Zhou, Ze; Roncoli, Claudio; Sipetas, Charalampos

Optimal matching for coexisting ride-hailing and ridesharing services considering pricing fairness and user choices

Published in:
Transportation Research Part C: Emerging Technologies

DOI:
10.1016/j.trc.2023.104326

Published: 01/11/2023

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY

Please cite the original version:
Optimal matching for coexisting ride-hailing and ridesharing services considering pricing fairness and user choices

Ze Zhou *, Claudio Roncoli, Charalampos Sipetas

Department of Built Environment, Aalto University, Espoo 02150, Finland

ARTICLE INFO

Keywords:
Dynamic ridesharing
Fleet operation
Pricing
Fairness

ABSTRACT

Mobility-on-demand (MoD) has the potential to revolutionise the patterns of urban mobility. Typically, an MoD platform provides both ride-hailing and ridesharing services, exacerbating the challenges of operating a city-scale real-time MoD system. Existing studies assume that travellers are fully compliant with the platform’s decisions regarding pricing and vehicle assignments, whereas, in reality, travellers can choose different modes based on monetary costs and travel experience, which may conflict with the results derived from the system perspective. In this study, we relax this assumption by accounting for pricing fairness and the travellers' modal choices within a framework designed to optimise vehicle–traveller matching when both ride-hailing and ridesharing services are provided by an MoD platform. Six fairness principles are defined to characterise fair pricing for shared rides. Computationally efficient optimisation problems are formulated accounting for co-existing ride-hailing and ridesharing services. In numerical experiments, we assess the effectiveness of our method and compare it with state-of-the-art ones using a dataset of taxi requests for New York City. The results show that our optimisation strategy can significantly increase the service ratio and profit without sacrificing the service quality.

1. Introduction

Mobility-on-demand (MoD) is a promising trend for providing affordable and efficient options to travellers. Instead of owning a private vehicle, travellers can easily access an on-demand service, which contributes to reducing car ownership and therefore improving urban mobility (Butler et al., 2020). Common MoD services that are making significant progress in many cities (Wang et al., 2019) include elements of ride-hailing and ridesharing, which are usually offered by private ride-sourcing companies (e.g., Uber, Lyft, DiDi). Although on-demand services are embraced by travellers due to their convenience, the providers of MoD services face various challenges ranging from addressing controversial congestion issues to daily operation problems (Amey et al., 2011) in order to ensure cost-effectiveness and sustainability.

Ridesharing offers opportunities for achieving the goals of cost-effectiveness and sustainability since it can improve the vehicle occupancy rate by allowing one vehicle to serve multiple travellers simultaneously. Deciding which travellers should share a trip depends on the spatiotemporal similarities of their trips. Even though travellers sacrifice time and comfort, they benefit from sharing the travel fare with their co-riders. Therefore, ride-sourcing companies commonly provide shared services to users as an economic option that complements the conventional ride-hailing non-shared trip service (Liu et al., 2021a).

Ridesharing can be organised either statically or dynamically, where static ridesharing refers to designing schedules for a group of travellers with all trip-related information known beforehand, while dynamic ridesharing deals with this information being disclosed...
over time (Luo et al., 2019). Compared to static, offering dynamic ridesharing services is a more challenging task, mostly in terms of matching travellers with vehicles and implementing a proper pricing scheme. Existing literature in the field includes several studies that aim at assisting the development of efficient dynamic ridesharing systems. Some of them make fundamental assumptions, e.g., that all travellers are willing to accept the suggested ride and fare (see, e.g., Santi et al., 2014). Others utilise heuristic algorithms to solve complex mathematical problems that result from accounting for travellers’ preferences (see, e.g., Haliem et al., 2021). Many studies in the field of dynamic ridesharing focus on studying matching and pricing schemes separately, although literature concludes that joint optimisation is promising for significant performance improvement (Özkan, 2020). Although some fairness-aware pricing methods, such as sectional (Zhang et al., 2020) or detour-based (Liu et al., 2021b) pricing, account for the shared trip distance or detour respectively, they fail to incorporate all the features of ridesharing into their models. Note that fairness in pricing can be defined from different perspectives. The fairness among ride-sharing travellers refers to a scheme in which a traveller who has a larger detour rate or shares with more travellers should receive a larger discount. All possible fairness principles are further defined from different perspectives. The fairness among ride-sharing travellers refers to a scheme in which a traveller who has a larger detour rate or shares with more travellers should receive a larger discount. All possible fairness principles are further summarised in Section 3.2.1.

Usually, an MoD platform operates fleets composed of more than one type of vehicle and provides both ride-hailing and ridesharing services simultaneously. For instance, Didi Chuxing Technology, LLC (2022) operates Express, Premier, and DiDi Hitch, whereas Uber Technologies, LLC (2022) provides both UberX Share and Uber Comfort services. By operating a mixed fleet of two or more types of vehicles, such on-demand mobility platforms could better satisfy different travel needs. In this paper, we assume that both ride-hailing and dynamic ridesharing services are provided by the proposed MoD platform. Besides, we propose a framework for optimising the daily operation of a dynamic MoD platform. The problem of matching vehicles with travellers under fair pricing schemes is considered, also taking into account travellers’ modal choices. The proposed framework aims at assisting shared MoD services to achieve high performance in terms of both providers’ and users’ costs. The former benefit from better utilisation of their fleet and drivers, among others, while the latter enjoy the benefits of an improved service that allows users to choose how much time and comfort they are willing to give up in exchange for fairly lower travel fares. Notably, most studies (Syed et al., 2021; Liu et al., 2021a; Jiao and Ramezani, 2022) in this field focus on maximising providers’ profit. Our study acknowledges the realistic need for ensuring profit for the service providers, but also considers users’ and societal needs, since a sole profit-oriented objective may deteriorate the system’s performance in terms of service ratio and quality. Moreover, there may be incentives, discounts and taxes that are imposed by policymakers or transport authorities aiming to provide more sustainable mobility services. Therefore, it contributes to making MoD services a valuable actor in a mobility ecosystem that requires efficient and sustainable solutions for constantly growing mobility needs.

1.1. Literature review

According to Fleckenstein et al. (2023), MoD belongs to generalised integrated demand management and vehicle routing problem. In systems where customers place orders online and services are provided offline, the efficient operation of a vehicle fleet is vital to ensure profitability for such services. There are numerous studies about ridesharing system operation in terms of matching and pricing. Several works (see, e.g., Santi et al., 2014; Alonso-Mora et al., 2017; Simonetto et al., 2019) proposed efficient algorithms for the real-time high-capacity large-scale ridesharing dispatching problem. Beirigo et al. (2022) proposed a method to handle the real-time dynamic ridesharing operation problem accounting for heterogeneous user classes and their needs for different service quality. The proposed method can significantly improve user satisfaction compared to conventional ridesharing systems. Besides, Pandey et al. (2019) deal with the competition among the ridesharing mobility providers, comparing the cooperative and competitive model with the baseline centralised model and finding that competition between companies can lower service quality. However, all these studies simply assume that travellers always accept the suggested ride even without considering the trip fare, which might deteriorate the system performance if this assumption does not hold (e.g., if the travellers decide to switch to another mode). In contrast, Engelhardt et al. (2022) introduced a simulation framework in which travellers are allowed to choose their travel mode, to assess the impact of competition and cooperation among multiple ridesharing providers. Their findings revealed that regulated competition benefits all stakeholders, including operators, cities (in terms of pooling efficiency), and customers. In contrast, the absence of regulation resulted in the lowest pooling efficiency and operator profit.

In another research branch, dynamic pricing is applied as a tool to maintain demand–supply market equilibrium (Liu and Li, 2017; Ke et al., 2020; Zhang and Nie, 2021). Meanwhile, surge and dynamic pricing models are prevailing in ridesharing. However, such pricing methods are designed for MoD platforms with freelance drivers and a two-sided market. Moreover, employing an opaque dynamic pricing algorithm and fare may confuse and annoy the ridesharing customers, thus preventing them from making a quick ride request or even using the system in the future (Yang et al., 2020; Tervo and Väyrynen, 2022). Furthermore, Haliem et al. (2021) designed a model-free demand-aware, dynamic ridesharing framework that unifies vehicle–traveller matching, pricing, and route planning. The framework allows customers to accept or reject ride offers based on their preferences, while a Deep Q-network is utilised to predict the future supply-and-demand distribution and thus update the initial price. However, only heuristic algorithms are employed and the matching and pricing problems are solved separately.

Other studies introduced the concept of stable matches in shared rides and set prices to attain such a stable state. Lu and Quadrifoglio (2019) formulate the sustainable ridesharing service design as a fair cost allocation problem using cooperative game theory. Foti et al. (2021) argue that the system optimum matching differs from fair matching where everybody is fully satisfied with her/his assignment. Fielbaum et al. (2022) propose three different protocols to split the cost of a ridesharing trip such that the optimal solution is also an equilibrium. Besides, a cost-sharing mechanism design is commonly used to set prices in an on-demand system. For instance, Chau et al. (2022) study various fair cost-sharing mechanisms and the corresponding stable matching results.
The results show that the induced social costs of fair matching are as comparably low as the social optimal outcomes. Furuhat et al. (2014) design cost-sharing mechanisms to update fare quotes in demand-responsive transport systems so that all travellers are treated fairly. Hu et al. (2021) extend Furuhat’s work by including driver’s costs and providing incentives to encourage travellers to submit their trip requests earlier. Nonetheless, such methods are not designed to provide ridesharing passengers with a fair price in real-time, which requires a computationally efficient solution for handling demand in large-scale scenarios.

Instead of investigating pricing and matching mechanisms separately, Özkan (2020) argues that joint optimisation of the pricing and matching in ridesharing systems can lead to significant performance improvement. Nonetheless, the macroscopic ridesharing model is formulated at the market level; thus, it is not applicable to the real-time operations of ridesharing systems at the trip level. Similarly, Yan et al. (2020) show that integrating matching with dynamic pricing under market equilibrium could reduce price and its volatility, thus increasing the system performance in terms of capacity utilisation, trip throughput, and welfare. In contrast, Liu et al. (2021a) propose a framework for jointly optimising the matching and pricing in MoD systems at the trip level. Although the challenge is amplified because of mixed fleets and stochastic traveller behaviour, a computationally efficient method for solving the integrated problem both in sequence and in batches is developed. However, the tractability of the method is challenged when the number of ridesharing travellers is greater than two. Kucharski and Cats (2020) proposed an efficient algorithm capable of identifying shared rides from large-scale travel demand for the purpose of evaluating strategic scenarios. Moreover, Jiao and Ramezani (2022) propose a dynamic discount pricing strategy to incentive customers to take shared rides; however, they assume that partially occupied vehicles are not considered for assignment and a maximum of two travellers share a trip. Azadeh et al. (2022) investigated the choice-driven dial-a-ride problem, taking into account pricing and passenger choice behaviour. To address this problem, they formulated a mixed integer linear program and developed a customised algorithm to solve it. Nonetheless, the computational efficiency of the proposed model is questionable when dealing with large-scale travel demand and fleet size.

In most shared mobility literature where system efficiency and profit are generally emphasised, ridesharing travellers are usually assumed to accept whatever the platform offers. This is, however, a strong assumption considering that travellers in reality can decline a ride offer and turn to other modes if the price is inappropriate. If we relax this assumption, then it is vital to present travellers with a customer-centred fare plan, which accounts for their detours and other ridesharing effects. However, most studies in existing literature have focused on maximising only the platform profit when it comes to pricing, thus ignoring fairness from the traveller’s viewpoint (Cao et al., 2021). Notice that the customers of the on-demand system are usually cost-sensitive and participate in shared trips in the hope of saving travel costs (U.S. Department of Transportation, 2017; Bulteau et al., 2019). Therefore, achieving customer satisfaction is not a trivial task, especially when the system accounts for travellers’ mode choice behaviour. If a customer finds the fare to be unfair, for instance, if the fare for ridesharing is higher than travelling alone or if they experienced a longer detour yet paid more than those who detoured less, they may decline the offer or even refuse to use the service again. Thus, a fairness-aware pricing scheme is critical to the sustainability of a ridesharing system. In this respect, there are two similar works. First, Zhang et al. (2020) develop algorithms to address fair pricing and matching challenges in dynamic ridesharing systems. Second, Liu et al. (2021b) propose a novel taxi ridesharing scheme that considers both online and offline travellers, introducing a payment model to fairly split the ridesharing benefits among travellers and drivers based on their detour rate. However, these studies do not provide a comprehensive fairness-aware pricing method for dynamic ridesharing, which should account for all potential fairness principles associated with ridesharing trips. In the numerical experiments section, we compare our fair pricing method with these two, demonstrating the achievable improvements.

To sum up, the motivation of this study is twofold and refers to the needs for (1) a complete ridesharing strategy that integrates pricing with matching in the context where the platform accounts for travellers’ mode choice decisions; and (2) a more flexible and fair fare structure for real-time dynamic ridesharing pricing that satisfies ridesharing travellers.

1.2. Contributions

In this paper, we target an MoD platform’s daily operation problem of how to match vehicles with travellers and price them in a dynamic ridesharing context where travellers can choose from ridesharing or non-shared (ride-hailing) services. The main contributions are summarised as follows.

- We design an MoD framework that enables the optimisation of vehicle–traveller matching while considering pricing fairness for co-existing ride-hailing and ridesharing services.
- We summarise six principles that a fair pricing method should satisfy and propose a fairness-aware dynamic ridesharing pricing method that features all the desired properties.
- We formulate a fair-price based vehicle–traveller matching operation problem that accounts for the travellers’ mode choice decisions, in the form of a linear assignment problem, which can be solved efficiently.

The current work builds upon a preliminary version published by Zhou and Roncoli (2023). We extend it by reformulating our methodology, integrating two service types into one optimisation problem and employing a real taxi dataset from New York City (NYC) to test the proposed methods. The remainder of this paper is structured as follows. Section 2 introduces the three components of an MoD system, as well as the interactions among them. In Section 3, we develop a fairness-aware pricing scheme and formulate optimal matching problem for co-existing ride-hailing and ridesharing services. Section 4 summarises the simulation setups for two series of numerical experiments. Then, we present the results of the NYC scenario in Section 5, also comparing the performance of the proposed pricing method with state-of-the-art ones. Next, in Section 6, we conduct sensitivity analysis experiments to reveal the effectiveness and the limitations of the proposed optimisation methods. Finally, the main findings and future research directions are summarised in Section 7.
2. MoD system

There are three main components in an MoD system, namely travellers, vehicles, and the MoD platform. In this study, we assume that the MoD platform manages a fleet of vehicles. We describe the interactions among these three components and elaborate on each component in the following sections.

2.1. System interactions

As shown in Fig. 1, new travellers submit their requests specifying the origin, destination, and latest pick-up and drop-off times. Together with previous unmatched requests, the platform is responsible to check if the travellers’ waiting time exceeds their maximum waiting time limits. Except for those travellers who leave the system because of long waiting times, the remaining requests are forwarded for matching. We further illustrate the matching process in detail in Fig. 2. The unmatched travellers return to the waiting pool and wait until the next round of matching. For the successfully matched travellers, the platform generates a service menu, which is comprised of discounts and detour time for a shared trip and trip fee, and expected waiting time for both solo (ride-hailing) and shared trips.

Upon receiving the service menu, travellers instantly choose the service based on their utilities. If a traveller chooses the ride-hailing service, then the trip offer comes into effect immediately. However, if a traveller chooses the ridesharing service, we assume that the operator would conduct a fairness check before confirming the offer.
On the other hand, the traveller might not be satisfied with either trip offer. In this case, the traveller actively rejects the offer and exits the system. In summary, there are three possible outcomes after the travellers submit their requests: (a) leaving the platform because their maximum waiting time is violated (b) rejecting the ride offer (no matter actively or passively) and exiting the platform immediately, or (c) choosing a service from the offered menu and getting served.

2.2. Heterogeneous travellers

Travellers are assumed to be cost and service-quality sensitive, with different levels of willingness to share, which are reflected by utility coefficients. When provided with an explicit service menu, each traveller chooses the mode that maximises her or his utility \( u_p \). Similarly to Jiao and Ramezani (2022), we employ the following formulations to measure the perceived utilities (Deaton and Muellbauer, 1980) of different modes to the individual traveller \( p \):

\[
\begin{align*}
    u^h_p &= \beta^h_p - \beta^h_p u^h_p - \beta^h_p f^h_p \\
    u^o_p &= \beta^o_p - \beta^o_p (u^o_p + \delta_p) - \beta^o_p f^o_p \\
    u^o_p &= \beta^o_p,
\end{align*}
\]

where \( \beta^h_p, \beta^o_p, \) and \( \beta^o_p \) are parameters to reflect the inherent trip qualities, e.g., safety and ride comfort of ride-hailing, ridesharing, and other travel modes, respectively; \( \beta^h_p \) and \( \beta^o_p \) are the utility coefficients to convert trip time and cost to the traveller’s perceived utility; \( u^h_p \) and \( f^h_p \) denote the traveller’s waiting time and trip fee if transported by a ride-hailing vehicle; while \( u^o_p, \delta_p \) and \( f^o_p \) indicate the waiting, detour time, and shared trip fare if served by a ridesharing vehicle.

Since travellers can select between ride-hailing and ridesharing mobility services, as well as leave the platform by turning to other modes, such as public transport or private vehicles, we utilise a multinomial logit model (Ben-Akiva and Lerman, 1985) to capture the travellers’ choice. Hence, the probability for traveller \( p \) to select a travel mode \( m \) is:

\[
Pr^m_p = \frac{\exp(u^m_p)}{\sum_{k \in \{h, o, o\}} \exp(u^k_p)}
\]

Note that in the current mode choice formulation (Eqs. (1)–(3)), a longer trip with larger travel time results also in a larger trip fare, leading to a smaller value of \( u^h_p \) and \( u^o_p \). Then, since the alternative modes are independent of the trip distance, services offered for long trips have a higher possibility of being rejected by travellers who may turn to alternative modes. This limitation could be overcome by extending the model to include other specific travel modes by utilising an explicit formulation that accounts for trip distance and duration. Finally, as the individual parameters are not known by the platform, due to inherent privacy issues, it is reasonable to assume that the platform knows only population-level estimates of the probability distribution of such parameters.

2.3. Vehicles

In this study, we consider two types of vehicles: ride-hailing and ridesharing vehicles. A ride-hailing vehicle is only assigned to solo trips, while ridesharing vehicles are allowed to serve one or multiple travellers at a time. Each vehicle is assumed to follow the platform instructions by travelling through a sequence of pick-up and drop-off nodes, denoted as a route. Whenever a traveller chooses a type of service and confirms the order, the vehicle route is updated simultaneously. Once a vehicle arrives at its last destination (i.e., when the last assigned traveller is dropped off), it parks nearby to avoid empty driving until reassigned to new travellers. Therefore the vehicle can only be either idle or on duty. The vehicle state, including the number of travellers on board, current path, and schedule are all accessible to the platform in real-time.

2.4. Mobility platform

The platform collects data regarding requests and service vehicle availability in a service area and performs calculations to (a) serve more travellers, and (b) maximise profit. In Fig. 2, we show how the platform processes travel requests and vehicle data and it outputs the so-called service menu. The service menu refers to the selected vehicles along with the trip information that is delivered to each potential traveller, such as wait and travel time, fare and discount. An initial fare is communicated to travellers before they choose a type of service and confirm the order, the vehicle route is updated simultaneously. Once a vehicle arrives at its last destination, it parks nearby to avoid empty driving until reassigned to new travellers. Therefore the vehicle can only be either idle or on duty. The vehicle state, including the number of travellers on board, current path, and schedule are all accessible to the platform in real-time.
problem. The formulated problem(s) can be solved very efficiently even for large-scale problems (Simonetto et al., 2019; Zhou and Roncoli, 2021, 2022), while it has been demonstrated to not affect significantly the performance. Nonetheless, re-assigning travellers to other vehicles is not allowed in this study. Consequently, the length of the tour only increases, leading to fares that are cheaper than the originally communicated fare.

3. Optimal matching for co-existing ride-hailing and ridesharing services

In this section, we formulate the optimal matching problem of ride-hailing and ridesharing trips with or without predicting travellers’ mode choices formulated as Integer Linear Programs (ILPs). The ride-hailing and ridesharing schedules are determined separately by the first two ILPs. Furthermore, the third one unifies the two service types and formulates them into one ILP.

3.1. Ride-hailing trips

For ride-hailing trips, each feasible traveller–vehicle pair is associated with a travel profit which is the difference between the trip fare and cost. We use a typical static fare structure adopted by taxis and UberX service (Uber, 2018) to calculate the ride-hailing trip fare $f_p^h$:

$$f_p^h = f_0 + f_t t_p + f_d d_p,$$

where $f_0$, $f_t$, and $f_d$ denote base, time, and distance fare rates, respectively, while $t_p$ and $d_p$ are the ride-hailing trip time and distance, respectively. Hence, we formulate the optimal vehicle–traveller matching for ride-hailing trips as the following ILP:

$$\max \sum_p \sum_v \delta (f_p^h - e^{h,v}_x) x_{p,v} - \gamma (|P| - \sum_p \sum_v x_{p,v})$$

subject to

$$\sum_v x_{p,v} \leq 1, \quad \forall p$$

$$\sum_p x_{p,v} \leq 1, \quad \forall v \in V_p$$

$$x_{p,v} \in \{0, 1\},$$

where $V_p$ is the available vehicle set for traveller $p$; $\delta$ is a parameter designed to scale the ride-hailing trip profit; $e^{h,v}_x$ denotes the trip cost, such as energy consumption, for vehicle $v$ to serve traveller $p$; $x_{p,v}$ is a binary variable, which equals 1 if vehicle $v$ is assigned to serve traveller $p$ and 0 otherwise; $\gamma$ is defined as the penalty cost for each unserved customer; and $|P|$ denotes the number of travellers that wait to be served. Therefore, the multiojective function (6) jointly maximises the profit and accounts for a penalty function for unserved travellers. Notice that $V_p$ is a subset of the entire fleet and is obtained from the batching and searching step. This subset facilitates the subsequent routing and optimisation procedure while reducing computational complexity. Both (7) and (8) ensure that a one-to-one match is achieved. Note that we introduce the parameter $\delta$ to provide flexibility for the operator to control which type of service should be promoted. Besides, $\delta$ only appears in the objective function and is not involved in the real profit calculation. We further discuss the effect of $\delta$ on prioritising which type of MoD service in Section 6.

By solving this profit-maximising linear assignment problem, the system determines the optimal dispatch plan for which traveller $p$ should be assigned to which ride-hailing vehicle $v$. We refer to this formulation as HVO, which is the acronym for ride-hailing Vehicle optimisation. Nonetheless, the formulation (6) overlooks the traveller’s mode choice decision.

3.2. Ridesharing trips

3.2.1. Fairness principles

The ridesharing optimisation problem can be defined by the same structure as the HVO model (Simonetto et al., 2019). The major challenge is defining a policy for each rider’s fare and updating it if a new traveller enters and shares the trip in a vehicle that has travellers already assigned. We first define the following principles to evaluate whether a pricing method is fair to each traveller.

- (P1) Price upper bound: for each ridesharing service user, the ride fee should be strictly less than the ride-hailing trip fare $f_p^h$ if she or he travels alone.
- (P2) Detour fairness: co-riders who have a larger detour should have a higher discount.
- (P3) Ridesharing benefit: the more travellers sharing the ride, the higher the discount ratio, resulting in less to be paid by each traveller.
- (P4) On-board traveller rationality: in a dynamic ridesharing context, vehicle routes may change frequently as new passengers show up. Consequently, the trip fare for previously assigned or already on-board passengers should be updated accordingly and must be lower than the previous quote. Otherwise, onboard travellers would not let a new traveller in.
- (P5) On-board traveller’s inconvenience cost: passengers who are already on board are less likely to accept the new offers when there are already many co-riders. This is expected because each new coming rider would inevitably bring inconvenience costs for the ones already on board, such as pick up and waiting time.
- (P6) Platform profit: The total fare paid by all travellers must be higher than the trip cost.
3.2.2. Discount function

We now proceed by designing a discount function \( \Phi(\cdot) \) that satisfies the aforementioned principles. Given a multiplicative discount \( \Phi(\cdot) \), the ridesharing price for each co-rider is

\[
\hat{f}_p = \Phi_p \cdot f^i_p,
\]

(10)

where \( f^i_p \) denotes the previous price before the new traveller joining. Since \( \Phi_p \) is strictly less than one, the updated price \( \hat{f}_p \) is always less than the previous price \( f^i_p \), which ensures that there is always an incentive for on-board passengers to allow ridesharing. Note that the initial fare for a passenger is also computed using (10), by replacing the previous fare \( f^i_p \) with the ride-hailing fare \( f^h_p \), derived from (5). As the discount is normally related to the traveller’s detour rate, the number of co-riders, and shared trip distance, we introduce two indicators that account for these quantities.

We first define the detour index \( d^d_p \), which is the ratio between the traveller’s detour due to the insertion of the new traveller pick-up and drop-off and the updated travel distance after insertion, namely

\[
d^d_p = \frac{d^u_p - d^s_p}{d^s_p}.
\]

(11)

where \( d^u_p \) denotes the updated travel distance after insertion, while \( d^s_p \) is the path length before insertion. For those already on board, \( d^u_p \) is the path length before inserting the new traveller’s origin and destination, while the updated travel distance is calculated after the insertion, as shown in Fig. 3. For the new traveller, \( d^u_p \) is the shortest path. In addition, we choose \( d^u_p \) to be the denominator in order to ensure that \( d^d_p \leq 1 \). Since each traveller has a maximum detour limit, we formally define the maximum detour distance as \( d^\text{max}_p = m \cdot d^d_p \), where \( m \) is the maximum detour rate. Hence we have \( d^d_p \in \left[0, \frac{m}{n+1}\right] \). Notice that the larger \( d^d_p \), the longer detour experienced by the traveller \( p \).

Furthermore, we define \( \zeta^r_p \) as the ridesharing index which reflects the distance and the number of travellers of shared rides experienced by traveller \( p \). A shared ride is composed of several nodes, which refer to co-riders’ pick-up or drop-off locations. Hence, a shared ride can be divided into legs, which are the parts of the shared ride between two nodes. For each leg \( \lambda \), we calculate the travel fare by \( f^\lambda_p = f^u_p \cdot t^\lambda + f_s \cdot d^\lambda \). Nonetheless, if a leg is shared by several co-riders, we divide the leg fare by the number of co-riders \( k^\lambda_p \) in this leg and name it \( \text{shared fare} f^\lambda_p = f^u_p / k^\lambda_p \). We introduce a binary parameter \( y_{p,\lambda} \), which is equal to 1 if traveller \( p \) travels through leg \( \lambda \) and 0 otherwise. Then we define

\[
\zeta^r_p = 1 - \frac{\sum_{\lambda} f^\lambda_p \cdot y_{p,\lambda}}{\sum_{\lambda} f^h_p \cdot y_{p,\lambda}}
\]

(12)

as the ridesharing index, which can reflect the variations of the number of co-riders and the length of the shared travelled distance. Note that a longer portion of the trip being shared and more co-riders sharing the trip lead to higher \( \zeta^r_p \). Based on (12), we can
further derive that \( \zeta_p^r \in [0, \frac{1}{n}] \), where \( n \) represents the vehicle traveller capacity. The lower bound 0 indicates that traveller \( i \) travels alone throughout the entire trip, while the upper bound shows that maximum \( n \) travellers travel together from their common origin to the same destination.

Eventually, we are ready to define a discount function of two variables \( \Phi(\tau_p^d, \zeta_p^r) \), which should satisfy the following properties:

\[
\begin{align*}
\Phi_p & \in [\Phi_{\text{min}}, \Phi_{\text{max}}], \text{ where } 0 < \Phi_{\text{min}} < \Phi_{\text{max}} < 1 \\
\frac{\partial \Phi}{\partial \tau} & < 0, \frac{\partial^2 \Phi}{\partial \tau^2} < 0 \\
\frac{\partial \Phi}{\partial \zeta} & > 0, \frac{\partial^2 \Phi}{\partial \zeta^2} > 0.
\end{align*}
\]

Notice that the first property is associated with previously defined P1 and P6. The second property guarantees that P2, P3, and P4 are satisfied, because the larger \( \tau_p^d \) and \( \zeta_p^r \), the more detours and more co-riders and shared distance. Therefore the \( \Phi \) is monotonically decreasing with the increase of \( \tau_p^d \) and \( \zeta_p^r \). Lastly, the second-order partial derivative strictly greater than zero reveals that function \( \Phi \) is a convex function. Compared to a linear function, \( \Phi \) declines faster at first and slow down with the increase of both independent variables. This property assures that co-riders are welcomed and benefit more from ridesharing when the vehicle is almost empty. In contrast, the customer who joins the shared rides when capacity is almost full would get less discount. This discourages too many travellers from being assigned to a single vehicle, such that the P5 is satisfied.

In this paper, we formulate a specific discount function that satisfies all the aforementioned properties, although other functions that comply with the three previously stated properties, i.e., that satisfy the predefined principles, can be employed. We select the exponential function \( y = e^{-x} \), which is monotonically decreasing and convex in the range of \([0,1]\). In the same range, the \( y \) value decreases from 1 to \( 1/e \approx 0.368 \), making it a suitable candidate for a discount function. By replacing the \( x \) with the sum of detour \( \tau_p^d \) and ridesharing \( \zeta_p^r \) index, the ultimate discount rate is

\[
\Phi_p = \exp(-a(\tau_p^d + \zeta_p^r)) - b,
\]

where the discount coefficient \( a \) is designed to control the level of discount and the base rate \( b \) is defined so that \( b > 0 \) if the vehicle is empty and \( b = 0 \) otherwise, in order to guarantee that all the ridesharing users can gain benefits even if they end up travelling alone.

### 3.2.3. ILP for ridesharing trips

Furthermore, we introduce the ridesharing profit \( \pi_{p,v} \), defined as the marginal profit whenever a new traveller enters, which is calculated via

\[
\pi_{p,v} = \left( \sum_{p \in P} f_p^i - C_{p,v}^i \right) - \left( \sum_{p \in \hat{P}} \hat{f}_p^i - C_{p,v}^{\hat{i}} \right),
\]

where \( P \) is the set of all assigned travellers except those who have arrived at their destination, \( \hat{P} \) represents the same set before the current traveller is joining, \( C_{p,v} \) and \( C_{p,v}^{\hat{i}} \) denote the entire trip cost for serving \( \hat{P} \) and \( P \). Therefore, the objective function of ridesharing trips is:

\[
\max \sum_{p \in P} \sum_{v \in P_p} \pi_{p,v} x_{p,v} - \zeta(|P|) \sum_{p \in P} x_{p,v}.
\]

Combining (15) with (7)–(9), we have a new ILP formulation for ridesharing trips, which is referred to as SVO. The HVO and SVO can be solved in parallel to determine alternatives in the service menu, which is forwarded to travellers and awaiting for final confirmation. Due to the HVO and SVO being solved separately, we refer to the combination of them as individual strategy.

Both HVO and SVO can be classified as the well-known linear assignment problem (Kennington and Helgason, 1980; Papadimitriou and Steiglitz, 1998). One prominent characteristic of the linear assignment problem is that its constraint matrix is a totally unimodular matrix, which guarantees the exactness of the relaxation of the binary variables \( x_{p,v} \) to continuous ones. Additionally, the constraint matrix is sparse due to the candidate vehicle searching process, as each customer is assigned to only a few vehicles. Exploiting the properties of unimodularity and sparsity, state-of-the-art solution algorithms, such as the auction algorithm (Bernard et al., 2016), are capable of solving large-scale instances in real-time (Bertsekas and Castanon, 1989; Simonetto et al., 2019). Therefore, the proposed model exhibits excellent scalability.

### 3.3. Integrated strategy

The two ILPs, HVO and SVO, proposed earlier determine not only the most convenient ride-hailing and ridesharing trips separately but also fail to incorporate the travellers’ mode choice decisions. To account for the travellers’ mode choice when provided with both ride-hailing and ridesharing service menus, we formulate a problem to find the best ride-hailing–ridesharing pair for each traveller.

Specifically, we calculate the probability of choosing ride-hailing or ridesharing vehicles, denoting them as \( Pr_{s,h}^{v} \) and \( Pr_{s,h}^{r} \), respectively. Note that, once potential ride-hailing and ridesharing vehicles are determined, it is straightforward to calculate \( Pr_{s,h}^{v} \) and \( Pr_{s,h}^{r} \).
and \( \Pr_{p,h,s} \) based on (4). Therefore, the expected profit for serving traveller \( p \) when he or she is provided with ride-hailing vehicle \( h \) and ridesharing vehicle \( s \) can be written as:

\[
E(\pi_{p,h,s}) = \Pr_{p,h,s}^1 \cdot \delta \cdot (f_{p,h} - c_{p,h}) + \Pr_{p,h,s}^2 \cdot \pi_{p,s},
\]

where \( \pi_{p,s} \) is calculated via (14). Hence, we can now formulate the integrated optimisation problem as:

\[
\max_s \sum_p \sum_{h \in H_s} \sum_{s \in S} E(\pi_{p,h,s}) x_{p,h,s} - r(P) - \sum_p \sum_{h \in H_s} x_{p,h,s}
\]

subject to

\[
\sum_p \sum_{h \in H_s} x_{p,h,s} \leq 1, \quad \forall p
\]

\[
\sum_p \sum_s x_{p,h,s} \leq 1, \quad \forall h \in H
\]

\[
\sum_p x_{p,h,s} \leq 1, \quad \forall s \in S
\]

\[
x_{p,h,s} \in \{0, 1\}
\]

To distinguish from individual strategy, we refer to the above formula system (17)–(21) as integrated strategy. Instead of solving ride-hailing and ridesharing problems separately, the integrated strategy pairs every potential ride-hailing vehicle with every ridesharing vehicle and calculates the corresponding ride-hailing–ridesharing probability pair. Eventually, a ridehailing–ridesharing vehicle pair is determined and delivered to customers along with other trip information.

However, by integrating ride-hailing and ridesharing vehicles into a single model, the problem (17)–(21) falls within the realm of three-dimensional or three-index assignment problem, which is known to be NP-complete (Frieze, 1983). In contrast to the linear assignment problem, there is no guarantee of an efficient solution algorithm for the three-dimensional problem. The branch and bound method (Balas and Saltzman, 1991) and heuristic algorithms are commonly used to solve the problem (Burkard and Cela, 1999). Nonetheless, we demonstrate in Section 5.3 and Appendix C that our specific formulation of the integrated model can be solved relatively fast and is applicable to real-time scenarios.

4. Experimental setup

In order to demonstrate and test the proposed methodology, we introduce two case studies: (a) a large-scale NYC scenario; and (b) a limited-size grid network that is employed for a series of sensitivity analysis experiments. Both the NYC and the grid network experiments are coded in Python 3.10. The NYC simulation scenarios are run on a shared server with 40 cores and 768 GB of memory. The grid network simulations are implemented in an i7-8750H 2.20 GHz computer with 8 GB RAM. The Gurobi optimiser (Gurobi Optimization, LLC, 2022) is employed to solve optimisation problems.

4.1. NYC network

4.1.1. Simulation setup

In this section, we utilise a processed dataset of taxi trips (Lesmana et al., 2019) for the New York Metropolitan region, shown in Fig. 4, to test the proposed methodology. The demand is extracted as a 2-hour evening peak data. The original dataset (Donovan and Work, 2016) is preprocessed to exclude invalid records, such as missing coordinates or records with either too short or too long travel time. Eventually, we generate 51,491 traveller requests with real OD pairs and time windows. Note that the original dataset only contains departure time and origin–destination information, while we synthesise the remaining quantities, including the traveller’s maximum waiting time, detour time, time and fare utility coefficients, as well as the utility constant for different modes (i.e., ride-hailing, share, and other), by extracting them from truncated normal distributions. Notice that each time a new traveller’s origin and destination are inserted into a vehicle’s route, we check the updated route’s feasibility using the following Eq. (22):

\[
r_d = r_s + g_r + \omega + h
\]

where \( r_d \) is the travel request deadline, which means that the traveller must arrive at her/his destination before it; \( r_s \) is the request submission time; \( g_r \) is the shortest travel time of the request; \( \omega \) and \( h \) are the traveller’s maximum waiting time and the acceptable slack time, respectively, where the latter term is introduced to provide additional time flexibility for ridesharing. The population mean \( \mu \) (the utility coefficients are based on Habib (2019) and Lavieri and Bhat (2019)), standard deviations \( \sigma \), and lower and upper bounds for these normal distributions are presented in Table 1. Similar to Jiao and Ramezani (2022), we only use these population mean values in the choice model when the operator tries to predict the travellers’ mode choice. However, each traveller in the experiments makes decisions based on the personalised coefficients package, which includes maximum wait and acceptable slack

1 A large amount of NYC taxi data is available at https://www.nyc.gov/site/tlc/about/tlc-trip-record-data.page.
Fig. 4. The network used in the NYC experiments. (a) Map of the city network (source: Google Map); (b) Sketch of the Manhattan network used in our case study.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
<th>Unit</th>
<th>$\mu$</th>
<th>$\sigma$</th>
<th>lb</th>
<th>ub</th>
</tr>
</thead>
<tbody>
<tr>
<td>$w$</td>
<td>Maximum waiting time</td>
<td>s</td>
<td>390</td>
<td>30</td>
<td>300</td>
<td>480</td>
</tr>
<tr>
<td>$h$</td>
<td>Acceptable slack time</td>
<td>s</td>
<td>420</td>
<td>60</td>
<td>300</td>
<td>540</td>
</tr>
<tr>
<td>$\beta_t$</td>
<td>Utility coefficient for travel time</td>
<td>1/min</td>
<td>0.6</td>
<td>0.05</td>
<td>0.5</td>
<td>0.7</td>
</tr>
<tr>
<td>$\beta_f$</td>
<td>Utility coefficient for trip fare</td>
<td>1/$$</td>
<td>3.2</td>
<td>0.2</td>
<td>2.8</td>
<td>3.6</td>
</tr>
<tr>
<td>$\beta_c$</td>
<td>Utility constant for ride-hailing trips</td>
<td>–</td>
<td>0</td>
<td>0</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>$\beta_s$</td>
<td>Utility constant for shared trips</td>
<td>–</td>
<td>–0.8</td>
<td>0.1</td>
<td>–1</td>
<td>–0.6</td>
</tr>
<tr>
<td>$\beta_o$</td>
<td>Utility constant for other modes</td>
<td>–</td>
<td>–30</td>
<td>2</td>
<td>–35</td>
<td>–25</td>
</tr>
</tbody>
</table>

4.1.2. Dynamic ridesharing pricing and optimisation strategies for comparison

As part of investigating the performance of the proposed methodology, it is important to compare (a) the two optimisation approaches presented in Section 3, i.e., the individual and the integrated; and (b) the proposed pricing scheme with well-established strategies discussed in Section 1.1. In order to achieve these comparisons, four scenarios for experimental analyses have been
developed. The first two refer to state-of-the-art pricing schemes and individual optimisation, hence representing the current status quo in the field of optimising MoD services. The following two scenarios incorporate the fairness-aware pricing scheme that is developed in this study, with scenario S3 adopting individual and scenario S4 joint optimisation. The four different strategies are compared using the NYC dataset. In Table 3, the differences among these strategies are explicitly described. It is shown that the same batching and routing methods are employed in all four scenarios. Specifically, the Greedy batching and searching approach is utilised to search for available vehicles for batched travellers. We elaborate on the four strategies as follows:

- **S1**: Sectional + Individual. Zhang et al. (2020) proposed a sectional pricing method. The method divides the vehicle’s path into several legs based on the traveller’s pick-up and drop-off nodes. For each leg, travellers on board share the fare equally. The final fare is the sum of all legs plus the base fee. The implementation details are described in Appendix A. In terms of optimisation, HVO and SVO are solved separately to determine the optimal assignment.

- **S2**: mT-Share + Individual. A mobility-aware dynamic taxi ridesharing system proposed in Liu et al. (2021b). Only the payment model is adopted here for comparison. The fundamental principle is to distribute the ridesharing benefit between taxi drivers and travellers. As we do not consider the driver’s profit, all benefit is allocated to ridesharing travellers proportionally to their detour rate. We further elaborate on the details of implementation in Appendix B. Besides, we employ the HVO and SVO models to decide the optimal assignment.

- **S3**: Fairness + Individual. Fairness-aware pricing is based on each user’s detour rate and ridesharing benefit. Same as above, the HVO and SVO are solved simultaneously to determine the best ride-hailing and ridesharing vehicle assignment without considering the travellers’ mode choice decision.

- **S4**: Fairness + Integrated. Fairness-aware pricing. Instead of assigning the ride-hailing and ridesharing vehicles separately, the Integrated strategy unifies two types of service in a single ILP, complementing the prediction of the travellers’ mode choice.

To emulate the process that each on-board passenger must grant permission whenever a new traveller wishes to join, we let the platform check if the updated prices for those already on-board are lower than their previous price, as shown in Fig. 1. This incentivises on-board passengers to permit others to join. If this criterion is met, the ridesharing offer is considered valid. If not, the new traveller passively exits the system, as the shared trip is rejected by their potential co-riders. By doing so, travellers who receive unfair prices have the option to refuse before starting the trip. Consequently, we can quantify the loss caused by unfair pricing. Note that unfair pricing only happens in the S1 and S2, while our method can guarantee fairness by design.

### 4.2. Grid network

Since the NYC scenario requires considerable computational time and resources, we employ a smaller grid network to conduct a series of sensitivity analysis experiments. The goal of sensitivity analysis is to investigate how different choices of internal (i.e., discount parameter, profit coefficient, fleet size) and external (i.e., population characteristics) parameters affect the proposed strategies. By doing so, we can reveal the effectiveness and limitations of the proposed strategies.

As shown in Fig. 5, the network consists of 16 nodes and 48 uni-directional homogeneous links. On the supply side, in the baseline scenario, 50 ride-hailing and 50 ridesharing vehicles with four seats are randomly distributed across the network nodes in the baseline scenario. On the demand side, we randomly generate 1200 travellers whose departure time is distributed across an hour horizon. As shown in Table 2, all travellers’ related quantities, except for departure time and origin–destination information, are set the same as in the NYC scenario. Again, travellers are grouped every 30 s and the base discount rate is fixed as $b = 0.05$.

### 5. Experimental results

#### 5.1. Results for the proposed strategies

In Table 4, we present simulation results from implementing the two proposed optimisation strategies, namely Individual and Integrated, with the proposed fairness-aware pricing scheme incorporated. These strategies refer to strategies S3 and S4 introduced.
Fig. 5. Grid network for sensitivity analysis.

Table 4
Aggregated results for the NYC scenario.

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Profit ($)</th>
<th>Served travellers</th>
<th>Other modes</th>
<th>Discount</th>
<th>Occupancy</th>
<th>Service</th>
<th>Waiting time (s)</th>
<th>Detour (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ridehailing</td>
<td>(pax)</td>
<td></td>
<td>Ratio</td>
<td>Ride-hailing</td>
<td>Shared</td>
</tr>
<tr>
<td>Individual</td>
<td>175,922</td>
<td>16,510</td>
<td>22,592</td>
<td>10,354</td>
<td>0.27</td>
<td>1.97</td>
<td>0.76</td>
<td>202</td>
</tr>
<tr>
<td>Integrated</td>
<td>174,625</td>
<td>15,811</td>
<td>24,377</td>
<td>8,430</td>
<td>0.29</td>
<td>2.18</td>
<td>0.78</td>
<td>238</td>
</tr>
</tbody>
</table>

in Section 4. Two types of metrics are used for performance evaluation. The first refers to system-related metrics, such as profit, service ratio, and occupancy. The occupancy of a single vehicle is calculated as the sum of travellers on board across the entire simulation horizon divided by the number of periods in which vehicles are not empty. We use occupancy to measure the degree of ridesharing. Thus, this indicator does not consider empty mileage. The final occupancy is determined as the average occupancy of all ride-sharing vehicles. The second includes traveller-related metrics, such as average waiting and detour time.

The individual strategy generates a total of 175,922 $ in profit covering both ride-hailing and ridesharing services. In terms of order cancellation, which refers to the number of travellers who turn down the ride offer actively, 10,354 travellers choose to switch to other travel modes, which corresponds to nearly 20% of the total demand. Additionally, the average discount rate in ridesharing trips is 0.27, which implies that each traveller only needs to pay about 70% of the full fare. The occupancy (pax/veh) indicator shows how many travellers are on board on average throughout the simulation horizon. This is a critical metric to measure the degree of ridesharing. In terms of traveller-related metrics, the average waiting time is 3.4 and 3.2 min for ride-hailing and ridesharing vehicles, respectively. Besides, the detour time due to ridesharing is 3.6 min on average.

As the ride-hailing trips are more profitable than ridesharing trips, more ride-hailing users under the individual strategy generates a slightly higher profit than the integrated strategy. Nonetheless, the integrated strategy yields 18.6% fewer travellers who opt for other modes, serving 2% more travellers than the individual strategy. Compared to the individual strategy, the integrated strategy arranges more travellers using the ridesharing service. Consequently, the integrated strategy leads to a higher occupancy. In terms of traveller-related metrics, travellers experience similar levels of waiting and detour times in both settings, except for a slightly longer waiting time of half a minute for the ride-hailing service provided by the integrated strategy. We further discuss how different settings of parameters affect both individual and integrated strategies in Section 6. To summarise, the integrated strategy enhances system efficiency by attracting more travellers using the ridesharing service without sacrificing service quality.

5.2. Analysis and comparison with state-of-the-art strategies

System-related metrics We proceed by comparing the four strategies presented in Section 4 in terms of vehicle kilometres travelled (VKT) and profit. In Fig. 6(a), one can observe that the VKT of different methods is quite close. There is only a maximum 2.6% difference between the S3 and S2. Nonetheless, with this slight increase of VKT, the proposed fairness-aware method manages to serve 15–18% more travellers and yields 20% more profit, as shown in Fig. 6(b). Besides, the VKT of S3 exceeds S4 shortly after time 120 because there are more accumulated travellers in S3 than in S4. As shown in Fig. 7(a), more travellers are assigned in S3 from time 120 to 125. Consequently, the VKT of S3 experiences a brief increase towards the end. In contrast, the cumulative profit
Z. Zhou et al.

Fig. 6. Vehicle kilometres travelled (VKT) and cumulative profit of different strategies.

Fig. 7. Number of assigned, exiting and leaving travellers. Note that the x axis ends at time 125 minutes instead of 150. This is because there is no newly arrived travel request after time 120 except for the travellers who submitted requests before 120 and are still waiting in the queue. However, the vehicles are still performing tasks to deliver travellers after time 125. Therefore, time 125 denotes the end of matching while time 150 is the time when all vehicles stop moving.

of S3 initially takes the lead before time 100, but is soon approached by S4 from time 100 to 120. This is because S4 prioritises ride-sharing vehicles, which are less profitable compared to ride-hailing ones. Nonetheless, as ridesharing vehicles operate near capacity, more ride-hailing vehicles are involved in providing mobility service, as shown in Fig. 9(b). Overall, more travellers are served in S4. As the majority of travellers reach their destinations, the cumulative profit of S4 is quite close to S3 shortly before the end. After time 125, as more travellers are served in S3, the ultimate profit of S3 is slightly higher than the S4. Note that the profit is the difference between the total fare paid by the travellers and the total vehicle travelled distance weighted by times an energy consumption parameter. Besides, the single trip profit is added to the cumulative profit when the traveller arrives at her destination.

Moreover, Fig. 7(a) shows the number of travellers assigned; Fig. 7(b) exhibits the trend of travellers exiting the platform, and Fig. 7(c) displays the number of travellers leaving the system due to exceeding their maximum waiting time. In Fig. 7, the green dashed line shows the number of newly arrived requests across time. It is evident that strategy S4, which considers the user’s
mode choice, has the most people assigned and the fewest people exiting the system. Unfair pricing is the primary reason why fairness-aware methods, i.e., both strategies $S3$ and $S4$, achieve better performance in terms of profit and assigned customers when compared to strategies $S1$ and $S2$. However, the operator can further reduce the occurrence of travellers exiting due to unfair pricing in both $S1$ and $S2$ by not delivering travellers unfair matching offers and, e.g., adding a penalty for previous unfair matches in the next optimisation step(s). It is worth mentioning that fairness-aware pricing helps prevent travellers from leaving the system and attracts more travellers to use more sustainable mobility, i.e., ridesharing. Fairness pricing can implicitly improve the travellers’ satisfaction and thus benefit the MoD system in the long run.

As shown in Fig. 8, the average number of passengers for strategy $S4$ increases rapidly at the beginning of the simulation and then maintains a high level of ride-sharing. The proposed fairness-aware ridesharing pricing method enables strategy $S4$ to serve significantly more travellers, especially ride-sharing travellers, than strategy $S1$ and $S2$. Compared to strategy $S3$, strategy $S4$ benefits from the prediction of customers’ mode choices, hence improving ridesharing efficiency, reducing the number of travellers opting for other modes, and resulting in better utilisation of a mixed fleet. Note that the average passengers metric differs from the occupancy metric that we introduced above. The average passengers metric is the ratio between the total number of passengers on board at time step $t$ and the fleet size, which thus reflects the average number of passengers on board at a fleet level during each time interval.

Fig. 9 presents the number of moving vehicles. The total number of moving vehicles (Fig. 9(a)) across different methods shows a similar trend, i.e., it increases rapidly in the initial stage, maintains a high level of utilisation rate, and drops steeply as no new
travellers are entering the platform. However, strategy $S4$ outperforms the other strategies and achieves the highest vehicle utilisation rate, which is due to its capability to account for both types of vehicles and customers’ mode choices and favours ridesharing vehicles over ride-hailing vehicles since the former brings more benefits, resulting in a slow growth of the number of moving ride-hailing vehicles (Fig. 9(b)). It is worth noting that the coefficient $\delta$ utilised for scaling the ride-hailing trip profit plays a crucial role in determining which type of vehicle is preferred by the platform. We will discuss further on this aspect in Section 6.

**Traveller-related metrics** In addition to the system-related metrics, we show in Figs. 10 and 11 how individual travellers perceive the service quality. Strategy $S4$ has the longest waiting time on average, which is approximately 4 min for ride-hailing trips (Fig. 10(a)) and 3 min for ridesharing trips (Fig. 10(b)). Nonetheless, the gaps among different methods are negligible. As shown in Fig. 11(a), travellers in shared trips averagely experienced four minutes of detours using strategy $S4$, which is reasonable considering nearly 18% more customers are served by the same number of vehicles compared to strategies $S1$ and $S2$. Note that there are outliers in Fig. 11(a) that have detours above the acceptable slack time ($h$) defined in Table 1. This is because we set a drop-off deadline for each traveller, which is calculated via (22). For example, let us assume a traveller’s maximum waiting time is 5 min and acceptable slack time is 6 min. If a traveller is picked up right after submitting the request, we implicitly assume that the maximum detour time she can tolerate increases to 11 min, as long as the traveller can arrive at her destination before the deadline. This explains why there are outliers with detours above 540 s but still below the upper bound 1020 s, which is the sum of maximum waiting and acceptable slack time.

Finally, we found that on average a higher discount is entitled to travellers when using the fairness-aware method, shown in Fig. 11(b). Yet, compared to strategies $S1$ and $S2$, the discount generated by the fairness-aware pricing strategies is more concentrated and within a reasonable range. Theoretically, both $S1$ and $S2$ can provide any discount to the travellers. Nonetheless, due to the unfair pricing, the ridesharing trips in $S1$ and $S2$ are usually being rejected as potential travellers are not satisfied with the price. Consequently, few people chose the ridesharing service and the service ratio is low. This is why the detour time of $S1$ and $S2$ is
concentrated around zero and generally has smaller discounts. This clearly explains why a fairness-aware pricing method is critical for ridesharing services. Without fairness (S1 and S2), travellers have no incentive to use the pooling service, resulting in smaller detours and discounts. It is worth mentioning that the platform can adjust the discount parameter $\alpha$ to further control the discount rate.

There is a trade-off between profit and occupancy and, compared to other strategies, the strategy $S4$ achieves a good balance between profit and occupancy since it does not mandate all travellers to choose ridesharing trips indefinitely and sacrifice profit. Instead, it intelligently arranges travellers and optimises the utilisation of the mixed fleet to maximise system profit while maintaining other system indicators, such as service ratio and occupancy, at a high level without compromising service quality.

### 5.3. Computation time

This section analyses how computation time varies with the change in the number of incoming requests under different optimisation methods. It is worth mentioning that the time span between the request submission and request response time is critical for real-time ridesharing operations. Therefore a practice-ready ridesharing model is required to be solved within a few seconds in real-time implementations. As shown in Fig. 2, there are three computational steps: batching and searching; routing and pricing; and optimisation. Note that, differently from Alonso-Mora et al. (2017) and Simonetto et al. (2019), where the total computation time was presented, we focus here only on the time required to solve the optimisation problem, which is related to the main contribution of this paper, while we assume that the other steps can be solved efficiently, e.g., employing the methods presented in the aforementioned papers. In particular, we investigate the computation time required to solve problems HVO, SVO, part of the Individual strategy and the problem formulated in the Integrated strategy. The results are reported in Fig. 12, where one can notice that both HVO and SVO are solved in less than 1 s, while the Integrated problem is solved within a maximum time of 5 s for the period with the highest requirements. Obviously, it takes a longer time to solve the Integrated problem compared to the problems within the Individual strategy, in particular when the demand is at its peak. This is because the number of decision variables $x_{p,h,s}$ in (17) is higher than the number of decision variables $x_{p,v}$ in (6) and (15), due to the additional dimension, which grows significantly as the number of requests increases. Besides, differently from the Individual problem, the Integrated model is combinatorial and its continuous relaxation cannot be guaranteed exact. To further explore the computational efficiency of the Integrated problem, we conducted an in-depth analysis that is presented in Appendix C.

However, it is worth noting that, despite the first two steps can be computationally intensive, advanced algorithms have been proposed to efficiently tackle them, such as state-of-the-art vehicle searching (Ma et al., 2013), insertion (Tong et al., 2018), and distributed implementation methods (Al-Abbasi et al., 2019).

Still, since the proposed methods do not add a heavy computational burden to state-of-the-art ridesharing frameworks, where modules for searching available vehicles and ridesharing routing have been proven viable for real-time applications (Simonetto et al., 2019; Liu et al., 2021b), we claim that the proposed pricing and matching method is efficiently applicable to large-scale ridesharing problems.

### 6. Sensitivity analysis

#### 6.1. Internal factors

**Impact of discount parameter $\alpha$.** As the parameter $\alpha$ in (13) determines the level of discount, we investigate the impact of $\alpha$ on the performance of the individual and integrated strategies. Table 5 presents simulation results for various values of $\alpha$, ranging...
As a result, ridesharing is implicitly discouraged. Both the individual and integrated strategies produce similar outcomes, except that the integrated strategy attracts more users to ridesharing services. Moreover, the total profit initially increases with rising values of \( \alpha \), but then reaches a peak before declining thereafter. Although fewer people choose other travel modes as \( \alpha \) grows, the increasing discount rate lowers profits overall. Compared to the individual, the integrated strategy attracts more users to ridesharing services. Moreover, the total profit initially increases with rising values of \( \alpha \), but then reaches a peak before declining thereafter. Although fewer people choose other travel modes as \( \alpha \) grows, the increasing discount rate lowers profits overall. Compared to the individual, the integrated strategy is more sensitive to changes in \( \alpha \). Specifically, the discount rate increases at a faster rate and the number of people exiting drops quicker in the integrated strategy than in the individual strategy. This is explained by the decision-optimization mechanism in the integrated strategy, which enables the service to be more competitive in reducing the number of people leaving the system. As \( \alpha \) increases, the integrated strategy would lower the price to keep travellers from leaving, while the individual strategy would consistently prioritise trips with higher profits due to its lack of traveller decision prediction. In brief, we observe that the integrated strategy consistently outperforms the individual strategy under all circumstances.

**Impact of the coefficient \( \delta \) on ride-hailing trip profit.** In (6) and (16), the coefficient \( \delta \) is designed to scale the ride-hailing trip profit. As the objective functions (6) and (17) are profit-oriented, the smaller \( \delta \), the less profit obtained from ride-hailing trips. Hence, the ridesharing service, which is usually less profitable yet more sustainable, is preferred by the platform. We investigate the performance of the individual and integrated strategies and show the results in Table 6. Specifically, the coefficient \( \delta \) varies from 1 to 0.5. Our findings indicate that the individual strategy is less sensitive to changes in \( \delta \), since \( \delta \) is only adopted in the objective function and does not affect the actual trip profit. Besides, since the ride-hailing and ridesharing service plans are determined separately, the platform cannot respond to the changes in profit obtained from different service types. Therefore the results across different settings are identical. In contrast, the integrated strategy is more adaptive and responsive to different values of \( \delta \). When \( \delta = 1 \), ride-hailing trips are more profitable, leading the platform to prioritise such trips over ridesharing. However, as \( \delta \) decreases, the platform shifts towards ridesharing trips. Notably, we find that the integrated strategy outperforms the individual strategy across all settings.

**Impact of fleet size.** We investigate the performance of the individual and integrated strategies across different fleet sizes, ranging from 50 to 150 vehicles. Since we have two types of vehicles, we denote each scenario by a code defined as ‘xx–yy’, where xx represents the number of ride-hailing vehicles and yy the number of ridesharing ones. For example, ‘25–75’ corresponds to a scenario with 25 ride-hailing vehicles and 75 ridesharing vehicles. Note that, although the fleet size may vary, we do not consider the vehicle fixed costs, such as the vehicle purchase fee. The profit is calculated as the difference between the total trip fare and the distance-based cost. The results are presented in Table 7. When the fleet size is small, all performance metrics of the two strategies are quite close except for the number of people exiting the system, which demonstrates that the integrated strategy could effectively prevent potential users from leaving the system. Furthermore, increasing the fleet size to 100 dramatically improves the service ratio and profit. Compared to the unbalanced scenarios where the number of ride-hailing and ridesharing vehicles is not equal,
more ridesharing vehicles (‘25–75’) can further boost the system’s profit. Moreover, if we increase the discount rate \( \alpha \), the service ratio as well as other performance indicators would be further improved. On the contrary, more ride-hailing vehicles (‘75–25’) lead to less desirable results, where not only system profit suffers from a decline, but also travellers have to experience longer waiting and detour time. Therefore, we conclude that increasing the number of ridesharing vehicles enhances the system service capacity while maintaining a relatively smaller fleet of ride-hailing vehicles for special needs. Lastly, we find that keeping increasing the fleet size (‘75–75’) has a limited effect on increasing the number of served customers. Because the discount rate \( \alpha \) is fixed in this case. A higher discount rate \( \alpha \) would incentive travellers and achieve better performance for larger fleet sizes. Again, the \textit{integrated} strategy consistently outperforms the \textit{individual} regardless of fleet size.

### 6.2. External factors

Except for these internal factors which the platform can control, there are two critical external parameters affecting the effectiveness of the proposed method, especially for the \textit{Integrated} strategy, which is based on the prediction of the traveller’s mode choice. Note the parameter \( \beta \) implicitly indicates which modes are considered for competing. For instance, if the taxi is the only option left for travelling excluding the on-demand service, then the mean \( \mu \) of \( \beta \) should be close to the average utility of ride-hailing trips \( u_h \) and the standard deviation \( \sigma \) of \( \beta \) is small. In contrast, if there is more than one type of competing mode, for example, public transit, taxi, and private vehicles are all involved, then the standard deviation \( \sigma \) of \( \beta \) is large. We conduct sensitivity analysis on the mean \( \mu \) and standard deviations \( \sigma \) of \( \beta \) to evaluate how the accuracy of mode choice prediction affects the proposed method.

#### Impact of Standard deviation \( \sigma \) of \( \beta \) on \textit{integrated} strategy

From (3), we can infer that the utility constant \( \beta^\sigma \) for other travel modes determines the degree of traveller satisfaction for these modes. Given that the \textit{integrated} strategy accounts for travellers’ mode choice, we evaluate its performance under different settings of \( \sigma \) and compare it with the \textit{individual} strategy for consistency. In Table 8, we show how the \textit{individual} and \textit{integrated} strategies react when \( \sigma \) varies from 1 to 10. When \( \sigma = 1 \), i.e., only one type of competing travel mode (e.g., private vehicles or taxis) is considered, the \textit{integrated} strategy is significantly superior to the \textit{individual} strategy. Specifically, we observe a 10.6% increase in total profit and a 21.0% decrease in the number of people exiting the system. Besides, the \textit{integrated} strategy attracts more ridesharing users and thus has a higher average number of travellers on board. However, as \( \sigma \) increases, the benefits of predicting travellers’ travel modes are less prominent. Thus the gap between the \textit{individual} and \textit{integrated} strategies is reduced. This decline in performance is expected since larger values of \( \sigma \) imply more significant fluctuations in \( \beta^\sigma \), which makes predicting a traveller’s mode choice based solely on the population mean \( \mu \) less accurate. Since the platform lacks access to individual preferences for other modes, the accuracy of the prediction decreases, particularly when \( \sigma \) is large.

#### Impact of the mean \( \mu \) of \( \beta^\sigma \) on \textit{integrated} strategy

The population mean \( \mu \) of \( \beta^\sigma \) (Table 2) shows a general degree of people’s satisfaction with the competing modes. If \( \mu \) is much higher than the on-demand service, then no one would choose the on-demand service and vice versa. Therefore, it is interesting to investigate how different optimisation method performs with different \( \mu \). In Fig. 13, we compare the two strategies under a different choice of \( \mu \). When \( \mu = -25 \), around half of the customers leave the
on-demand system. Compared to the individual, the integrated yields 11.4% more profits and increases the service ratio as well as the average traveller rate. When $\mu = -30$, both individual and integrated strategies manage to serve more travellers. However, the gap between the two methods reaches a peak. Compared to the individual strategy, the service ratio of the integrated strategy increases by approximately 7% due to the customer mode choice prediction. By attracting more travellers to use ridesharing services, the average traveller rate grows significantly. As a result, the detour time and discount rate increase inevitably. Nonetheless, further lowering $\mu$ narrows the gap between the individual and integrated strategies since a smaller $\mu$ indicates that other travel modes are less attractive than the on-demand service. Hence, people are most likely choosing the MoD when making the mode choice. In that case, the platform can emphasise profit without worrying about people leaving the system; consequently, the benefit of adopting the integrated strategy is less prominent. Although the integrated method experiences fewer traveller exits when $\mu = -40$, the service ratio for both methods remains almost the same, as illustrated in Fig. 13(b) and (d). This is because the integrated method attracts more ridesharing users. Consequently, it takes longer for the integrated method to complete a single trip. As a result, more travellers in the integrated method choose to leave the system when the waiting time exceeds their limits. Nonetheless, this rarely happens in practice, since in urban traveller transport there are always similar and competitive travel modes, such as private vehicles, taxis, and vehicles operated by other MoD platforms. Hence the utilities of these competing modes are quite close. It suggests that the second ($\mu = -25$) and third ($\mu = -30$) scenarios are most likely prevalent in practice.

7. Conclusion

This study presents a fair-price based optimisation for vehicle–traveller matching in on-demand mobility systems, where both ride-hailing and ridesharing services are provided simultaneously. Particularly, the travellers are assumed to be cost-sensitive and allowed to choose travel modes from ride-hailing and ridesharing, or switch to external alternatives. Besides, we defined six fairness principles and a discount function for dynamic ridesharing so that the travellers’ trip fares can be updated in real-time. Two ILPs are designed to determine the ride-hailing and ridesharing vehicle–traveller schedules separately. Furthermore, an advanced optimisation model is developed to integrate the ride-hailing and ridesharing assignment considering the travellers’ mode choices. It is worth mentioning that our method is suitable also in the case of shared automated vehicles. Although we did not consider drivers’ commissions, the method can be extended to include them, e.g., by assigning a fixed percentage of the profit to drivers, regardless of whether they are company-employed or freelance drivers.
A series of numerical experiments are performed to investigate the proposed method’s effectiveness, robustness, and limitations. To begin with, the proposed fairness-aware pricing scheme considering both strategies of individual and integrated optimisation is compared with state-of-the-art pricing methods using the NYC taxi dataset. We find our fairness-aware pricing generally has a better performance. Compared to state-of-the-art pricing strategies, the integrated optimisation strategy can achieve 20% more profit by serving 18% more customers using the same fleet, while only increasing 2% VKT. This is because our approach can significantly reduce the occurrence of unfair pricing and thus reduce the number of exiting travellers. Meanwhile, the average number of travellers in ridesharing trips increased by 50% with a slight increase in travellers’ waiting and detour time. Moreover, a series of simulations are implemented in a small grid network for sensitivity analysis. Regardless of different choices of parameters, the integrated optimisation strategy consistently outperforms the individual optimisation strategy. However, the magnitude of differences between the two methods depends on the knowledge of travellers’ preferences. Experimental results indicate that a moderate discount and profit coefficient, more ridesharing vehicles, and a concentrated distribution of population utility of other travel modes can boost the performance of the integrated optimisation strategy.

There are lots of interesting directions for extensions. Since the effectiveness of the proposed integrated optimisation strategy relies on the accuracy of predicting travellers’ mode choice decisions, improving the prediction accuracy would enhance its performance. One potential method is to apply adaptation and prediction methods, e.g., based on AI-based techniques, for travellers’ choices based on their feedback or order cancellation rates. Additionally, we assume that the link travel time is constant, and it would be interesting to consider dynamic travel time and integrate matching and pricing with real-time time-dependent routing (Zhou and Roncoli, 2022). Another direction is combining pricing with the repositioning of idle vehicles (Filipovska et al., 2022), as pricing is a powerful tool to tackle the supply-and-demand imbalance problem. Finally, involving drivers in the decision-making process and pricing considering future supply-and-demand imbalance (Syed et al., 2021) are also promising research avenues.

CRediT authorship contribution statement

Ze Zhou: Conceptualisation, Methodology, Software, Validation, Formal analysis, Writing – review & Editing. Claudio Roncoli: Conceptualisation, Methodology, Writing – review & editing, Supervision, Funding acquisition. Charalampos Sipetas: Conceptualisation, Methodology, Writing – review & Editing.

Data availability

Data will be made available on request.

Acknowledgements

This research is funded by the FINEST Twins Center of Excellence (H2020) under Grant 856602 and the Academy of Finland projects ALCOSTO (no. 349327).

We also acknowledge the computational resources provided by the Aalto Science-IT.

Appendix A. Implementation details of the sectional pricing method

Our implementation of the method by Zhang et al. (2020) is presented in Algorithm 1, which shows how we adapted it to our fare notions and settings. Instead of utilising payment budgets and pure distance-based pricing, we employ, for consistency, the time fare rates defined in (5). The functions \( \text{Dis}(\cdot) \) and \( \text{Time}(\cdot) \) are defined to calculate the travel distance and time between two arbitrary points. For each trip leg, we rewrite the leg cost as \( f_d \cdot \text{distance} + f_t \cdot \text{time} \). Then all the active travellers who travel through the leg \( (x_{i-1}, x_i) \) share the cost of this leg equally. Finally, each traveller pays the sum of all sectional fares.

Algorithm 1: The sectional pricing method implemented in this paper

<table>
<thead>
<tr>
<th>Data: Schedule ((x_0, x_1, ..., x_{2l})), Request profiles of a collaborator tuple ((r_1, r_2, ..., r_i)), time ((f_t)) and distance ((f_d)) fare rate</th>
<th>Result: Fare paid by each traveller ((c_1, c_2, ..., c_l) = (b, b, ..., b))</th>
</tr>
</thead>
<tbody>
<tr>
<td>for (i=1) to (2l) do</td>
<td>(c_i = c_{i-1} + \frac{\text{price}}{\sum \text{time}})</td>
</tr>
<tr>
<td>(\text{distance} = \text{Dis}(x_{i-1}, x_i)); (\text{time} = \text{Time}(x_{i-1}, x_i)); (\text{price} = f_d \cdot \text{distance} + f_t \cdot \text{time}); for each active (r_j) do</td>
<td>end</td>
</tr>
<tr>
<td>end</td>
<td></td>
</tr>
<tr>
<td>return (c_1, c_2, ..., c_l)</td>
<td></td>
</tr>
</tbody>
</table>
Appendix B. Implementation details of the mT-Share pricing method

We adopt a Liu et al. (2021b) payment model to fairly split the ridesharing benefit based on the detour, according to

\[ B = \sum_{i=1}^{n} f_{ri} - F \]  \hspace{1cm} (B.1)
\[ \rho_i = \nu + \frac{\text{cost}(R_{ri}) - \text{cost}(R_{ri}^s)}{\text{cost}(R_{ri}^s)} \]  \hspace{1cm} (B.2)
\[ f_{ri} = f_{ri}^s - \beta \cdot B \cdot \frac{\rho_i}{\sum_{z=1}^{n} \rho_z} \]  \hspace{1cm} (B.3)

In (B.1), we calculate the ridesharing benefit \( B \) by subtracting a shared taxi fare \( F \) from the sum of \( n \) travellers’ fares if they travel alone. For consistency, (5) is utilised for the fare calculation. We derive traveller’s detour rate \( \rho \) from (B.2), where \( \nu \) denotes the base rate (set as \( \nu = 0.01 \) as in Liu et al., 2021b), \( R_{ri} \) is the shared route that traveller \( r_i \) has travelled, while \( R_{ri}^s \) is the shortest path. In our implementation, the function \( \text{cost}(\cdot) \) is assumed equivalent to the trip length. In (B.3), \( f_{ri} \) is the final price, which is the difference between the original taxi fare \( f_{ri}^s \) and a portion of the ridesharing benefit \( B \). Finally, \( \beta \) is the split rate to partition the benefit \( B \) between the taxi driver and travellers and, since in this work we ignore the driver cost, we set \( \beta = 1 \).

Appendix C. Computation time analysis for the integrated optimisation model

As the optimisation problem (17)–(21) is combinatorial, we are interested in testing the proposed method with different demand densities (req/s) to identify the computational limitations of the proposed approach. We employ the same dataset and set all the parameters as we did in the NYC scenarios. Instead of running the whole two-hour simulation, we only implement the integrated method using one batch of requests. We manually increase the number of requests in one batch from 254 to 3220. The code is running on the same shared server with 40 cores and 768 GB of memory, and we utilise Gurobi to solve the integrated optimisation model. Note that the server we used is a standard server that we share with other concurrent applications, and, actually, we found out that the computation time is even shorter while solving the same optimisation problem on a dedicated laptop.

As shown in Fig. C.14, the integrated model can be solved in less than 30 s even with up to 1500 requests within a batch. On the other hand, we argue that the dataset we use is considerably large enough. In the NYC scenario, we group requests every 30 s and a maximum of 700 requests are accumulated in one batch. Thus, the integrated model can be solved within a few seconds. Referring to Alonso-Mora et al. (2017) and Simonetto et al. (2019), both of them conducted real-time ridesharing simulations using the NYC taxi dataset, which contains around 400,000 requests per day. Thus, the average demand density is 4.63 req/s. If we consider a batch of 30 s, then it will include \( \sim 139 \) requests. By checking Fig. C.14, we can see that this number of requests could be solved within less than 1 s with our model. Of course, the demand is much higher during peak hours. Even assuming a much higher demand during the peak hours, e.g., a request density of 20 req/s, this would result in 600 requests in a 30-second batch, which, according to Fig. C.14, would imply a computation time of less than 2 s. Furthermore, choosing a smaller sampling (batching) period would allow us to control the number of requests in a batch, further reducing the computational burden, if necessary. Although the scalability of the integrated model is not guaranteed due to its combinatorial formulation, we demonstrate that the proposed problem can be solved efficiently enough in real implementations and thus virtually applicable in a real-time large-scale ridesharing system.
References


Cao, Y., Wang, S., Li, J., 2021. The optimization model of ride-sharing route for ride hailing considering both system optimization and user fairness. Sustainability 13 (2), 902.


