its centroid is (0,0) (as shown in Figure 14). The delivery destination is uniformly distributed among the rest of zones ($\mathcal{S}'$). The edge length of that delivery zone in the FZ policy is $R/2$. Applying the result in Lemma 3, conditional on that the last delivery destination is uniformly distributed in that zone, the probability density function of that ride origin is

$$\rho_{R/2}(x,y)=\begin{cases}4/R^2,\ \text{if } |x|,|y|\le(1-\alpha)R/4\\ ((1+\alpha)R/4-|x|)/(\alpha(R/2)^3),\ \text{if } |x|\le(1-\alpha)R/4 \text{ and } |y|>(1-\alpha)R/4\\ ((1+\alpha)R/4-|y|)/(\alpha(R/2)^3),\ \text{if } |x|>(1-\alpha)R/4 \text{ and } |y|\le(1-\alpha)R/4\\ ((1+\alpha)R/4-|x|)\left((1+\alpha)R/4-|y|\right)/(\alpha^2(R/2)^4),\ \text{if } |x|,|y|>(1-\alpha)R/4.\end{cases}$$

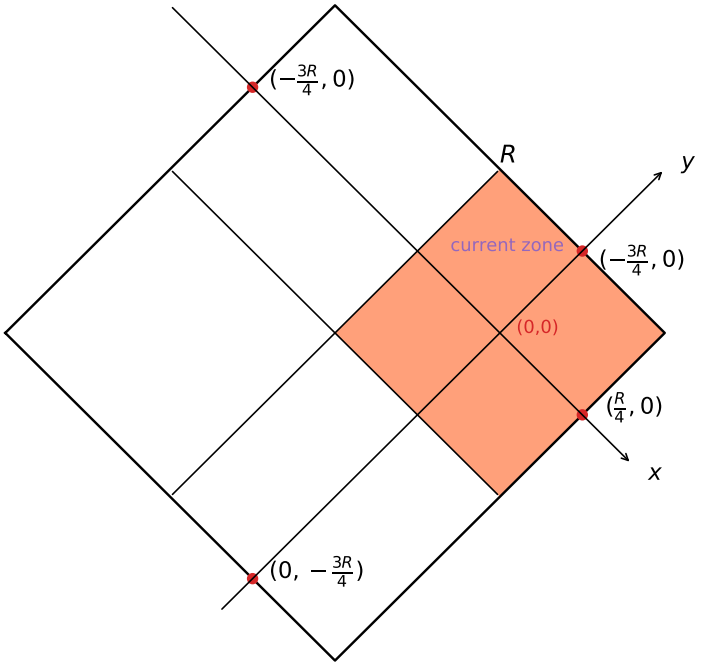


Figure 14 Four-zone pick-up areas. The orange zone is the current zone.

Suppose ϱ is the probability density function of the destination. That is, $\varrho = 1/((3/4)R^2)$ for all (x, y) belonging to $\mathcal{A}'$. Then the expected passenger ride distance L_1 is

$$L_1(K) = \iint_{(x_1,y_1)\in\mathcal{S}_O,(x_2,y_2)\in\mathcal{S}'} \frac{4}{3R^2}\rho(x_1,y_1)D((x_1,y_1),(x_2,y_2))d(x_1,y_1)d(x_2,y_2).$$

□

Appendix D: Proofs of Main Results

D.1. Proofs for Section 3

Proof of Proposition 1. We first calculate the first-part distance, i.e., the distance from the central to a uniformly distributed point in the diamond with edge length αR, as shown in Figure 15. The integration area is a diamond with edge length αR.

Then the expected pick-up distance equals

$$4\int_0^{\frac{\alpha R}{\sqrt{2}}}\int_0^{-x+\frac{\alpha R}{\sqrt{2}}}(x+y)\cdot\frac{1}{\alpha^2R^2}dydx = \frac{\sqrt{2}\alpha R}{3}.$$

Next we calculate the second-part distance, i.e., the distance from a uniformly distributed point to its nearest neighbor $D_{\min}$. We denote the fixed point as $w = (x_0, y_0)$ and define the set of neighbors that are within r distance $\mathcal{N}_w(r) = \{x_j \in \mathcal{A} \mid d(x, x_j) < r, j = 1, .., n\}$. Since the package location

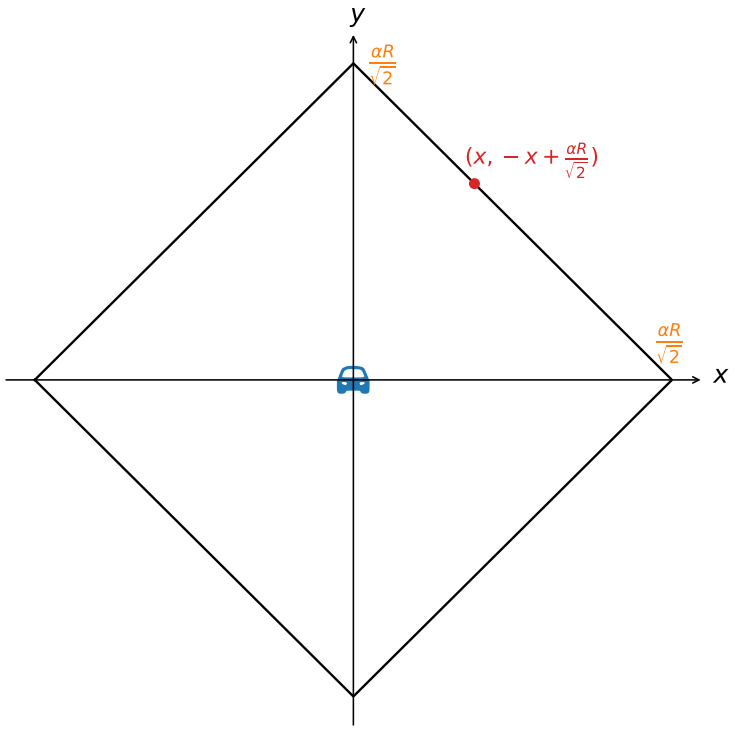


Figure 15 The region for picking up passengers.

follows a homogeneous spatial Poisson process with rate γ, the probability that the minimum distance is shorter than r equals the probability that there is at least one point in $\mathcal{N}_w(r)$, i.e.,

$$\mathbb{P}(D_{\min}(w) \le r) = 1 - \mathbb{P}(|\mathcal{N}_w(r)| = 0) = 1 - \exp(-2\gamma r^2)\,.$$

Therefore, the probability density for the distance to the nearest neighbor is

$$f_{D_{\min}}(r) = 4\gamma r \exp(-2\gamma r^2)\,.$$

Taking expectation with ignoring the boundary, we have

$$\mathbb{E}[D_{\min}] = \int_0^\infty 4\gamma r^2 \exp(-2\gamma r^2) dr = \frac{\sqrt{\pi}}{2\sqrt{2}\sqrt{\gamma}} = \frac{\sqrt{\pi}}{2\sqrt{2}\sqrt{n/R^2}} = \frac{R\sqrt{\pi}}{2\sqrt{2}\sqrt{n}}.$$

Thus, we reach our conclusion that

$$L_2 = \frac{\sqrt{2}}{3}\alpha R + \frac{\sqrt{\pi}}{2\sqrt{2}}\frac{R}{\sqrt{n}} = \alpha_2 R + c_2\frac{R}{\sqrt{n}},$$

where $\alpha_2 = \frac{\sqrt{2}}{3}\alpha$ and $c_2 = \frac{\sqrt{\pi}}{2\sqrt{2}}$. □

Proof of Proposition 2. (I) To see the relation between $R^{FZ}(n,\lambda)$ and λ, we take the derivative with respect to λ. According to the definition,

$$R^{FZ}(n,\lambda) = v\frac{nr_p/R + \psi_0(\lambda)r_t c_1(\alpha)}{\beta\sqrt{n} - 4\beta\sqrt{1/n} + \psi_0(\lambda)((c_1(\alpha) + \alpha_2 - c_3) + c_2\frac{1}{\sqrt{n}}) + 3c_3}\,, \tag{12}$$

where $\psi_0(\lambda) = \sum_{i=1}^{3}(1-\exp(-(4-i)\alpha^2\lambda t_0/4))$. Then both $\psi_0(\lambda)$ and $c_1(\alpha)$ increase in λ. Therefore, $R^{FZ}(n,\lambda)$ can be also written as

$$R^{FZ}(n,\lambda) = v\frac{nr_p/R + r_t h_1(\lambda)}{\beta\sqrt{n} - 4\beta\sqrt{1/n} + h_1(\lambda) + h_2(\lambda)},$$

where $h_1(\lambda) = \psi_0(\lambda)c_1(\alpha)$, $h_2(\lambda) = \psi_0(\lambda)((\alpha_2 - c_3) + c_2/\sqrt{n}) + 3c_3$.

Taking derivative with respect to λ, we have

$$\frac{\partial R^{FZ}(n,\lambda)}{\partial\lambda} = v\frac{r_t h_1'(\lambda)(\beta\sqrt{n} - 4\beta\sqrt{1/n} + h_1(\lambda) + h_2(\lambda)) - (nr_p/R + r_t h_1(\lambda))(h_1'(\lambda) + h_2'(\lambda))}{(\beta\sqrt{n} - 4\beta\sqrt{1/n} + h_1(\lambda) + h_2(\lambda))^2}.$$

Note that

$$\begin{aligned}
&r_t h_1'(\lambda)(\beta\sqrt{n} - 4\beta\sqrt{1/n} + h_1(\lambda) + h_2(\lambda)) - (nr_p/R + r_t h_1(\lambda))(h_1'(\lambda) + h_2'(\lambda))\\
=&r_t h_1'(\lambda)(\beta\sqrt{n} - 4\beta\sqrt{1/n} + h_2(\lambda)) - (nr_p/R + r_t h_1(\lambda))h_2'(\lambda) - nr_p h_1'(\lambda)/R\\
=&r_t h_1'(\lambda)(\beta\sqrt{n} - 4\beta\sqrt{1/n} + h_2(\lambda)) - nr_p/R(h_1'(\lambda) + h_2'(\lambda)) - r_t h_1(\lambda)h_2'(\lambda)\\
=&r_t c_1(\alpha)\psi_0'(\lambda)(\beta\sqrt{n} - 4\beta\sqrt{1/n} + \psi_0(\lambda)((\alpha_2 - c_3) + c_2/\sqrt{n}) + 3c_3)\\
&- (nr_p/R)(c_1(\alpha) + (\alpha_2 - c_3) + c_2/\sqrt{n})\psi_0'(\lambda) - r_t c_1(\alpha)\psi_0(\lambda)\psi_0'(\lambda)((\alpha_2 - c_3) + c_2/\sqrt{n})\\
=&\psi_0'(\lambda)\left(r_t c_1(\alpha)(\beta\sqrt{n} - 4\beta\sqrt{1/n} + 3c_3) - nr_p/R(c_1(\alpha) + (\alpha_2 - c_3) + c_2/\sqrt{n})\right).
\end{aligned}$$

Therefore, $\frac{\partial R^{FZ}(n,\lambda)}{\partial\lambda}$ is positive if and only if

$$\begin{aligned}
&r_t c_1(\alpha)(\beta\sqrt{n} - 4\beta\sqrt{1/n} + 3c_3) \geq nr_p/R(c_1(\alpha) + (\alpha_2 - c_3) + c_2/\sqrt{n})\\
\Leftrightarrow&(c_1(\alpha) + (\alpha_2 - c_3))\frac{r_p}{R}n + \left(c_2\frac{r_p}{R} - r_t c_1(\alpha)\beta(1 - 4/n)\right)\sqrt{n} - 3r_t c_1(\alpha)c_3 \leq 0.
\end{aligned}$$

Thus, we reach the conclusion that for any fixed $n > 0$, $R^M(n,\lambda)$ is either increasing or decreasing in λ. Note that

$$\begin{aligned}
&(c_1(\alpha) + (\alpha_2 - c_3))\frac{r_p}{R}n + \left(c_2\frac{r_p}{R} - r_t c_1(\alpha)\beta(1 - 4/n)\right)\sqrt{n} - 3r_t c_1(\alpha)c_3\\
\leq&(c_1(\alpha) + (\alpha_2 - c_3))\frac{r_p}{R}n + c_2\frac{r_p}{R}\sqrt{n} - 3r_t c_1(\alpha)c_3. \qquad (13)
\end{aligned}$$

Let

$$\tilde{n}_1 = \frac{-c_2\frac{r_p}{R} + \sqrt{(c_2\frac{r_p}{R})^2 + 12\,(c_1(\alpha) + (\alpha_2 - c_3))\,r_t c_1(\alpha)c_3}}{2\,(c_1(\alpha) + (\alpha_2 - c_3))}.$$

When $0 \leq n \leq \tilde{n}_1$, it holds that (13)$\leq 0$, which implies that the derivative is positive. Thus, $R^M(n,\lambda)$ is increasing in λ when $n \leq \tilde{n}_1$.

On the other hand,

$$\begin{aligned}
&(c_1(\alpha) + (\alpha_2 - c_3))\frac{r_p}{R}n + \left(c_2\frac{r_p}{R} - r_t c_1(\alpha)\beta(1 - 4/n)\right)\sqrt{n} - 3r_t c_1(\alpha)c_3\\
\geq&(c_1(\alpha) + (\alpha_2 - c_3))\frac{r_p}{R}n + \left(c_2\frac{r_p}{R} - r_t c_1(\alpha)\beta\right)\sqrt{n} - 3r_t c_1(\alpha)c_3. \qquad (14)
\end{aligned}$$

Let

$$\tilde{n}_2 = \frac{-\left(c_2\frac{r_p}{R} - r_t c_1(\alpha)\beta\right) + \sqrt{\left(\left(c_2\frac{r_p}{R} - r_t c_1(\alpha)\beta\right)\right)^2 + 12\left(c_1(\alpha) + (\alpha_2 - c_3)\right) r_t c_1(\alpha) c_3}}{2\left(c_1(\alpha) + (\alpha_2 - c_3)\right)}.$$

When $n \geq \tilde{n}_2$, it holds that (13)≥ 0, which implies that the derivative is negative. Thus, $R^M(n,\lambda)$ is decreasing in λ when $n \geq \tilde{n}_2$.

(II) A more general version of this result is proved in Theorem 1. □

Proof of Proposition 3. The four-zone policy yields a higher expected revenue rate than the pure delivery policy if and only if

$$\begin{aligned}
&\frac{nr_p + \psi_0(\lambda) r_t c_1(\alpha) R}{\beta\sqrt{nA} - 4\beta\sqrt{A/n} + \psi_0(\lambda)((c_1(\alpha) + \alpha_2 - c_3)R + c_2\frac{R}{\sqrt{n}}) + 3c_3 R} > \frac{\sqrt{n} r_p}{\beta R} \\
\Leftrightarrow\quad & \beta r_p R n + \beta\psi_0(\lambda) r_t c_1(\alpha) R^2 > \beta r_p R(n-4) + \psi_0(\lambda) r_p((c_1(\alpha) + \alpha_2 - c_3)\sqrt{n} + c_2)R + 3r_p c_3 R\sqrt{n} \\
\Leftrightarrow\quad & \beta\psi_0(\lambda) r_t c_1(\alpha) R > \psi_0(\lambda) r_p((c_1(\alpha) + \alpha_2 - c_3)\sqrt{n} + c_2) + 3r_p c_3\sqrt{n} - 4\beta r_p R \\
\Leftrightarrow\quad & \psi_0(\lambda)(\beta r_t c_1(\alpha) R - r_p c_2) > (\psi_0(\lambda)(c_1(\alpha) + \alpha_2 - c_3) + 3c_3) r_p\sqrt{n} - 4\beta r_p R.
\end{aligned}$$

Note that $\psi_0(\lambda)(c_1(\alpha) + \alpha_2 - c_3) + 3c_3 > 0$. Therefore, the four-zone policy always yields a higher revenue rate when $\psi_0(\lambda)(\beta r_t c_1(\alpha) R - r_p c_2) + 4\beta r_p R > 0$ and

$$n < \left[\frac{(\beta R r_t c_1(\alpha) - r_p c_2)\psi_0(\lambda) + 4\beta r_p R}{r_p(\psi_0(\lambda)(c_1(\alpha) + \alpha_2 - c_3) + 3c_3)}\right]^2.$$

□

D.2. Proofs for Section 4

We first present a technical supporting result for $L_3(i)$, followed by the remaining proofs for the results in Section 4.

PROPOSITION 7. *For a fixed point that is uniformly distributed in the diamond area with edge length R, the expected shortest distance from this point to a set of $N-i$ points which are also uniformly distributed in the diamond, is $c_2 R/\sqrt{N-i}$ where $c_2 = \frac{\sqrt{\pi}}{2\sqrt{2}}$.*

Proof of Proposition 7. The result can be deduced from Lemma 3 by setting $\alpha = 0$. □

Proof of Proposition 4. For any $K \geq 2$, denote the optimal solution to problem (6) by $\mathcal{X}^* = \{\mathbf{x}_1^*, \ldots, \mathbf{x}_K^*\}$, and compute $d_k^* = \min_{\mathbf{x}\in\mathcal{X}^*\setminus\mathbf{x}_k^*} ||\mathbf{x} - \mathbf{x}_k^*||$ as the distance between $\mathbf{x}_k^*$ and its nearest neighbor among $\mathcal{X}^*$ for $k = 1, \ldots, K$. Let $r_k^* = (d_k^*/2)\cdot\sqrt{2}/2$ be the inradius of the square around centroid $\mathbf{x}_k^*$, as illustrated in Figure 16 (where we assume $\mathbf{x}_{k'}^*$ is closest to $\mathbf{x}_k^*$). By construction, the squares $\mathcal{S}(\mathbf{x}_k^*, r_k^*)$, $k = 1, \ldots, K$, are non-overlapping, thereby satisfying the constraints of problem (7). Therefore, $\mathcal{X}^*$ and $\{r_1^*, \ldots, r_K^*\}$ are a feasible solution to problem (7). Note that $2\sqrt{2}\sum_{k=1}^K r_k^* = \sum_{k=1}^K d_k^*$, so $G_2(K) \geq 2\sqrt{2}\sum_{k=1}^K r_k^*/K = G_1(K)$. □

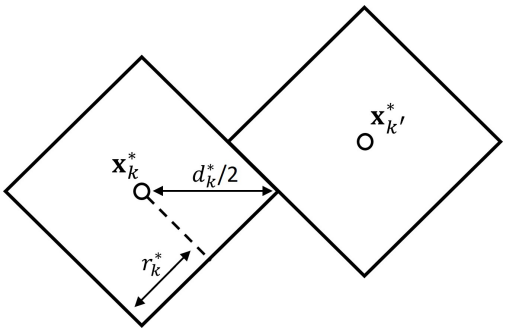


Figure 16 Two squares that are closest to each other with centroids $\mathbf{x}^*_k$ and $\mathbf{x}^*_{k'}$, respectively. Here, $d^*_k = ||\mathbf{x}^*_k - \mathbf{x}^*_{k'}||$ is measured in ℓ_1-distance.

Proof of Lemma 1. According to the definition,

$$\begin{aligned}\psi_1(\lambda, N) &= \int_0^N (1 - \exp(-\hbar\lambda(N-x)/N))dx \\ &= \int_0^N (1 - \exp(-\hbar\lambda x/N))dx \\ &= N + \frac{N}{\hbar\lambda}\exp(-\hbar\lambda x/N)|_0^N \\ &= N + \frac{N}{\hbar\lambda}\left(\exp(-\hbar\lambda) - 1\right).\end{aligned}$$

For ψ_2, we similarly have

$$\begin{aligned}\psi_2(\lambda, N) &= \int_1^N \exp(-\alpha^2\lambda t_0(N-x)/N)/\sqrt{N+1-x}dx \\ &= \exp(-\alpha^2\lambda t_0)\int_1^N \exp(\alpha^2\lambda t_0 x/N)/\sqrt{N+1-x}dx \\ &= \exp(\alpha^2\lambda t_0/N)\int_1^N \exp(-\alpha^2\lambda t_0 x/N)/\sqrt{x}dx.\end{aligned}$$

For notation simplicity, we let $c = \alpha^2\lambda t_0/N$. We compute the integral. Let $u = \sqrt{x}$ so $x = u^2$. Then, $dx = 2udu$. By substituting into the integral,

$$\int_a^b \exp(-cx)/\sqrt{x}dx = 2\int_{\sqrt{a}}^{\sqrt{b}} \exp(-cu^2)du = 2\cdot\frac{\sqrt{\pi}}{2\sqrt{c}}(erf(\sqrt{cb}) - erf(\sqrt{ca})) = \frac{\sqrt{\pi}}{\sqrt{c}}(erf(\sqrt{cb}) - erf(\sqrt{ca})).$$

Therefore,

$$\psi_2(\lambda, N) = \exp(\alpha^2\lambda t_0/N)\cdot\frac{\sqrt{\pi N}}{\sqrt{\alpha^2\lambda t_0}}(erf(\sqrt{\alpha^2\lambda t_0}) - erf(\sqrt{\alpha^2\lambda t_0/N})).$$

Thus, we can write ψ_2 as

$$\psi_2(\lambda, N) = \varsigma_N(\lambda)\sqrt{N},$$

where

$$\varsigma_N(\lambda) = \exp(\alpha^2\lambda t_0/N)\cdot\frac{\sqrt{\pi}}{\sqrt{\alpha^2\lambda t_0}}(erf(\sqrt{\alpha^2\lambda t_0}) - erf(\sqrt{\alpha^2\lambda t_0/N})).$$

□

Proof of Lemma 2. For notation simplicity, let $c = \alpha^2 \lambda t_0$. Define

$$h(N) = \exp(c/N)\left(erf(\sqrt{c}) - erf(\sqrt{c/N})\right).$$

To prove the conclusion, we only need to show that $h(N)$ monotonically increases in N when $N \geq 4$. Taking the derivative with respect to N,

$$\begin{aligned} h'(N) &= -cN^{-2}\exp(c/N)\left(erf(\sqrt{c}) - erf(\sqrt{c/N})\right) + \exp(c/N)\left(-\frac{2}{\sqrt{\pi}}\exp(-c/N)\cdot(-\sqrt{c}\frac{1}{2}N^{-3/2})\right) \\ &= \exp(c/N)N^{-2}\left(-c\left(erf(\sqrt{c}) - erf(\sqrt{c/N})\right) + \frac{\sqrt{c}}{\sqrt{\pi}}\exp(-c/N)N^{1/2}\right). \end{aligned}$$

Note that $\exp(c/N)N^{-2}$ is positive. Let

$$u(N) = -c\left(erf(\sqrt{c}) - erf(\sqrt{c/N})\right) + \frac{\sqrt{c}}{\sqrt{\pi}}\exp(-c/N)N^{1/2}.$$

Taking derivative with respect to N again,

$$\begin{aligned} u'(N) &= \frac{2c}{\sqrt{\pi}}\exp(-c/N)\cdot(-\sqrt{c}\frac{1}{2}N^{-3/2}) + \frac{\sqrt{c}}{\sqrt{\pi}}\left(\exp(-c/N)(cN^{-2})N^{1/2} + \exp(-c/N)\cdot\frac{1}{2}N^{-1/2}\right) \\ &= \frac{\sqrt{c}}{\sqrt{\pi}}\exp(-c/N)N^{-3/2}(-c + 2c + \frac{1}{2}N) \geq 0, \quad \forall N \geq 1. \end{aligned}$$

Since $u(N) > 0$ for $N \geq 1$, we have $u(N) \geq 0$ for all $N \geq 1$. Thus, $h'(N) \geq 0$ for all $N \geq 1$, which implies that $h(N)$ monotonically increases for $N \geq 1$. When N increases from 4 to infinity, it increases from ζ_4 to $\zeta_\infty := \frac{\sqrt{\pi}\cdot erf(\sqrt{\alpha^2\lambda t_0})}{\sqrt{\alpha^2\lambda t_0}}$. Thus, we reach our conclusion. □

Proof of Corollary 1. Fix any N. Recall that

$$f_\lambda(N) := v\frac{nr_p/R + C_1 r_t \psi_1(\lambda, N)}{\beta\sqrt{n} - \beta N/\sqrt{n} + C_2\psi_1(\lambda, N) + C_3\psi_2(\lambda, N)} = v\frac{nr_p/R + C_1 r_t \psi_1(\lambda, N)}{\beta\sqrt{n} - \beta N/\sqrt{n} + C_2\psi_1(\lambda, N) + C_3\varsigma_N(\lambda)\sqrt{N}},$$

and

$$\underline{f}_\lambda(N) = v\frac{nr_p/R + C_1 r_t \psi_1(\lambda, N)}{\beta\sqrt{n} - \beta N/\sqrt{n} + C_2\psi_1(\lambda, N) + C_3\varsigma_\infty(\lambda)\sqrt{N}}.$$

Note that

$$\begin{aligned} \underline{f}_\lambda(N) &= v\frac{nr_p/R + C_1 r_t \psi_1(\lambda, N)}{\beta\sqrt{n} - \beta N/\sqrt{n} + C_2\psi_1(\lambda, N) + C_3\varsigma_\infty(\lambda)\sqrt{N}} \\ &= v\frac{nr_p/R + C_1 r_t \psi_1(\lambda, N)}{\beta\sqrt{n} - \beta N/\sqrt{n} + C_2\psi_1(\lambda, N) + C_3\varsigma_N(\lambda)\sqrt{N}\cdot\frac{\varsigma_\infty}{\varsigma_N}} \\ &\geq \frac{\varsigma_N}{\varsigma_\infty} f_\lambda(N), \end{aligned}$$

where

$$\begin{aligned} \frac{\varsigma_N}{\varsigma_\infty} &= \frac{\exp(\alpha^2\lambda t_0/N)\cdot(erf(\sqrt{\alpha^2\lambda t_0}) - erf(\sqrt{\alpha^2\lambda t_0/N}))}{erf(\sqrt{\alpha^2\lambda t_0})} \\ &\geq \frac{\exp(\alpha^2\lambda t_0/N_0)\cdot(erf(\sqrt{\alpha^2\lambda t_0}) - erf(\sqrt{\alpha^2\lambda t_0/N_0}))}{erf(\sqrt{\alpha^2\lambda t_0})}, \end{aligned}$$

The last inequality holds from monotonicity of ζ_N in Lemma 2. □

Proof of Proposition 5. Recall that

$$\psi_1(\lambda, N) = \kappa N, \quad \text{and} \quad \bar{\psi}_2(\lambda, N) = \varsigma_\infty \sqrt{N},$$

where $\kappa = 1 + (\exp(-\hbar\lambda) - 1)/(\hbar\lambda)$ and $\varsigma_\infty = \frac{\sqrt{\pi} erf(\sqrt{\alpha^2 \lambda t_0})}{\sqrt{\alpha^2 \lambda t_0}}$. The derivative of ψ_1 with respect to N is

$$\frac{\partial \psi_1}{\partial N} = \kappa > 0,$$

and the derivative of $\bar{\psi}_2$ with respect to N is

$$\frac{\partial \bar{\psi}_2}{\partial N} = \frac{1}{2} \varsigma_\infty N^{-1/2} > 0.$$

Since the lower bound of revenue rate of the mixed delivery policy is

$$\underline{f}_\lambda(N) := v \frac{n r_p / R + C_1 r_t \psi_1(\lambda, N)}{\beta\sqrt{n} - \beta N/\sqrt{n} + C_2 \psi_1(\lambda, N) + C_3 \bar{\psi}_2(\lambda, N)},$$

the numerator of the derivative with respect to N equals

$$\begin{aligned}
& v \left(C_1 r_t \frac{\partial \psi_1}{\partial N} (\beta\sqrt{n} - \beta N/\sqrt{n} + C_2 \psi_1(\lambda, N) + C_3 \bar{\psi}_2(\lambda, N)) \right. \\
& \left. - (n r_p / R + C_1 r_t \psi_1(\lambda, N)) \left(-\beta/\sqrt{n} + C_2 \frac{\partial \psi_1}{\partial N} + C_3 \frac{\partial \bar{\psi}_2}{\partial N} \right) \right) \\
=& v\kappa \left((C_1 r_t \beta\sqrt{n} - C_2 n r_p / R + r_p/(\kappa R)\beta\sqrt{n}) + C_1 C_3 r_t \varsigma_\infty \sqrt{N} - C_3 \left(\frac{1}{2} \frac{\varsigma_\infty}{\kappa} n r_p N^{-1/2} / R + \frac{1}{2} C_1 r_t \varsigma_\infty \sqrt{N} \right) \right) \\
=& v\kappa \left((C_1 r_t \beta\sqrt{n} - C_2 n r_p / R + r_p/(\kappa R)\beta\sqrt{n}) + \frac{1}{2} C_1 C_3 r_t \varsigma_\infty \sqrt{N} - \frac{1}{2} \frac{\varsigma_\infty C_3 n r_p}{\kappa R \sqrt{N}} \right). \qquad (15)
\end{aligned}$$

The derivative is monotonically increasing in N. Therefore, there exists a threshold N^Δ such that the lower bound of expected revenue function $\underline{f}_\lambda(N)$ is decreasing for $N \le N^\Delta$ and increasing for $N \ge N^\Delta$. □

Proof of Theorem 1. (a) From Equation (15), while the first part of the derivative is independent of N, the second part of the numerator of the derivative equals

$$\frac{1}{2} C_1 C_3 r_t \varsigma_\infty \sqrt{N} - \frac{1}{2} \frac{\varsigma_\infty C_3 n r_p}{\kappa R \sqrt{N}}.$$

It is monotonically increasing in N. Therefore, when N increases from 4 to infinity, the profit rate either monotonically increases, or it decreases first and then increases. Thus, the optimal N is either 1, 4, or n.

(b) Note that the revenue rate of pure delivery is

$$R^{PD}(n, \lambda) = v \frac{n r_p}{\beta\sqrt{n}} = \frac{v\sqrt{n} r_p}{\beta R},$$

and the lower bound of revenue rate of mixed delivery with $N=4$ is

$$\underline{f}_\lambda(4) := \frac{nr_p/R + C_1 r_t \psi_1(\lambda,4)}{\beta\sqrt{n} - 4\beta/\sqrt{n} + C_2\psi_1(\lambda,4) + C_3\bar{\psi}_2(\lambda,4)},$$

and the lower bound of revenue rate of mixed delivery with $N=n$ is

$$\underline{f}_\lambda(n) := \frac{nr_p/R + C_1 r_t \psi_1(\lambda,n)}{C_2\psi_1(\lambda,n) + C_3\bar{\psi}_2(\lambda,n)}.$$

We first compare the revenue rate of $N=4$ and $N=n$ (switching policy). The inequality $\underline{f}_\lambda(n) \geq \underline{f}_\lambda(4)$ holds if and only if

$$\begin{aligned}
&\frac{nr_p/R + C_1 r_t \psi_1(\lambda,n)}{C_2\psi_1(\lambda,n) + C_3\bar{\psi}_2(\lambda,n)} \geq \frac{nr_p/R + C_1 r_t \psi_1(\lambda,4)}{\beta\sqrt{n} - 4\beta/\sqrt{n} + C_2\psi_1(\lambda,4) + C_3\bar{\psi}_2(\lambda,4)} \\
\Leftrightarrow\ & n\frac{r_p}{R}(C_2\psi_1(\lambda,4) + C_3\bar{\psi}_2(\lambda,4)) + \beta r_p/R(n^{\frac{3}{2}} - 4\sqrt{n}) + C_1 r_t \psi_1(\lambda,n)\beta(\sqrt{n} - 4/\sqrt{n}) + C_1 r_t \psi_1(\lambda,n)\cdot C_3\bar{\psi}_2(\lambda,4) \\
&\geq n\frac{r_p}{R}(C_2\psi_1(\lambda,n) + C_3\bar{\psi}_2(\lambda,n)) + C_1 C_3 \psi_1(\lambda,4)\bar{\psi}_2(\lambda,n) r_t \\
\Leftrightarrow\ & n\frac{r_p}{R}(C_2\cdot 4\kappa + C_3\cdot 2\varsigma_\infty) + \beta r_p/R(n^{\frac{3}{2}} - 4\sqrt{n}) + C_1 r_t \kappa n\beta(\sqrt{n} - 4/\sqrt{n}) + C_1 r_t\cdot \kappa n\cdot C_3\cdot 2\varsigma_\infty \\
&\geq n\frac{r_p}{R}(C_2\kappa n + C_3\varsigma_\infty\sqrt{n}) + C_1 C_3 r_t\cdot 4\kappa\cdot\varsigma_\infty\sqrt{n} \\
\Leftrightarrow\ & n\frac{r_p}{R}((C_2^1 + C_2^2/\sqrt{n})\cdot 4\kappa + C_3\cdot 2\varsigma_\infty) + \beta r_p/R(n^{\frac{3}{2}} - 4\sqrt{n}) + C_1 r_t\kappa n\beta(\sqrt{n} - 4/\sqrt{n}) + C_1 r_t\cdot\kappa n\cdot C_3\cdot 2\varsigma_\infty \\
&\geq n\frac{r_p}{R}((C_2^1 + C_2^2/\sqrt{n})\kappa n + C_3\varsigma_\infty\sqrt{n}) + C_1 C_3 r_t\cdot 4\kappa\cdot\varsigma_\infty\sqrt{n} \\
\Leftrightarrow\ & -\kappa C_2^1\frac{r_p}{R}n^2 + (\beta r_p/R + C_1 r_t\kappa\beta - C_2^2\kappa r_p/R - C_3\varsigma_\infty\frac{r_p}{R})n^{3/2} + (4\kappa C_2^1 r_p/R + 2C_1 C_3\kappa\varsigma_\infty r_t + 2C_3\varsigma_\infty r_p/R)n \\
&+ (4\kappa C_2^2 r_p/R - 4\beta r_p/R - 4C_1 r_t\kappa\beta - 4C_1 C_3 r_t\kappa\varsigma_\infty)\sqrt{n} \geq 0 \\
\Leftrightarrow\ & \kappa C_2^1 n^{3/2} + (\kappa C_2^2 + C_3\varsigma_\infty - \beta - \beta C_1\kappa R r_t/r_p)n \\
&\leq 4(\kappa C_2^2 - \beta - C_1\kappa\beta R r_t/r_p - C_1 C_3\kappa\varsigma_\infty R r_t/r_p) + (4\kappa C_2^1 + 2C_1 C_3\kappa\varsigma_\infty R r_t/r_p + 2C_3\varsigma_\infty)\sqrt{n}.
\end{aligned}$$

Let $t=\sqrt{n}$ and define function

$$\begin{aligned}
g(t) =& C_2^1\kappa t^3 + (\kappa C_2^2 + C_3\varsigma_\infty - \beta - \beta R C_1\kappa r_t/r_p)t^2 \\
&- 4\kappa(C_2^2 - \beta/\kappa - \beta R C_1 r_t/r_p - C_1 C_3\varsigma_\infty R r_t/r_p) - (4\kappa C_2^1 + 2C_1 C_3\kappa\varsigma_\infty R r_t/r_p + 2C_3\varsigma_\infty)t.
\end{aligned}$$

From the derivation above, we know that $\underline{f}_\lambda(n) \geq \underline{f}_\lambda(4)$ holds if and only if $g(t)\leq 0$. We note that $g(0) = -4\kappa(C_2^2 - \beta/\kappa - \beta R C_1 r_t/r_p - C_1 C_3\varsigma_\infty R r_t/r_p)$ and $g(2)=0$ (corresponding to the case where $n=4$). The derivative of $g(t)$ is

$$g'(t) = 3C_2^1\kappa t^2 + 2(\kappa C_2^2 + C_3\varsigma_\infty - \beta - \beta R C_1\kappa r_t/r_p)t - (4\kappa C_2^1 + 2C_1 C_3\kappa\varsigma_\infty R r_t/r_p + 2C_3\varsigma_\infty),$$

which implies that $g'(0) = -(4\kappa C_2^1 + 2C_1 C_3\kappa\varsigma_\infty R r_t/r_p + 2C_3\varsigma_\infty) < 0$ and

$$\begin{aligned}
g'(2) &= 12C_2^1\kappa + 4(\kappa C_2^2 + C_3\varsigma_\infty - \beta - \beta R C_1\kappa r_t/r_p) - (4\kappa C_2^1 + 2C_1 C_3\kappa\varsigma_\infty R r_t/r_p + 2C_3\varsigma_\infty) \\
&= 8C_2^1\kappa + 4\kappa C_2^2 + 2(C_3\varsigma_\infty - 2\beta - 2\beta R C_1\kappa r_t/r_p - C_1 C_3\kappa\varsigma_\infty R r_t/r_p).
\end{aligned}$$

Recall that $\omega = r_t/r_p$, so by combining the above properties, we conclude that there are two possibilities:

(I) When $g'(2) \geq 0$, we have $g(t) \geq 0$ for all $t \geq 2$. This is because $g(t)$ has at most two local minimum or maximum, and $g(t)$ goes to (negative) infinity when t goes to (negative) infinity. See Figure 17 as an illustration. Therefore, when

$$\omega \leq \frac{4C_2^1 + 2C_2^2 + (C_3\varsigma_\infty - 2\beta)/\kappa}{(2\beta C_1 + C_1 C_3 \varsigma_\infty)R} =: \eta_1, \tag{16}$$

we have $\underline{f}_\lambda(n) \leq \underline{f}_\lambda(4)$ for all $n \geq 4$.

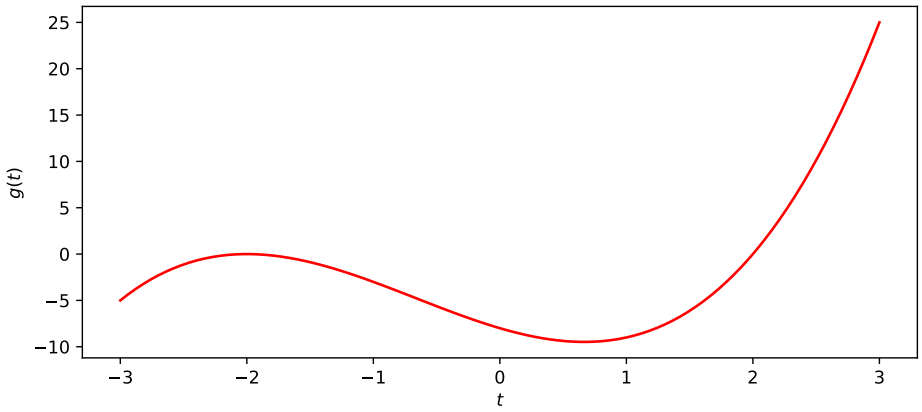


Figure 17 Scenario (I) of $g(t)$.

(II) When $g'(2) < 0$, i.e., when

$$\omega > \frac{4C_2^1 + 2C_2^2 + (C_3\varsigma_\infty - 2\beta)/\kappa}{(2\beta C_1 + C_1 C_3 \varsigma_\infty)R} = \eta_1,$$

there exists some $t' > 2$ such that $g(t) \leq 0$ for all $2 \leq t \leq t'$ and $g(t) > 0$ for all $t > t'$. This is because $g(t)$ has at most two local minimum or maximum, and $g(t)$ goes to (negative) infinity when t goes to (negative) infinity. In addition, $g'(2), g'(0) < 0$. Therefore, there exists n_1 which depends on ω such that $\underline{f}_\lambda(n) \geq \underline{f}_\lambda(4)$ if $n \leq n_1(\omega)$ and otherwise $\underline{f}_\lambda(n) \leq \underline{f}_\lambda(4)$. Here, $n_1(\omega)$ is the largest root of $g(t)$.

Next, *we compare the revenue rate of $N = n$ (switching policy) and $N = 1$ (pure delivery policy).* $\underline{f}_\lambda(n) \geq \underline{f}_\lambda(1)$ if and only if

$$\begin{aligned}
&\frac{nr_p/R + C_1 r_t \kappa n}{C_2 \kappa n + C_3 \varsigma_\infty \sqrt{n}} \geq \frac{\sqrt{n} r_p}{\beta R} \\
\Leftrightarrow\ &\frac{r_p/R + C_1 \kappa r_t}{(C_2^1 + C_2^2/\sqrt{n})\kappa\sqrt{n} + C_3\varsigma_\infty} \geq \frac{r_p}{\beta R} \\
\Leftrightarrow\ &\frac{r_p/R + C_1 \kappa r_t}{\kappa(C_2^1\sqrt{n} + C_2^2) + C_3\varsigma_\infty} \geq \frac{r_p}{\beta R} \\
\Leftrightarrow\ &r_p(\kappa(C_2^1\sqrt{n} + C_2^2) + C_3\varsigma_\infty) \leq \beta R C_1 \kappa r_t + \beta r_p.
\end{aligned}$$

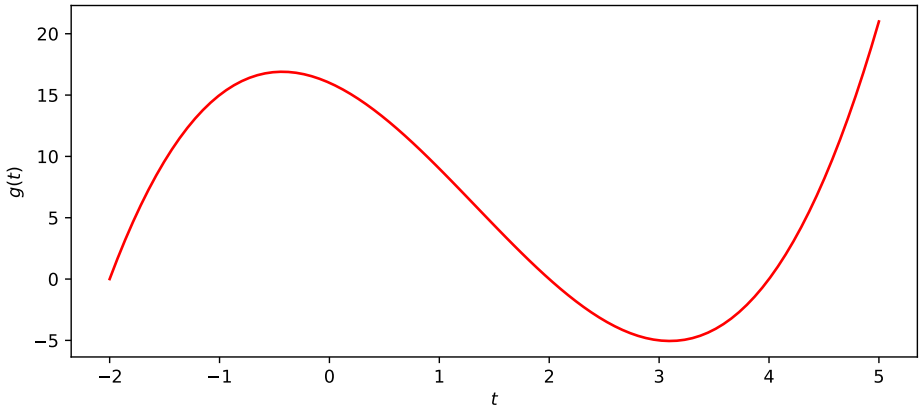


Figure 18 Scenario (II) of $g(t)$.

If $\beta RC_1\kappa r_t/r_p + \beta - C_3\varsigma_\infty - \kappa C_2^2 < 0$, i.e.,

$$\omega = \frac{r_t}{r_p} < \frac{C_2^2 + C_3\varsigma_\infty/\kappa - \beta/\kappa}{\beta RC_1} =: \eta_2,$$

then $\underline{f}_\lambda(1) \geq \underline{f}_\lambda(n)$ for all n. Otherwise,

$$\underline{f}_\lambda(n) \geq \underline{f}_\lambda(1) \text{ if and only if } n \leq \left[\frac{\beta RC_1 r_t/r_p + \beta/\kappa - (C_3\varsigma_\infty/\kappa + C_2^2)}{C_2^1}\right]^2 =: n_2(\omega).$$

Lastly, *we compare the revenue rate of* $N = 4$ *and* $N = 1$ *(pure delivery policy).* From Proposition 3, $\underline{f}_\lambda(4) \geq \underline{f}_\lambda(1)$ if and only if

$$\begin{aligned}
&\frac{nr_p/R + C_1 r_t \psi_1(\lambda, 4)}{\beta\sqrt{n} - 4\beta/\sqrt{n} + C_2\psi_1(\lambda,4) + C_3\bar{\psi}_2(\lambda,4)} \geq \frac{\sqrt{n}r_p}{\beta R} \\
\Leftrightarrow &\frac{nr_p/R + C_1 r_t \cdot 4\kappa}{\beta\sqrt{n} - 4\beta/\sqrt{n} + (C_2^1 + C_2^2/\sqrt{n})\cdot 4\kappa + C_3\cdot 2\varsigma_\infty} \geq \frac{\sqrt{n}r_p}{\beta R} \\
\Leftrightarrow &-4\beta r_p + 4C_2^1\kappa r_p\sqrt{n} + 4C_2^2\kappa r_p + 2C_3\varsigma_\infty\sqrt{n}r_p \leq 4C_1\kappa r_t\beta R \\
\Leftrightarrow &(2C_2^1\kappa + C_3\varsigma_\infty)\sqrt{n} + 2C_2^2\kappa \leq 2C_1\kappa\beta R r_t/r_p + 2\beta
\end{aligned}$$

Thus, if

$$\omega = \frac{r_t}{r_p} \leq \frac{C_2^2 - \beta/\kappa}{C_1\beta R} =: \eta_3$$

then $\underline{f}_\lambda(4) \leq \underline{f}_\lambda(1)$ for all number of packages. If

$$\omega = \frac{r_t}{r_p} \geq \frac{C_2^2 - \beta/\kappa}{C_1\beta R} = \eta_3,$$

$\underline{f}_\lambda(4) \geq \underline{f}_\lambda(1)$ if and only if $n \leq n_3(\omega)$ where

$$n_3(\omega) = \left[\frac{2C_1\kappa\beta R r_t/r_p + 2\beta - 2C_2^2\kappa}{2C_2^1\kappa + C_3\varsigma_\infty}\right]^2.$$

Therefore, we conclude that

(I) $\underline{N}^* = 1$ when both (a) either $\omega \le \eta_2$ or $n \ge n_2(\omega)$; and (b) either $\omega \le \eta_3$ or $n \ge n_3(\omega)$ are satisfied.

(II) $\underline{N}^* = 4$ when both (a) either $\omega \le \eta_1$ or $n \ge n_1(\omega)$; and (b) $\omega \ge \eta_3$ and $n \le n_3(\omega)$ are satisfied.

(III) $\underline{N}^* = n$ when (a) $\omega \ge \max\{\eta_1, \eta_2\}$ and (b) $n \le \min\{n_1(\omega), n_3(\omega)\}$.

□

Proof of Corollary 2. From Corollary 1, we can bound $\underline{f}_\lambda(N)$ by

$$\frac{\varsigma_N}{\varsigma_\infty} f_\lambda(N) \le \underline{f}_\lambda(N) \le f_\lambda(N), \quad \forall N.$$

If we show that $N = \underline{N}^*$ is optimal for $\underline{f}_\lambda(N)$, then

$$f_\lambda(\underline{N}^*) \ge \underline{f}_\lambda(\underline{N}^*) \ge \underline{f}_\lambda(N) \ge \frac{\varsigma_N}{\varsigma_\infty} f_\lambda(N), \quad \forall N.$$

Therefore, for N^* that maximizes $f_\lambda(N)$, it holds that

$$f_\lambda(\underline{N}^*) \ge \frac{\varsigma_{N^*}(\lambda)}{\varsigma_\infty} f_\lambda(N^*).$$

□

Proof of Theorem 2. We consider the full-information benchmark policy as follows: the driver takes m passenger rides whose origination and destination are generated uniformly among the region, are known beforehand. The parameter m will be optimized later. The expected revenue of the best coordination policy is

$$R^\diamond(n, \lambda) = \frac{n r_p + l_2 r_t}{l_1 + l_2}.$$

Here, l_1 represents the expected length of the route corresponding to package deliveries, while l_2 represents the length of the route associated with passenger transport. More specifically, l_1 includes all the routes that connect packages to packages and connect packages to passengers; l_2 includes all the routes that connect passenger's drop-off and pick-up locations. When $m = 0$, then $R^\diamond(n, \lambda) = \frac{r_p}{\beta R}\sqrt{n}$. Now we investigate the scenario where $1 \le m \le n$. Then, under the assumption that all packages and passenger locations are uniformly distribution, we have

$$l_1 \ge n \cdot c_2 \frac{R}{\sqrt{n+m}} = c_2 R \sqrt{\frac{n^2}{n+m}},$$

where $c_2 \frac{R}{\sqrt{n+m}}$ is the expected distance to the nearest location when $n+m$ locations are uniformly distributed within the zone, as shown in Proposition 1. Moreover, because passengers' origins and destinations are uniformly distributed, we have

$$l_2 = m c_l R,$$

where $c_l = \frac{7\sqrt{2}}{15}$. Then,

$$R^\diamond(n,\lambda) = v\frac{nr_p + l_2 r_t}{l_1 + l_2}.$$

Note that $l_1 + l_2 \geq \beta R\sqrt{n}$, so we have

$$R^\diamond(n,\lambda)/v \leq \max_{1\leq m\leq n} \frac{nr_p + mc_l R r_t}{\max\{c_2 R\sqrt{n^2/(n+m)} + mc_l R, \beta R\sqrt{n}\}} \leq \max_{1\leq m\leq n} \frac{nr_p + mc_l R r_t}{\max\{c_2 R\sqrt{n/2} + mc_l R, \beta R\sqrt{n}\}}.$$

– When $c_2 R\sqrt{n/2} + mc_l R \leq \beta R\sqrt{n}$, i.e., when $m \leq \frac{(\beta R - c_2 R/\sqrt{2})\sqrt{n}}{c_l R}$, it holds that

$$R^\diamond(n,\lambda)/v \leq \frac{nr_p + \frac{(\beta R - c_2 R/\sqrt{2})\sqrt{n}}{c_l R} c_l R r_t}{\beta R\sqrt{n}} = \frac{nr_p + (\beta R - c_2 R/\sqrt{2})\sqrt{n} r_t}{\beta R\sqrt{n}}.$$

– On the other hand, when $m \geq \frac{(\beta R - c_2 R/\sqrt{2})\sqrt{n}}{c_l R}$, it holds that

$$R^\diamond(n,\lambda)/v \leq \max_{\frac{(\beta R - c_2 R/\sqrt{2})\sqrt{n}}{c_l R} \leq m\leq n} \frac{nr_p + mc_l R r_t}{c_2 R\sqrt{n/2} + mc_l R} = r_t + \max_{\frac{(\beta R - c_2 R/\sqrt{2})\sqrt{n}}{c_l R} \leq m\leq n} \frac{nr_p - c_2 R\sqrt{n/2} r_t}{c_2 R\sqrt{n/2} + mc_l R}. \quad (17)$$

We optimize for m. If $nr_p - c_2 R\sqrt{n/2} r_t \geq 0$, i.e., $n \geq \frac{1}{2}(\frac{C_2 R r_t}{r_p})^2$, then $m^* = \frac{(\beta R - c_2 R/\sqrt{2})}{c_l R}\sqrt{n}$. In this case,

$$R^\diamond(n,\lambda)/v \leq r_t + \frac{nr_p - c_2 R\sqrt{n/2} r_t}{\beta R\sqrt{n}} = \frac{nr_p + (\beta R - c_2 R/\sqrt{2})\sqrt{n} r_t}{\beta R\sqrt{n}}. \quad (18)$$

Otherwise, if $nr_p - c_2 R\sqrt{n/2} r_t \leq 0$, i.e., $n \leq \frac{1}{2}(\frac{C_2 R r_t}{r_p})^2$, then from (17), we have

$$R^\diamond(n,\lambda)/v \leq r_t.$$

To summarize from (17) and (18), we have

$$R^\diamond(n,\lambda)/v \leq \max\left\{r_t, \frac{nr_p + (\beta R - c_2 R/\sqrt{2})\sqrt{n} r_t}{\beta R\sqrt{n}}\right\}. \quad (19)$$

Next we compare two quantities in (19). Note that

$$\frac{nr_p + (\beta R - c_2 R/\sqrt{2})\sqrt{n} r_t}{\beta R\sqrt{n}} = \frac{r_p}{\beta R}\sqrt{n} + \frac{(\beta - c_2/\sqrt{2})}{\beta} r_t.$$

For ease of notation, we define $\xi = (\beta - c_2/\sqrt{2})R$ and $\xi' = \frac{\beta - c_2/\sqrt{2}}{\beta}$. Then, when

$$n \leq (\frac{(1-\xi')\beta R r_t}{r_P})^2,$$

we have $r_t \geq \frac{nr_p + (\beta R - c_2 R/\sqrt{2})\sqrt{n} r_t}{\beta R\sqrt{n}}$, so it holds that

$$R^\diamond(n,\lambda)/v \leq r_t.$$

On the other hand, when $n \geq (\frac{(1-\xi')\beta R r_t}{r_P})^2$, we have

$$R^\diamond(n,\lambda)/v \leq \frac{nr_p + (\beta R - c_2 R/\sqrt{2})\sqrt{n} r_t}{\beta R\sqrt{n}}.$$

Next, we analyze our proposed zoning strategy. From Equation (4), by setting $N=1$, we obtain the lower bound

$$R^* := \max_N \; R^*(N,\lambda) \geq R^*(1,\lambda) = \frac{v r_p}{\beta R}\sqrt{n}.$$

To obtain a tighter lower bound, by plugging $N=n$ in Equation (9), we have

$$\begin{aligned} R^* = \max_N \; R^*(N,\lambda) \geq R^*(n,\lambda) &= v\frac{nr_p/R + C_1 r_t \psi_1(\lambda,n)}{C_2\psi_1(\lambda,n) + C_3\psi_2(\lambda,n)} \\ &\geq v\frac{nr_p/R + C_1 r_t \kappa n}{C_2\kappa n + C_3\varsigma_\infty\sqrt{n}}, \end{aligned}$$

where the last inequality is derived from Lemma 1. When λ goes to infinity, since $\kappa \to 1$ and $\varsigma_\infty \to 0$, the revenue rate is lower bounded by

$$v\frac{r_p + C_1 r_t R}{C_2 R}.$$

Recall $\xi = (\beta - c_2/\sqrt{2})R$. From Equation (19),

(I) When $n \geq (\frac{(1-\xi')\beta R r_t}{r_P})^2$, the performance ratio can be bounded by

$$\frac{R^*(n,\lambda)}{R^\diamond(n,\lambda)} \geq \frac{\frac{r_p}{\beta R}\sqrt{n}}{\frac{r_p}{\beta R}\sqrt{n} + \xi' r_t} = 1 - \frac{\xi r_t}{r_p\sqrt{n} + \xi r_t},$$

which increases in n and converges to 1 when n goes to infinity.

(II) When $n \leq (\frac{(1-\xi')\beta R r_t}{r_P})^2$, the performance ratio can be bounded by

$$\frac{R^*(n,\lambda)}{R^\diamond(n,\lambda)} \geq \frac{r_p}{\beta R r_t}\sqrt{n}.$$

(III) Further, when λ goes to infinity, when $n \leq (\frac{(1-\xi')\beta R r_t}{r_P})^2$, we have

$$\lim_{\lambda\to\infty} \frac{R^*(n,\lambda)}{R^\diamond(n,\lambda)} \geq \frac{\frac{r_p + C_1 r_t R}{C_2 R}}{r_t} = \frac{r_p + C_1 r_t R}{C_2 R r_t}.$$

Thus, we can conclude that for all n and λ,

$$\frac{R^*(n,\lambda)}{R^\diamond(n,\lambda)} \geq \min\left\{\tau\sqrt{n}, 1 - \frac{\xi r_t}{r_p\sqrt{n} + \xi r_t}\right\},$$

where $\tau = \frac{r_p}{\beta R r_t}$ and $\xi = (\beta - c_2/\sqrt{2})R$. In particular, when λ goes to infinity, we can conclude that

$$\lim_{\lambda\to\infty} \frac{R^*(n,\lambda)}{R^\diamond(n,\lambda)} \geq \min\left\{\max\left\{\tau\sqrt{n}, \frac{r_p + C_1 r_t R}{C_2 R r_t}\right\}, 1 - \frac{\xi r_t}{r_p\sqrt{n} + \xi r_t}\right\}.$$

□

Appendix E: Impact of α on the Choice between the Four-Zone and Pure Delivery Policy

In this section, we report the optimal operating regimes of the four-zone policy versus the pure delivery policy when α is optimized to maximize the revenue rate. Specifically, following the same setting as in Figure 2, we enumerate over $\{0.1, 0.15, \ldots, 0.3\}$ to identify the best α for each n and λ. In general, the insights from this figure are similar to those from Figure 2, while the boundary separating the two policies is smoother when α is optimized. We observe that an optimized α would favor the four-zone policy (and generally the coordination policy) because it enables another lever to manage the revenue performance in response to the market condition. Figure 20 illustrates the optimal value of α as a function of n and λ. Intuitively, as λ increases, the ride-hailing opportunities are more abundant, and it is beneficial to choose a lower α to minimize pickup distance and detours. With a similar logic, a lower α is preferred with a larger n to prioritize delivery efficiency, while the value of optimal α is less sensitive to n.

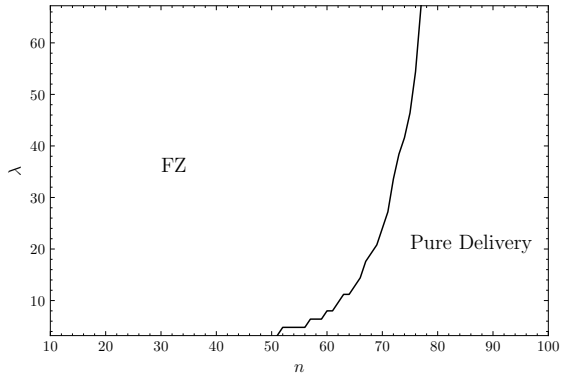


Figure 19 Optimal operating regimes of the four-zone policy and the pure delivery policy when α is optimized over $\{0.1, 0.15, 0.2, 0.25, 0.3\}$

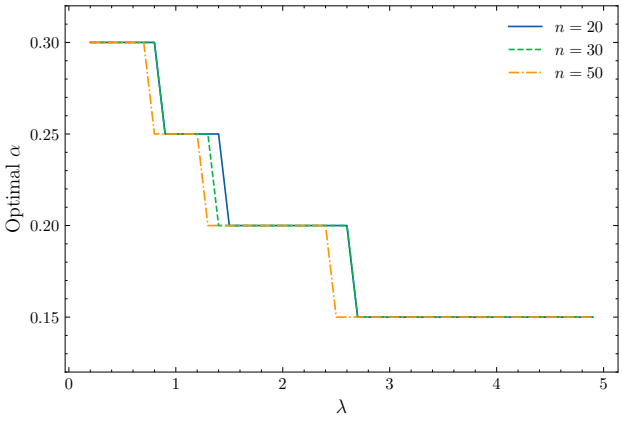


Figure 20 The value of optimal α with respect to n and λ in the considered four-zone policy

Appendix F: Additional Formulations for Section 4

This section presents detailed formulations for approximations used in Section 4.

F.1. MILP Formulation for $G_1(K)$

We consider an MILP formulation to solve for $G_1(K)$:

$$\max_{\mathcal{X}=\{\mathbf{x}_1,\ldots,\mathbf{x}_K\}\subset\mathcal{A}} \frac{1}{K}\sum_{k=1}^{K}\min_{\mathbf{x}\in\mathcal{X}\setminus\mathbf{x}_k}||\mathbf{x}-\mathbf{x}_k||.$$

The key is to replace the inner minimization problem with additional binary variables. Specifically, let $d_k = \min_{\mathbf{x}\in\mathcal{X}\setminus\mathbf{x}_k}||\mathbf{x}-\mathbf{x}_k||$, and we rewrite the objective function as $\frac{1}{K}\sum_{k=1}^{K} d_k$. For the constraints, we introduce for each k,

$$d_k \le d_{kl}^1 + d_{kl}^2, \quad \forall l \in \{1,\ldots,K\}/k \tag{20}$$

$$d_{kl}^1 \ge x_k^1 - x_l^1, \quad \forall l \in \{1,\ldots,K\}/k \tag{21}$$

$$d_{kl}^1 \ge x_l^1 - x_k^1, \quad \forall l \in \{1,\ldots,K\}/k \tag{22}$$

$$d_{kl}^2 \ge x_k^2 - x_l^2, \quad \forall l \in \{1,\ldots,K\}/k \tag{23}$$

$$d_{kl}^2 \ge x_l^2 - x_k^2, \quad \forall l \in \{1,\ldots,K\}/k \tag{24}$$

$$d_{kl}^1 = x_k^1 - x_l^1 + 2z_{kl}^1(x_l^1 - x_k^1), \quad \forall l \in \{1,\ldots,K\}/k \tag{25}$$

$$d_{kl}^2 = x_k^2 - x_l^2 + 2z_{kl}^2(x_l^2 - x_k^2), \quad \forall l \in \{1,\ldots,K\}/k \tag{26}$$

where d_{kl}^1 and d_{kl}^2 denote the absolute difference in the x-coordinate and y-coordinate between points $\mathbf{x}_k$ and $\mathbf{x}_l$, respectively; x_k^1 and x_k^2 represent the x-coordinate and y-coordinate of point $\mathbf{x}_k$. Constraints (21)-(24) ensure the differences are nonnegative (in ℓ_1-norm). Additionally, constraints (25)-(26) enforce the definition of d_{kl}^1 and d_{kl}^2, respectively: $d_{kl}^1 = x_k^1 - x_l^1$ if $z_{kl}^1 = 0$ and $d_{kl}^1 = x_l^1 - x_k^1$ if $z_{kl}^1 = 1$; $d_{kl}^2 = x_k^2 - x_l^2$ if $z_{kl}^2 = 0$ and $d_{kl}^2 = x_l^2 - x_k^2$ if $z_{kl}^2 = 1$. Then, following the standard linearization technique, we can further linearize the bilinear terms $z_{kl}^1 x_l^1$, $z_{kl}^1 x_k^1$, $z_{kl}^2 x_l^2$, and $z_{kl}^2 x_k^2$ in constraints (25)-(26), which leads to an MILP formulation. In summary, the resulting MILP includes $O(K^2)$ continuous variables and $O(K^2)$ binary variables.

We tested the above MILP using Gurobi on an Intel i7-8700 3.20GHz Windows desktop with 16GB of RAM. We find that it can not be solved for $K \ge 10$ with a time limit of six hours.

F.2. Supporting Results and Formulations for the Approximation of $L_3(i)$

Below, we demonstrate the key structural properties of $G_2(K)$ that can be leveraged to formulate a tractable mathematical program to bound $G_2(K)$.

Proposition 8. *An optimal solution $\mathcal{X}^*$ to problem (7) satisfies:*

(I) *The outermost points (the points closest to the boundary edges) of $\mathcal{X}^*$ must lie on the boundary of $\mathcal{A}$;*

(II) *When $K \geq 4$, the four vertices of $\mathcal{A}$ must belong to $\mathcal{X}^*$.*

Proof of Proposition 8. We prove the result in two parts.

(I) To show that outermost points of $\mathcal{X}^*$ must lie on the boundary of $\mathcal{A}$, it suffices to prove that for any solution $\mathcal{X}'$ (with square inradiuses $\{r'_1, \ldots, r'_K\}$) whose outermost points do not lie on the boundary, we can move those points to the boundary edge and keep the square inradiuses the same, so the squares $\mathcal{S}(\mathbf{x}'_k, r'_k)$, $k = 1, \ldots, K$, are still feasible to problem (7). To see this, note that the outermost points are closest to the boundary, so moving the corresponding squares toward the boundary (along the direction that is perpendicular to the boundary edge) would not violate the non-overlapping constraints with respect to other squares. Therefore, the new solution with the outermost points lying on the boundary shares the same objective value as the original solution.

(II) According to (I), when $K \geq 4$, there exists an optimal solution $\mathcal{X}^*$ with at least four points lying on the boundary of $\mathcal{A}$ (corresponding to the four edges). If there are four points on the boundary that coincide with the vertices, the desired result follows. Suppose there is a vertex of $\mathcal{A}$ that does not belong to $\mathcal{X}^*$, then let $\mathbf{x}^*_{k_1}, \mathbf{x}^*_{k_2} \in \mathcal{X}^*$ be the two points on the two intersecting edges of $\mathcal{A}$ that are closest to this vertex, respectively. Without loss of generality, assume this vertex is the left vertex of the service region (as shown in Figure 21). We make the following argument.

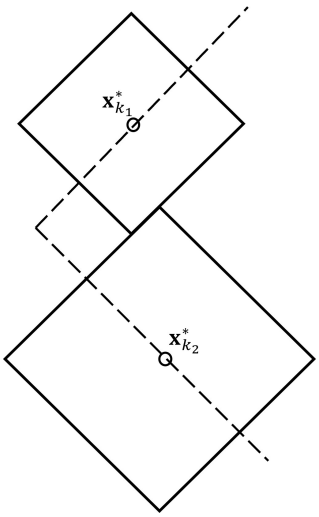


Figure 21 Illustration of $\mathcal{S}(\mathbf{x}^*_{k_1}, r^*_{k_1})$ and $\mathcal{S}(\mathbf{x}^*_{k_2}, r^*_{k_2})$. The two dashed lines are the two boundary edges of the service region, of which the intersection is a vertex.

Claim: Moving either $\mathcal{S}(\mathbf{x}^*_{k_1}, r^*_{k_1})$ or $\mathcal{S}(\mathbf{x}^*_{k_2}, r^*_{k_2})$ toward the vertex can maintain that these two squares do not overlap with each other.

To prove this claim, let $\mathbf{x}^*_{k_1}$ be such that it is closer (or equidistant) to the vertex than $\mathbf{x}^*_{k_2}$ in the x-coordinate. Then moving $\mathbf{x}^*_{k_1}$ toward the vertex (along the edge) does not change $||\mathbf{x}^*_{k_1} - \mathbf{x}^*_{k_2}||$

because the difference in their x-coordinates increases by the same amount as the decrease of their difference in the y-coordinates. It follows that the two squares maintain the same distance from each other, and therefore, the non-overlapping property holds.

Now, we need to check if moving one square (let it be $\mathcal{S}(\mathbf{x}^*_{k_1}, r^*_{k_1})$) toward the vertex would violate the non-overlapping constraints with other squares that do not lie on the boundary. If not, the desired result follows. Otherwise, suppose $\mathcal{S}(\mathbf{x}^*_{k_3}, r^*_{k_3})$ is the one that is closest to the edge among all the squares that will intersect $\mathcal{S}(\mathbf{x}^*_{k_1}, r^*_{k_1})$ along its moving path. Similar to the argument in (I), $\mathbf{x}^*_{k_3}$ can be moved to the boundary edge without violating the non-overlapping constraints. Then we can repeat the above operations for the pair $\mathbf{x}^*_{k_2}$ and $\mathbf{x}^*_{k_3}$ because they are now on the two intersecting edges (and $\mathbf{x}^*_{k_3}$ is closer to the vertex than $\mathbf{x}^*_{k_1}$). This procedure will stop once all four vertices belong to $\mathcal{X}^*$. Because the objective function value and feasibility conditions do not change, the new set of squares is also optimal to problem (7), and the result is proved. □

The intuition behind Proposition 8 is that maximizing the average edge length should lead to a set of squares that are spread out in the service region, pushing the centroids to the boundary. Furthermore, locating squares on the vertices of the service region would take minimum space from $\mathcal{A}$, so there is more room for enlarging squares (and increasing their inradiuses) inside $\mathcal{A}$.

Based on Proposition 8, we can evaluate $G_2(K)$ by conditioning on the number of centroids on the boundary, which is denoted by $4 \le K_b \le K$ (including the vertices). Furthermore, the non-overlapping constraints can be relaxed to admit a convex program that can be efficiently solved. Without loss of generality, we normalize R to be 1. We formulate a relaxed version of problem (7) as follows. According to Proposition 8, for any $K \ge 4$, an optimal solution to problem (7) should include the four vertices of $\mathcal{A}$. Conditioning on the number of points on the boundary (including the vertices) $4 \le K_b \le K$, we solve

$$\max_{r_1,\dots,r_K \ge 0} \quad \frac{2\sqrt{2}}{K} \sum_{k=1}^{K} r_k \tag{27}$$

$$\text{s.t.} \quad \sum_{k=1}^{4} r_k^2 + 2\sum_{k=5}^{K_b} r_k^2 + 4 \sum_{k=K_b+1}^{K} r_k^2 \le 1, \tag{28}$$

$$2\sum_{k=1}^{K_b} r_k \le 4, \tag{29}$$

where the objective function is the same as problem (7), constraint (28) ensures that the total area of squares overlapping with $\mathcal{A}$ is less or equal to 1, and constraint (29) ensures the total edge length of squares on the boundary do not exceed the perimeter of the service region ($= 4R = 4$). These two constraints are satisfied if the squares are non-overlapping. We use a key observation that a square of inradius r on the boundary takes $2r$ from the periphery if it is located on a vertex or does not overlap

with the vertices at all, which is the case for the optimal solution due to Proposition 8. Because both constraints are convex, the above problem can be solved efficiently. We enumerate the possible values of K_b and report the optimal objective function value that serves as an upper bound of $G_2(K)$.

Appendix G: Parameter Calibration

In this section, we provide the details about the parameter calibration for the numerical results reported in Section 4.5. The area of the service region is chosen so it matches that of Manhattan, New York, USA. The average travel speed is assumed to be $v = 7.1$ mph ($= 0.19$ km/minute) based on the average taxi speed profiles summarized by NYC Department of Transportation (2018). The potential on-demand ride arrival rate is estimated based on the maximum pick-up rate (507 per minute), assuming the number of competing drivers is 50, from the 2013 yellow taxi trip record (Ata et al., 2019). Regarding the ride-hailing revenue rate (r_t, per km of trip distance), we base our estimation on the per-minute trip fare of the NYC yellow taxi in 2020 ($1.3/minute) and the average travel speed (0.19 km/minute), assuming a 20% commission rate: $r_t = (\$1.3/0.19) \times 80\% = \$5.5/\text{km}$ (Schneider, 2023). The package revenue r_p is estimated using the average hourly wage ($20) of crowdsourcing drivers in NYC and assuming a driver can deliver 25 packages per hour (A Martínez, 2023).[9]

[9] A driver may deliver 200-300 packages per day, see, e.g., a job description from https://www.ziprecruiter.com/c/Get-Set-Go-Logistics-LLC/Job/Amazon-Delivery-Associate-Get-Set-Go-Logistics/-in-Woodside,NY?jid=c61cb43977436625&lvk=w3zOaeu9mIP8tyA2XtTVVQ.--N0rhCvOEN