Transportation Research Part C 155 (2023) 104284

Contents lists available at ScienceDirect

Transportation Research Part C

journal homepage: www.elsevier.com/locate/trc

Leveraging ride-hailing services for social good: Fleet optimal routing and system optimal pricing

Zemian Ke [a], Sean Qian [a,b,*]

[a] Department of Civil and Environmental Engineering, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, 15213, PA, USA
[b] Heinz College, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, 15213, PA, USA

ARTICLE INFO

Dataset link: https://github.com/Jimmy-Ke404/ORHP

Keywords:
Ride-hailing
Optimal pricing
Mixed equilibrium

ABSTRACT

With the penetration of ride-hailing services, the impacts of transportation network companies (TNC) on network performance grow. TNC comes with a ubiquitous sensing and pricing system that may be leveraged by public agencies to improve transportation system performance. This study first formulates and solves a mixed equilibrium (ME) of personal driving vehicles and ride-hailing vehicles, where TNCs centrally assign routes to ride-hailing vehicles to achieve fleet-wide optimum. We propose a novel fleet behavior desired by TNCs, named fleet-optimal behavior with service constraint (FOSC), which provides a good compromise between total fleet cost minimization and fairness among riders. However, we show that the system state of FOSC can be far from the system optimum state. To this end, we propose a novel Optimal Ride-hailing Pricing (ORHP) scheme for public agencies as an efficient manner to intervene ride-hailing online platforms. The essential idea of ORHP is to regulate and subsidize TNCs, in exchange for guaranteed network performance improvement. Under ORHP, public agencies set the value of a subsidy for each link for any TNC rider using this link. The subsidies are provided to TNCs, not directly to riders. TNC receives subsidies, and determines the best way to provide compensations for each rider who deviates from his/her "shortest" route, determined by FOSC. TNC's ultimate goal is to reduce their total cost, including total fleet vehicle travel time and total compensations, subtracted by the total subsidy received from public agencies. The ORHP is likely less controversial than other pricing schemes, since it calls for voluntary participation of travelers who are provided with multiple route options. Because it is built into the TNCs' fare system, it is cost effective to implement and hard to game. This would be a win-win: a win for public agencies to cost-effectively improve system performance leveraging TNC's platform without building physical infrastructure or services; and a win for TNC to profit from subsidies and improve service quality. ORHP is formulated as a bi-level optimization problem, solved with a sensitivity analysis based algorithm and tested on two networks implying the ORHP scheme can be effective: a small total subsidy provided to TNC can lead to significant improvement in system performance.

1. Introduction

Transportation systems are designed for all: to meet the travel needs of individuals, as well as to connect and support regional economies. Transportation managers, however, face unprecedented challenges due to increasing congestion, emissions, energy use and infrastructure deterioration. Many solutions have been proposed and deployed to address those challenges, which however, can

* Corresponding author at: Department of Civil and Environmental Engineering, Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh, 15213, PA, USA.
E-mail address: seanqian@cmu.edu (S. Qian).

https://doi.org/10.1016/j.trc.2023.104284
Received 1 November 2022; Received in revised form 30 July 2023; Accepted 30 July 2023
Available online 12 August 2023

hardly please all. Transportation systems are largely driven by uncoordinated (and selfish) travelers' decisions that result in traffic states at the busiest times and locations that can be far away from a social system optimum. This calls for controlling demand, oftentimes in the form of incentivizing travelers to align their behaviors, with broad objectives such as minimizing system-wide travel delays, mitigating environmental impacts, and minimizing social costs. Unfortunately, existing incentives, such as congestion pricing (Newell, 1987), tradable credits (Xiao et al., 2013) and parking pricing (Qian et al., 2012), unavoidably bring in technical, financial and social equity concerns. For example, it is often questioned that optimal incentives are difficult to derive, and existing ones lead to inefficient outcomes and disproportionate impacts on various users. Furthermore, implementing incentives in the real world could be very costly. What is missing is an easily deployable and ubiquitous system that can control demand (e.g.,via incentives) in an inexpensive, effective, and fair manner, with voluntary participation from travelers.

Transportation network companies (TNC) offering ride-hailing services, e.g., Uber, Lyft, Didi, Grab, hold great potential to become elements of an ubiquitous demand control system. The premises are based on voluntary participation from travelers who have money exchange through a platform. If any incentives or disincentives are added to those platforms by either the private platform owners or public agencies' regulations, potential controversy that mandating all populations to participate, such as in congestion pricing or tradable credits, would have been mitigated. It is worth noting that such efforts via the platform would be cost effective to public agencies. Instead of procuring from private contractors for a ubiquitous charging and sensing system (e.g. tolling systems, or online trading systems), all incentives would be provided and regulated in an existing platform. Those service platforms are led and maintained by the private sector in the first place, and potentially profitable to the service providers under many business models.

Next we argue that TNCs could willingly work with public agencies to provide incentives, because this can be profitable to TNCs and benefit all riders in respective platforms with a higher service quality. Prior research implies that, with even a small fraction of market penetration, TNCs have motives to coordinate all vehicles/riders within their respective platform, routing in different ways from individual selfish routing (i.e. User Equilibrium) (Harker and Pang, 1988; Yang and Zhang, 2008; Battifarano and Qian, 2023). This is to TNCs' benefit by reducing their own travel time from individual selfish routing, so to offer highest service quality and enable more orders per time unit. This, however, comes at the price of dis-proportionally benefiting individual riders. In a low TNC's market penetration, this may even harm some riders as they can be directed to a route that benefits TNCs, but not necessarily those riders. This paper precisely addresses this concern. We show that in theory, a TNC platform would have motives to offer a set of incentives that would benefit all riders equally, and the incentives can be regulated and passed along to TNCs from public agencies that come more cost-effectively reducing total system congestion (not only TNCs' total travel time, but population total travel time) than other pricing schemes. The same theory and design could be applied to any platform that has significant market share of vehicles, such as delivery fleets (e.g., UPS) or information dissemination platforms (e.g., Google Maps). Ride-hailing services are one of those good examples.

The potential of regulating those platforms (TNCs as an example) is clear: These TNC platforms offer unprecedented opportunities to align travel behaviors with TNC-wide goals that, interestingly, would turn out benefiting all stakeholders: users, ride-sharing providers, and society. This promise is contingent on that, in many urban areas, ride-sharing systems have grown to represent a considerable fraction of traffic (e.g., over 10% in some cases) and are still growing quickly. Ride-hailing (shared rides or non-shared rides) creates two major paradigm shifts: (i) unlike atomistic travelers, ride-sharing operations can have substantial impacts on traffic flows and resulting system performance, and (ii) they are by nature an ideal platform for ubiquitous sensing, information provision and real-time control, e.g., Uber's "surge price" or ride-pooling option ("uberPOOL") through built-in technology-enabled price incentives. With those great features and advantages, TNCs can thus be leveraged by the communities to design and implement novel control schemes that ultimately promote the profitability of the service providers and reduce the social costs of transportation. In a simple example, a service provider, with subsidies from the city, incentivizes users (via surcharge or credit to their respective fares) for intended choices of departure time, route and pooling that aim for improving both its profits and system-wide performance. Meanwhile, the city can effectively improve system-wide performance through a subsidy to the service provider in exchanged for guaranteed performance improvement. This paper shows this could be possible in theory.

The literature acknowledges that the popularization of mobile internet has prompted the rapid expansion of such ride-hailing services, but their impact to traffic systems is unclear. For example, Uber offers ride-hailing services to more than 91 million users with 15 million daily trips as of mid-2019 (DMR, 2021). Impacts of ride-hailing systems on network are not trivial but increasingly significant (Wang and Yang, 2019). Some literature study impacts of ride-hailing systems on other transportation services, such as public transit (Zhang and Zhang, 2018; Babar and Burtch, 2017; Pham et al., 2020), and conventional taxis (Wallsten, 2015; Harding et al., 2016; Nie, 2017), while others analyze social and environmental impacts of ride-hailing systems, such as private car ownership (Anderson, 2014), energy consumption and emission (Yu et al., 2017). In general, as Jin et al. (2018) argue, the impacts of ride-hailing systems on traffic congestion is still unclear. This paper focuses on direct service impacts on network traffic flow through routing, provided with exogenous vehicle trips without modal choices.

Ride-hailing vehicles compete with personal driving vehicles for limited road network resources, which was mathematically formulated as mixed equilibrium (Harker and Pang, 1988; Yang and Zhang, 2008; Battifarano and Qian, 2023). The mixed equilibrium may lead to a different network congestion pattern from UE, depending on the travel behavior of ride-hailing vehicles. Since ride-hailing vehicles from one TNC can be centralized routed, the ride-hailing vehicles from one TNC are regarded as one vehicle fleet, and the behavior of vehicles within the fleet is denoted as the fleet behavior. Present studies indicate the ride-hailing vehicle fleet might follow UE (Geisberger et al., 2009) or fleet optimum (FO) fleet behavior (Harker, 1988; Yang and Zhang, 2008). UE behavior means ride-hailing vehicles would choose the best path to minimize individual travel time/cost, which benefits passengers the most but not necessarily for TNC as a whole fleet. In this case, the mixed equilibrium is exactly same as the system

UE. However, if the ride-hailing fleet has influences on network congestion patterns, then a TNC would be motivated to route its vehicles differently from UE to benefit itself the most. Ride-hailing vehicles following FO fleet behavior would cooperate to minimize total travel time of all vehicles within the fleet, which implies overall better service quality to riders and more opportunities for taking additional orders. In this case, ride-hailing vehicles are assigned to paths with minimum marginal fleet cost. With the growth of ride-hailing services, such effects of ride-hailing vehicles would continue to grow.

FO is optimal for TNC since it minimizes TNC's total fleet travel time and opens more vehicle times to take additional ride-hailing orders. FO may benefit the entire system including all drivers, because FO may drive the system from selfish UE towards System Optimum (SO), as indicated by a number of studies (Harker and Pang, 1988; Yang and Zhang, 2008; Battifarano and Qian, 2023). In an extreme case, if a TNC owns all vehicles on the roadway, it would operate all its fleet vehicles to achieve SO. However, FO is not necessarily optimal for passengers because paths with minimum marginal fleet cost are not necessarily the minimum cost paths to an individual.

In this study, we propose a new fleet route choice behavior under (dis) incentives, called fleet optimum with service constraints (FOSC). With FOSC, TNC would still assign ride-hailing vehicles to minimize its generalized total fleet cost with one additional consideration: to ensure its mobility service remains attractive to passengers, namely, passengers not assigned to their respective minimum cost paths would be compensated for their 'sacrifice'. Note those compensations are also included in the generalized total fleet cost, counted towards the operation cost for TNCs. We show that FOSC with compensations are feasible, and thus would benefit both TNC and ride-hailing passengers. By modeling a mixed equilibrium with personal driving vehicles and ride-hailing vehicles, we find ride-hailing vehicles, even with a small penetration rate, are able to affect network congestion substantially if following goals towards a fleet optimum. Hence, another question is, can public agencies leverage such impacts of ride-hailing vehicles to mitigate network congestion in a more cost effective way than creating its own (dis) incentivization system?

Existing online TNC platforms along with a build-in fare system provide an inexpensive and efficient way to implement travel demand management measures. For example, one can utilize existing platforms to implement tolls/subsidies to affect ride-hailing vehicle behavior to improve network efficiency. Toll/subsidy information and payment can be easily integrated into the online platforms, and there is no need to install any additional infrastructure such as toll gantry or toll booths. More importantly, tolls/subsidies can be set for any road segments on such platforms. Recently some airports have already implemented a surcharge for ride-hailing vehicles, which is one example of this ubiquitous traffic management system. One key benefit on top of all pricing flexibility is that traveling with TNC is voluntary rather than mandatory. This is likely to be less controversial than other incentivization designs. In a nutshell, this study proposes an optimal ride-hailing pricing scheme (ORHP) with subsidy from public agencies to intervene with TNC aiming at improving network performance, while considering ride-hailing service's quality and profitability.

Zhang and Nie (2018) shows when a fraction of vehicles (e.g., a fleet of autonomous vehicles (AVs)) follow SO behavior, the system total travel time can be effectively reduced. If the fleet size is sufficiently large (but not necessarily 100%), the SO system state can be realized (Chen et al., 2020). This study is fundamentally different. In practice, ride-hailing vehicles would not follow SO behavior without incentives, since the SO behavior is not necessarily optimal for TNC as a whole fleet nor individual ride-hailing users. We argue public agencies can work collaboratively with TNCs, and reduce generalized system cost by leveraging existing online ride-hailing platforms and implementing link-based subsidies/surcharge exclusively for ride-hailing vehicles. The generalized system cost to public agencies includes the system total travel time and the subsidy cost. With the parameter weighing the system total travel time over the subsidy cost, the public agency is able to balance among the savings of system total travel time and the cost of subsidies paid to ride-hailing vehicles. The ORHP will then lead to a win–win: a win for public agencies to effectively reduce system cost leveraging TNC's platform; and a win for TNC to profit from collected fares with subsidies from public agencies and meanwhile reducing TNC's own total travel time.

Our intention is not to scrutinize ride-hailing fleet operation. Rather, we choose to use OD demand of ride-hailing vehicles exogenously that would have built in searching, idling and service timing behaviors. The novelty of our paper is that, by working with an arbitrary set of ride-hailing demand (exogenous demand), even at a small penetration across all O–D pairs, public agencies can leverage that to achieve an improved services rather than investing its own sensing/pricing infrastructure. We fully acknowledge this is not a perfect model for ride-hailing operations, and any intervention to the network would in turn change the searching, idling and service timing behaviors and demand of ride-hailing vehicles, so to impact the system performance endogenously. However, this would rely on sophisticated network-level simulation of behaviors of drivers, riders and system managers, which is not necessary at the stage of proof of concept for our innovative traffic management strategy leveraging public–private partnership. It would also lose the mathematical tractability of system-level optimizations, thus not helping us identify analytical insights for the newly proposed strategies. Our study, as its first step, intends to propose a new concept, shows its promises under arbitrary network flow demand of both private and ride-hailing vehicles, and demonstrates how it can be solved analytically in real-world networks.

The contributions of this study are four-fold.

- We analyze the impacts of ride-hailing vehicles on network performance with different penetration rates. Cases where TNC choose different fleet behaviors for ride-hailing vehicles are studied and compared.
- A novel fleet behavior desired by TNCs, fleet optimum with service constraints (FOSC), is proposed to consider compensations provided to riders to ensure service quality, and a heuristic algorithm is developed to solve the respective system state under mixed equilibrium. UE fleet behavior benefits riders but may not be optimal for TNC, while FO behavior would trade longer travel times of some riders for fleet improvement, which leads to fairness issues among riders. In contrast, the proposed FOSC reduces fleet travel time while ensuring fairness among riders, which would be desired by TNCs.

- Most importantly, we design a novel Optimal Ride-hailing Pricing (ORHP) scheme for public agencies as an efficient manner to manage network traffic and improve system performance. It intervenes and leverages ride-hailing online platforms without requiring any additional capital investments on infrastructure, services or systems. The essential idea of ORHP is for public agencies to regulate and subsidize TNCs, in exchange for guaranteed network performance improvement. The ORHP is likely less controversial than other pricing schemes, since it calls for voluntary participation of travelers, provided with multiple route options. Because it is built into the TNCs' fare system, it is cost effective to implement and may eliminate scamming behavior that may occur in other credit-based schemes. The ORHP benefits both the public agencies for effectively reducing system cost and TNC for being more profitable.
- To solve ORHP, we extend the sensitivity analysis method in Yang and Huang (2005) to the multi-class traffic assignment problems. The sensitivity analysis-based solution algorithm converges fast even though ORHP problem is highly non-convex. Using the mathematical formulations and numerical examples in real-world networks, we show that this is practically plausible to benefit the entire system, TNC riders, non-TNC private drivers, and the TNC as well: everyone wins.

The remainder of this paper is organized as follows. In Section 3, we first introduce FO fleet behavior, and then the proposed FOSC fleet behavior is illustrated and the solution algorithms for its mixed equilibrium are developed. Section 4 formulates the proposed ORHP as a bi-level optimization problem, and then a sensitivity analysis-based algorithm is developed to solve for optimal pricing schemes. Section 5 gives two numerical examples to illustrate the impacts of ride-hailing vehicles on network and the effects of ORHP and solution algorithms. The paper is concluded with a summary of main findings and future research in Section 6.

2. Related work

For realism of network modeling with different types of vehicles, Dafermos (1972) proposed a multi-class traffic assignment model by adopting different cost functions for vehicle class, and the corresponding equilibrium state is also known as the mixed equilibrium. Since then, multi-class traffic assignment has been utilized to model heterogeneity among vehicles, such as travel modes (Florian, 1977), travel behaviors (Dafermos, 1972; Harker, 1988; Van Vuren et al., 1990; Yang et al., 2007; Yang and Zhang, 2008; He et al., 2013; Zhang and Nie, 2018; Chen et al., 2020; Delle Site, 2021), value of time (Yang and Huang, 2004; Nagurney and Dong, 2002), perceptions of travel time (Van Vuren and Watling, 1991; Yang, 1998; Wang et al., 2019), and desired arrival times (Lo et al., 2006; Shao et al., 2006). Though personal driving and ride-hailing trips are two different travel modes, driving vehicles and ride-hailing vehicles have the same impacts on traffic flow and consequently link travel time functions. The key impact to network performance varied by driving vehicles and ride-hailing vehicles is that they may have different route choice behavior, which is the focus of this paper.

In the literature, routing behaviors considered in the mixed equilibrium mainly include UE behavior, SO behavior and FO behavior. UE behavior means vehicles choose paths with minimum travel time/cost of all possible paths, while SO (FO) behavior means vehicles choose path with minimum marginal system cost (marginal fleet cost). Dafermos (1972) and Van Vuren et al. (1990) studied the mixed UE and SO behaviors, and Dafermos (1972) showed the multi-class traffic assignment problem can be reduced to a single class traffic assignment problem. Harker (1988) argued vehicles (e.g., motor carriers and mass transit) from the same mobility company would cooperate to minimize the total cost of the fleet vehicles from the same company, namely FO behavior, while the other vehicles follow UE behavior. The mixed equilibrium with UE and FO behaviors has unique solution if link cost function is a strictly monotone and linear function with regard to total link flow, and variational inequalities methods can be used to solve this mixed equilibrium (Harker, 1988). Yang et al. (2007) studied the mixed equilibrium with UE, SO and FO behaviors, and also used variational inequity methods to solve for the mixed equilibrium. A Stackelberg routing game where a SO player is the leader and UE and FO players are the followers were proposed, which is then compared to a Nash routing game in which UE, SO and FO players play against each other. Numerical results suggest the SO player can further improve efficiency with full knowledge of how other players would react in the Stackelberg game. Zhang and Nie (2018) and Chen et al. (2020) studied the mixed equilibrium where autonomous vehicles hypothetically follow the SO behavior and human-driving vehicles (HVs) follow the UE behavior. A small fraction of vehicles directed by SO behavior can drive the network system performance towards SO. In this research, we consider a more realistic network setting, a mobility service provider, e.g. TNC, would have a fleet of vehicles to route to its own benefit. Thus, a mixed equilibrium with UE and FO behaviors is studied, where driving vehicles follow UE and ride-hailing vehicles follow FO. Intuitively, this mixed equilibrium without system interventions (e.g. management measures) is not optimal for the network performance, because the fleet objective does not necessarily align with the system objective. This is exactly what is happening or is predicted to take place as TNCs, delivery companies, or navigation system markets continue to grow. Accordingly, we investigate what this mixed equilibrium entails for ride-hailing riders and how to provide incentives or compensations to ensure they are not worse off under the fleet objectives. Most importantly, we derive methods to suggest public agencies to regulate and subsidize TNC, in exchange for guaranteed network performance improvement, namely ORHP, that leverages ride-hailing vehicles to reduce system total travel time/cost.

To minimize system cost, different from the control methods (Ni and Cassidy, 2020; Xu et al., 2022), the proposed ORHP utilizes link-based subsidies (or equivalently surcharges) to influence vehicles routing. This is very relevant to the congestion pricing problem introduced in Yang and Huang (2005) from the public view. Ride-hailing fare pricing problems are summarized in Wang and Yang (2019), but is not the focus of this paper. The idea of congestion pricing was firstly proposed by Pigou (1920). The core concept behind congestion pricing is quite intuitive: highly congested road is charged a high price for road users accounting for high externalities, and less congested road is charged a low price or free of charge. By applying tolls (or equivalently credits) on roads, vehicles can be distributed in a desired efficient way. The well-known first-best pricing problem considers a scenario, where all the links can be tolled. In order to minimize the system cost, the toll equals the marginal external cost which is the difference between marginal cost and the individual cost (Pigou, 1920; Vickrey, 1963). When not all the links can be tolled, it comes to the second-best pricing problem. Most second-best pricing literature studies how to calculate the optimal tolls given a small set of links that are allowed to toll. A simple case in which the network has two parallel links and only one link can be tolled is extensively studied (Verhoef et al., 1996; Liu and McDonald, 1999).

In general, there are four types of congestion pricing schemes (May and Milne, 2000), travel-distance based scheme, travel-time based scheme, link-based scheme, and cordon-based scheme. In travel-distance or travel-time based schemes, the toll is proportional to travel distance or travel time. The toll in link-based scheme is implemented on selected links, such as bottlenecks and bridges. The cordon-based scheme aims to reduce the travel demand of a central area, and vehicles are charged when crossing the cordon (May et al., 2002; Mun et al., 2003). In our study, subsidies (or equivalent surcharges) are applied to all the links at the discretion of public agencies, but only for ride-hailing vehicles. This is analogous to surcharges to ride-hailing vehicles at airports or selected curb spaces. To our best knowledge, the most related work in literature is Delle Site (2021) which studied a congestion pricing scheme where only connected autonomous vehicles (CAVs) are charged. Our study differs from Delle Site (2021) in two ways. First, Delle Site (2021) assumed CAVs are willing to sacrifice individual benefits to improve fleet/system efficiency, while this study relaxed this assumption by proposing a novel cooperative fleet behavior (i.e., FOSC) that ensures fairness among travelers while reducing fleet travel time through cooperative routing. On the other hand, Delle Site (2021) studied two individual scenarios, including a zero-toll-sum scenario and a minimum-toll-sum scenario. In this study, we introduced a continuous parameter to trade off the system travel time reduction and the subsidy cost. Furthermore, Pareto frontiers for system travel time reduction and subsidy cost are provided, from which public agencies are able to adapt to a range of subsidy budgets or expected travel time reductions by setting up customized trade-off goals.

Intuitively, the congestion pricing problem except for the first-best pricing lends itself a classical bi-level optimization problem, where the upper level optimizes the tolls/subsidies, and the lower level calculates the network equilibrium given a set of tolls/subsidies. The bi-level optimization problem is commonly formulated as the mathematical programs with equilibrium constraints (MPEC), in which the lower level problem is regarded as an equilibrium constraint of the upper level problem (Harker and Pang, 1988). One approach to solve the MPEC is the sensitivity analysis based method (Tobin and Friesz, 1988; Yang, 1997; Yang and Huang, 2005). The gradients of equilibrium flows with regard to decision variables (i.e., tolls/subsidies) are calculated using sensitivity analysis of the restricted equilibrium problem. Then, the decision variables are updated using the gradients to optimize the objective function of the upper level. Another approach is the gap function based method (Chen and Florian, 1995; Marcotte and Zhu, 1996; Meng et al., 2001). The lower level problem is represented by a gap function, and then the gap function is treated as a penalty term of the objective function of the upper level problem. By doing so, the MPEC is reformulated as a differentiable single level optimization problem which can be solved by the standard optimization algorithms, such as the augmented Lagrangian algorithm. The sensitivity analysis based method has been widely used to solve bi-level problem with the low level problem being a single class traffic assignment (Tobin and Friesz, 1988; Yang, 1997). In this study, we adapt the sensitivity analysis based algorithm to solve the bi-level problem, however, with the low level problem being a multi-class mixed traffic assignment.

3. Mixed equilibrium of driving vehicles and ride-hailing vehicles

3.1. Problem formulation

Suppose there are two classes of vehicles, personal driving vehicles and ride-hailing vehicles, on the network. We do not consider travel mode choices or stochastic demands (e.g., Xu et al., 2023). Rather, we assume the personal driving demand and ride-hailing demand are exogenous and fixed for simplicity. Our study, as its first step, intends to propose a new concept and traffic management strategy, show its promises under arbitrary network flow demand of both private and ride-hailing vehicles, and demonstrate how it can be solved analytically in real-world. The driving vehicles behave per UE principle, while the ride-hailing vehicles follow a certain fleet behavior which is decided by a TNC platform as a whole. If the ride-hailing vehicles act on their own and there is no coordination among the TNC platform, then ride-hailing vehicles follow UE fleet behavior, and the whole system would reach UE. TNC would be motivated to coordinate its fleet vehicles to save total travel time, also known as FO behavior. The mixed equilibrium with driving vehicles following UE and ride-hailing vehicles following FO is denoted as ME-FO thereafter. Notations used in this paper are summarized in Table 1.

Note we are assuming a simultaneous game between the TNC platform vehicles and personal driving vehicles. Both the fleet of ride-hailing vehicles and personal driving vehicles are considered as players in a non-cooperative game where each player seeks to minimize their costs. Given the competition between ride-hailing fleet and personal driving vehicles, each group of vehicles is likely to adjust their behavior in response to the other group's behavior until a stable state is reached, and because they are infinitesimal players, this is known as Wardrop Equilibrium. At a Wardrop Equilibrium, no player can improve their generalized costs by unilaterally changing their respective strategies.

We denote $\mathbf{x}_D = (\dots, x_a^D, \dots)^T$ and $\mathbf{x}_R = (\dots, x_a^R, \dots)^T$ as the personal driving link flows and the ride-hailing link flows, where x_a^D and x_a^R denote the driving link flow and the ride-hailing link flow of the link a. Similarly, the driving demands and the ride-hailing demands are denoted as $\mathbf{q}_D = (\dots, q_D^{rs}, \dots)^T$ and $\mathbf{q}_R = (\dots, q_R^{rs}, \dots)^T$, where q_D^{rs} and q_R^{rs} denote the driving demand and the ride-hailing demand of OD pair rs. $\mathbf{f}_D = (\dots, f_D^{rs,k}, \dots)^T$ and $\mathbf{f}_R = (\dots, f_R^{rs,k}, \dots)^T$ denote the driving path flows and the ride-hailing path flows,

Table 1
Table of notations.

A	Set of links
R, S	Sets of origins and destinations
r, s	Indices of origins and destinations
K^{rs}	Set of paths between OD pair rs
k	Index of a path
a	Index of a link
$\delta_a^{rs,k}$	$\delta_a^{rs,k} = 1$ if link a is on path $k \in K^{rs}$ and 0 otherwise
x_a^D, x_a^R	Driving link flow and ride-hailing link flow of link a
q_D^{rs}, q_R^{rs}	Driving demand and ride-hailing demand of OD pair rs
$f_D^{rs,k}, f_R^{rs,k}$	Driving path flow and ride-hailing path flow of path k of OD pair rs
$x_a, q^{rs}, f^{rs,k}$	Total link flow of link a, total demand of OD pair rs, and total path flow of path k of OD pair rs
t_a	Link travel time of link a
τ_a	Subsidy of link a for ride-hailing vehicles
$c_D^{rs,k}$	Cost of path k of OD pair rs for driving vehicles
$c_R^{rs,k}$	Cost of path k of OD pair rs for ride-hailing vehicles (the cost function used in ride-hailing routing)
$cp_R^{rs,k}$	Cost of path k of OD pair rs for ride-hailing passengers
$d^{rs,k}$	Deviating compensation of path k of OD pair rs for ride-hailing passengers paid by TNC
FC	Generalized fleet cost for the fleet of ride-hailing vehicles
μ_u	The value of time of the ride-hailing passengers (e.g., 0.5 \$/min)
μ_t	The operating cost of travel time for the ride-hailing company (e.g., 1.5 \$/min)
μ_p	The average price of ride-hailing per unit travel time (e.g., 2.0 \$/min)
α	Demand ratio of ride-hailing vehicles such that $q_R^{rs} = \alpha q^{rs}$
τ_a	Subsidy of link a for ride-hailing vehicles
γ	Trade-off parameter of ORHP objective function
$\mathbf{x}_D, \mathbf{x}_R, \mathbf{q}_D, \mathbf{q}_R, \mathbf{f}_D, \mathbf{f}_R$	Vectors of $x_a^D, x_a^R, q_D^{rs}, q_R^{rs}, f_D^{rs,k}, f_R^{rs,k}$. For example $\mathbf{x}_D = (\dots, x_a^D, \dots)^T$
$\mathbf{x}, \mathbf{q}, \mathbf{f}$	Vectors of link flows, demands and path flows among two classes. For example, $\mathbf{x} = (\mathbf{x}_D^T, \mathbf{x}_R^T)^T$
$\mathbf{d}$	Vector of path-based compensations, $\mathbf{d} = (\dots, d^{rs,k}, \dots)^T$
$\boldsymbol{\tau}$	Vector of link-based subsidies, $\boldsymbol{\tau} = (\dots, \tau_a, \dots)^T$

where $f_D^{rs,k}$ and $f_R^{rs,k}$ denote the driving path flow and the ride-hailing flow on path k between OD pair rs. Given the driving demands $\mathbf{q}_D$ and ride-hailing demands $\mathbf{q}_R$, the feasible sets of driving link flows $\mathbf{x}_D$ and ride-hailing link flows $\mathbf{x}_R$ are given below:

$$\Omega_D = \{\mathbf{x}_D | x_a^D = \sum_{r\in R}\sum_{s\in S}\sum_{k\in K^{rs}} f_D^{rs,k}\delta_a^{rs,k}, q_D^{rs} = \sum_{k\in K^{rs}} f_D^{rs,k}, f_D^{rs,k} \geq 0, \forall a \in A, r \in R, s \in S, k \in K^{rs}\} \tag{1}$$

$$\Omega_R = \{\mathbf{x}_R | x_a^R = \sum_{r\in R}\sum_{s\in S}\sum_{k\in K^{rs}} f_R^{rs,k}\delta_a^{rs,k}, q_R^{rs} = \sum_{k\in K^{rs}} f_R^{rs,k}, f_R^{rs,k} \geq 0, \forall a \in A, r \in R, s \in S, k \in K^{rs}\} \tag{2}$$

where $\delta_a^{rs,k} = 1$ if link a is on path k of OD pair rs and 0 otherwise, A is the set of links, R is the set of origins r, S is the set of destinations s, and K^{rs} is the set of paths of OD pair rs. Path flows, link flows and demands satisfy following constraints:

$$x_a^D = \sum_{r\in R}\sum_{s\in S}\sum_{k\in K^{rs}} f_D^{rs,k}\delta_a^{rs,k}, x_a^R = \sum_{r\in R}\sum_{s\in S}\sum_{k\in K^{rs}} f_R^{rs,k}\delta_a^{rs,k}, \forall a \in A, r \in R, s \in S, k \in K^{rs} \tag{3a}$$

$$q_D^{rs} = \sum_{k\in K^{rs}} f_D^{rs,k}, q_R^{rs} = \sum_{k\in K^{rs}} f_R^{rs,k}, \forall r \in R, s \in S, k \in K^{rs} \tag{3b}$$

$$f_D^{rs,k} \geq 0, f_R^{rs,k} \geq 0, \forall r \in R, s \in S, k \in K^{rs} \tag{3c}$$

Driving vehicles aim to minimize individual travel time given $\mathbf{x}_R$, so the driving flows $\mathbf{x}_D$ can be obtained by solving:

$$\min_{\mathbf{x}_D\in\Omega_D} \sum_{a\in A} \int_0^{x_a^D} t_a(x + x_a^R)dx \tag{4}$$

where $t_a(\cdot)$ is the travel time function of link a with respect to the total link flow across two classes. If TNC adopts an FO behavior for ride-hailing vehicles, the ride-hailing flows are solved by minimizing fleet travel time:

$$\min_{\mathbf{x}_R\in\Omega_R} \sum_{a\in A} x_a^R t_a(x_a^R + x_a^D) \tag{5}$$

where $\mathbf{x}_D$ is given.

Solving optimization problems (4) and (5) simultaneously leads to a solution to ME-FO. To proceed, we introduce the following assumptions:

Assumption 1. Link travel time function is first-order and second-order differentiable. In particular, $t_a'(x_a) > 0$ and $t_a''(x_a) \geq 0, \forall a \in A, \forall x_a \geq 0$.

With Assumption 1, the objective functions of optimization problems (4) and (5) both are convex, implying congestion increases on each link more drastically when the link flow grows. Solving the optimization problems (4) and (5) is equivalent to solving variational inequalities (Nagurney, 1998). More specifically, Harker (1988) proved the following proposition.

Proposition 1. *(Harker, 1988) With Assumption 1, the vector $\mathbf{x}^* = (\mathbf{x}_D^{*T}, \mathbf{x}_R^{*T})^T$ is ME-FO if and only if the following variational inequality (VI) holds:*

$$\sum_{a\in A}[t_a(x_a^*)(x_a^D - x_a^{D*}) + (t_a(x_a^*) + x_a^{R*}t_a'(x_a^*))(x_a^R - x_a^{R*})] \geq 0, \forall \mathbf{x}^* \in \Omega \tag{6}$$

where $x_a^* = x_a^{D*} + x_a^{R*}$, and $\Omega = \Omega_D \times \Omega_R$.

Path cost function for personal driving vehicles reads,

$$c_D^{rs,k} = \sum_{a\in A} t_a(x_a)\delta_a^{rs,k} \tag{7}$$

where $x_a = x_a^D + x_a^R$ is the total flow of link a. Ride-hailing vehicles choose their respective routes to minimize the fleet travel time, so path cost function for ride-hailing vehicles in ME-FO reads,

$$c_R^{rs,k} = \sum_{a} \left(t_a(x_a) + x_a^R t_a'(x_a)\right)\delta_a^{rs,k} \tag{8}$$

A thorough discussion regarding VI formulation (6) and when its solution $\mathbf{x}^*$ meanwhile attains classical UE or SO is provided in Battifarano and Qian (2023). Now consider a general case where ME-FO is attained somewhere between the classical UE and SO.

3.2. Fleet optimal behavior with service constraint

If TNC hypothetically dictates ride-hailing vehicles to follow FO path assignments, clearly some passengers may be worse off deviating away from their respective optimal path. This degrades service quality, which can be perceived by passengers. In the short term, TNCs can benefit from saving total travel time, but this is not the best interest to TNCs in the long term if they could lose some consumers. Thus, we first propose a concept of service constraint (SC) for TNCs. With SC, a TNC pays compensations to those deviating passengers accounting for their individual-level additional travel time. Without loss of generality, we assume the average number of passengers in ride-hailing vehicles is one. Now, the generalized cost of ride-hailing passengers is given by:

$$cp_R^{rs,k} = \mu_u \sum_{a\in A} t_a(x_a)\delta_a^{rs,k} + p^{rs,k} - d^{rs,k} \tag{9}$$

where μ_u is the value of time (VOT) of riders, $p^{rs,k}$ is the ride-hailing fare calculated based on travel time and/or travel distance, and $d^{rs,k}$ is the compensation for anyone on path k between OD pair rs. For simplicity, we use a ride-hailing fair structure based on travel time, denoted as μ_p, thus $p^{rs,k} = \mu_p \sum_{a\in A} t_a(x_a)\delta_a^{rs,k}$. Note that $p^{rs,k}$ can always be generalized to embed other factors in the following formulations.

Assumption 2 (*Equal Generalized Travel Costs for all Riders*)**.** To guarantee service quality and fairness, we set that the generalized cost of any ride-hailing riders within each O–D pair must be equalized to the generalized cost of riders on the path(s) with the minimum travel time:

$$\mu_u \sum_{a\in A} t_a(x_a)\delta_a^{rs,k} + p^{rs,k} - d^{rs,k} = \min_{k\in K^{rs}}\left(\mu_u \sum_{a\in A} t_a(x_a)\delta_a^{rs,k} + p^{rs,k}\right), \forall k \in K^{rs} \tag{10}$$

Then, the deviation compensation reads,

$$d^{rs,k} = (\mu_u + \mu_p)\sum_{a\in A} t_a(x_a)(\delta_a^{rs,k} - \delta_a^{rs,k^*_{rs}}) \tag{11}$$

where $k^*_{rs} = \arg\min_{k\in K^{rs}} \sum_{a\in A} t_a(x_a)\delta_a^{rs,k}$.

With SC in place, the optimization problem for TNC is to minimize the total generalized fleet cost, denoted as FC, consisting of the fleet total travel time and the total compensation cost:

$$\min_{\mathbf{f}_R} \; FC = \mu_t \sum_{a\in A} x_a^R t_a(x_a) + \sum_r \sum_s \sum_k f_R^{rs,k} d^{rs,k} \tag{12a}$$

$$\text{s.t. Eqs. (3) and (11)} \tag{12b}$$

where personal driving demand $\mathbf{x}_D$ is given. We call such fleet assignment (in other words TNC's fleet routing behavior) fleet optimal with service constraint (FOSC). If optimization problems (4) and (12) are solved simultaneously, the solution is a mixed equilibrium of the UE and FOSC players, denoted as ME-FOSC. Note we assume a TNC is able to centrally assign routes to ride-hailing vehicles with compensations, so the path cost function used in TNC's routing is the marginal fleet cost:

$$\begin{aligned} c_R^{rs,k} &= \frac{\partial FC}{\partial f_R^{rs,k}} \\ &= \mu_t \sum_a \left(t_a(x_a) + x_a^R t_a'(x_a)\right)\delta_a^{rs,k} \end{aligned} \tag{13}$$

$$+ d^{rs,k} + (\mu_u + \mu_p) \sum_{r'} \sum_{s'} \sum_{k'} \left(f_R^{r's',k'} \sum_a t'_a(x_a) \delta_a^{rs,k} (\delta_a^{r's',k'} - \delta_a^{r's',k^*_{r's'}}) \right) \tag{14}$$

In theory, the TNC's goal may vary by the total cost or total revenue. However, in practice, most of those cost items boil down to the total travel time of the fleet, as any time savings can be directly translated to additional orders, trips and thus revenues and high review ratings. Uber, for instance, has KPIs on average ETA, the total number of trips completed given a period of time are directly related to total travel time, among many other KPIs. Therefore, travel time is the main consideration in the fleet cost function of FO fleet behavior. While with the FOSC fleet behavior, the fleet cost function also consider the compensation that ensures high service quality and balance fare/service time, in order to retain consumers in the long run.

3.3. Solution algorithms for mixed equilibrium

3.3.1. Mixed equilibrium-FO

ME-FO is obtained by solving two convex problems (4) and (5) simultaneously. There are mainly two types of methods to solve it, including diagonalization methods (Harker, 1988; Van Vuren et al., 1990) and method of successive average (MSA) (Sheffi and Powell, 1982; Van Vuren and Watling, 1991). In diagonalization methods, the problem is decomposed into two sub-problems. At a sub-problem, the ride-hailing (driving) flows are calculated assuming fixed driving (ride-hailing) flows. Two sub-problems are solved successively until convergence is reached. In the MSA methods, the optimal paths for vehicles are calculated at each iteration, which are used for all-or-nothing path flow assignments. The optimal paths for driving vehicles are paths with minimum path cost, while the optimal paths for ride-hailing vehicles are paths with minimum path marginal cost in Eq. (8). Here we used a path-based MSA to solve ME-FO for its simplicity, summarized in Algorithm 1.

Note that ME-FO can also be solved through the VI problem (6) that is then cast into an optimization model. This is particularly compelling if the link travel time function is not separable among flows on different links. However, solving ME-FO is not the focus of this study.

Algorithm 1: Path-based MSA for ME-FO

Initialization: Iteration $i = 0$, given demands $\mathbf{q}_D$ and $\mathbf{q}_R$, initialize path sets, path flows $(\mathbf{f}_D^0, \mathbf{f}_R^0)$, gap tolerances ϵ_D and ϵ_R, maximum iteration I

do

1. Given path flows $(\mathbf{f}_D^i, \mathbf{f}_R^i)$, update path costs
2. Find optimal paths k_D^{rs*} and k_R^{rs*}. Generate auxiliary path flow patterns $\mathbf{g}_D(\mathbf{f}_D^i, \mathbf{f}_R^i)$ and $\mathbf{g}_R(\mathbf{f}_D^i, \mathbf{f}_R^i)$ by assigning all demands onto k_D^{rs*} and k_R^{rs*}
3. $\mathbf{f}_D^{i+1} = \frac{i}{i+1}\mathbf{f}_D^i + \frac{1}{i+1}\mathbf{g}_D(\mathbf{f}_D^i, \mathbf{f}_R^i)$, $\mathbf{f}_R^{i+1} = \frac{i}{i+1}\mathbf{f}_R^i + \frac{1}{i+1}\mathbf{g}_R(\mathbf{f}_D^i, \mathbf{f}_R^i)$, and $i = i + 1$

while $(\text{GAP}_D > \epsilon_D$ **or** $\text{GAP}_R > \epsilon_R)$ **and** $i \le I$

Output $\mathbf{f}^* = (\mathbf{f}_D^i, \mathbf{f}_R^i)$

The gap functions in Algorithm 1 are defined as:

$$\text{GAP}_D = \frac{\sum_{r,s} \sum_k f_D^{rs,k} (c_D^{rs,k} - c_D^{rs,*})}{\sum_{r,s} \sum_k f_D^{rs,k} c_D^{rs,*}}, \ \text{GAP}_R = \frac{\sum_{r,s} \sum_k f_R^{rs,k} (c_R^{rs,k} - c_R^{rs,*})}{\sum_{r,s} \sum_k f_R^{rs,k} c_R^{rs,*}} \tag{15}$$

where * is the index for the optimal path. $c_D^{rs,k}$ is path travel time given in Eq. (7), whereas $c_R^{rs,k}$ is ride-hailing path cost in ME-FO given in Eq. (8).

3.3.2. Mixed equilibrium-FOSC

For ME-FOSC, Problem (12) is not convex because compensation $d^{rs,k}$ is dependent on flows. As such, ME-FOSC cannot be solved by applying existing solution algorithms, such as diagonalization methods and MSA. Thus, we propose a heuristic solution algorithm for ME-FOSC. Conceptually, we first fix compensations $\mathbf{d}$ to solve ME-FOSC for paths flows $\mathbf{f}(\mathbf{d})$. Then compensations $\mathbf{d}$ are updated using Eq. (11) given path flows $\mathbf{f}^*(\mathbf{d})$. These two steps are conducted iteratively until a convergence criterion is satisfied.

We first introduce how to solve ME-FOSC with fixed compensations $\mathbf{d}$. In this case, the total fleet cost is dependent on compensations $\mathbf{d}$. The optimization problem (12) becomes:

$$\min_{\mathbf{f}_R} \ FC(\mathbf{d}) = \mu_t \sum_{a \in A} x_a^R t_a(x_a) + \sum_r \sum_s \sum_k f_R^{rs,k} d^{rs,k} \tag{16a}$$

$$\text{s.t.} \ \ \mathbf{f}_R \text{ satisfies constraint (3)} \tag{16b}$$

Then, we write the path cost function of ride-hailing vehicles dependent on $\mathbf{d}$:

$$c_R^{rs,k}(\mathbf{d}) = \mu_t \sum_a \left(t_a(x_a) + x_a^R t'_a(x_a) \right) \delta_a^{rs,k} + d^{rs,k} \tag{17}$$

Proposition 2. *Under Assumptions 1 and 2, given compensations* $\mathbf{d}$*,* $FC(\mathbf{d})$ *is convex with respect to path flows* $\mathbf{f}_R$.

Proof. We show the convexity of $FC(\mathbf{d})$ by showing its Hessian matrix $\mathbf{H}$ is positive semi-definite. The secondary partial derivative of $FC(\mathbf{d})$ is given by:

$$\frac{\partial c_R^{rs,k}(\mathbf{d})}{\partial f_R^{r's',k'}} = \mu_t \sum_a \left(2t_a'(x_a) + x_a^R t_a''(x_a)\right) \delta_a^{rs,k} \delta_a^{r's',k'} \tag{18}$$

For simplicity, we use $\delta_{a,i}$ and $\delta_{a,j}$ to denote $\delta_a^{rs,k}$ and $\delta_a^{r's',k'}$. Thus, the Hessian matrix $\mathbf{H}$ of $FC(\mathbf{d})$ is given by:

$$\mathbf{H} = \begin{bmatrix} h_{11} & h_{12} & \cdots & h_{1n} \\ h_{21} & h_{22} & \cdots & h_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ h_{n1} & h_{n2} & \cdots & h_{nn} \end{bmatrix} \tag{19}$$

where $h_{ij} = \mu_t \sum_a \left(2t_a'(x_a) + x_a^R t_a''(x_a)\right) \delta_{a,i}\delta_{a,j}$.

As $\mathbf{H}$ is symmetric, $\mathbf{H}$ is positive semi-definite if $\mathbf{v}^T\mathbf{H}\mathbf{v} \geq 0, \forall \mathbf{v} \in \mathbb{R}^n$. $\mathbf{v}^T\mathbf{H}\mathbf{v}$ is given by:

$$\begin{aligned}
\mathbf{v}^T\mathbf{H}\mathbf{v} &= v_1 \sum_i^n h_{i1} v_i + v_2 \sum_i^n h_{i2} v_i + \cdots + v_n \sum_i^n h_{in} v_i \\
&= \sum_i^n \sum_{j \neq i}^n v_i^2 h_{ii} + 2 v_i v_j h_{ij} + x_j^2 h_{jj} \\
&= \mu_t \sum_i^n \sum_{j \neq i}^n \sum_a \left(2t_a'(x_a) + x_a^R t_a''(x_a)\right) \left(v_i^2 \delta_{a,i}^2 + 2 v_i v_j \delta_{a,i}\delta_{a,j} + v_j^2 \delta_{a,j}^2\right) \\
&= \mu_t \sum_i^n \sum_{j \neq i}^n \sum_a \left(2t_a'(x_a) + x_a^R t_a''(x_a)\right) \left(v_i \delta_{a,i} + v_j \delta_{a,j}\right)^2
\end{aligned}$$

By Assumption 1, $2t_a'(x_a) + x_a^R t_a''(x_a) > 0$, thus $\mathbf{v}^T\mathbf{H}\mathbf{v} \geq 0, \forall \mathbf{v} \in \mathbb{R}^n$. □

Proposition 3. *Under Assumptions 1 and 2, given fixed compensations* $\mathbf{d}$*, the vector* $\mathbf{f}^*$ *is the network equilibrium if and only if the following VI holds:*

$$\langle \mathbf{c}(\mathbf{f}^*), \mathbf{f} - \mathbf{f}^* \rangle \geq 0, \forall \mathbf{f} \textit{ satisfies constraint (3)} \tag{20}$$

where

$$\mathbf{c} = (\mathbf{c}_D^T, \mathbf{c}_R^T(\mathbf{d}))^T \tag{21}$$

$$\mathbf{f} = (\mathbf{f}_D^T, \mathbf{f}_R^T)^T \tag{22}$$

Proof. With Assumption 1 and fixed compensations $\mathbf{d}$, $FC(\mathbf{d})$ is convex with respect to $\mathbf{f}_R$ based on Proposition 2. Thus, the objective functions of the optimization problems (4) and (16) both are convex. Given the feasible spaces of problems (4) and (16) are convex and closed, solving problems (4) and (16) are equivalent to solving two corresponding VI problems. Besides, the feasible sets of $\mathbf{f}_D$ and $\mathbf{f}_R$ are disjoint, the two VI problems can be combined into a single VI problem (20). □

Then, ME-FOSC with fixed compensations can be obtained by solving the VI problem (20) using a path-based MSA algorithm similar to Algorithm 1. The solution algorithm for ME-FOSC with fixed compensations is summarized in Algorithm 2.

Algorithm 2: Path-based MSA for ME-FOSC with fixed compensations

---

Initialization: Iteration $i = 0$, given fix compensations $\mathbf{d}$, initialize path sets, path flows $(\mathbf{f}_D^0, \mathbf{f}_R^0)$, gap tolerances ϵ_D and ϵ_R, maximum iteration I

do

1. Given path flows $(\mathbf{f}_D^i, \mathbf{f}_R^i)$ and compensations $\mathbf{d}$, update path costs
2. Find optimal paths k_D^{rs*} and k_R^{rs*}. Generate auxiliary path flow patterns $\mathbf{g}_D(\mathbf{f}_D^i, \mathbf{f}_R^i)$ and $\mathbf{g}_R(\mathbf{f}_D^i, \mathbf{f}_R^i)$ by assigning all demands onto k_D^{rs*} and k_R^{rs*}
3. $\mathbf{f}_D^{i+1} = \frac{i}{i+1}\mathbf{f}_D^i + \frac{1}{i+1}\mathbf{g}_D(\mathbf{f}_D^i, \mathbf{f}_R^i)$, $\mathbf{f}_R^{i+1} = \frac{i}{i+1}\mathbf{f}_R^i + \frac{1}{i+1}\mathbf{g}_R(\mathbf{f}_D^i, \mathbf{f}_R^i)$, and $i = i + 1$

while $(\text{GAP}_D > \epsilon_D$ **or** $\text{GAP}_R > \epsilon_R)$ **and** $i \leq I$

Output $\mathbf{f}^*(\mathbf{d}) = (\mathbf{f}_D^i, \mathbf{f}_R^i)$

Note Algorithm 2 is the same as Algorithm 1, except that the optimal paths for ride-hailing vehicles are paths with minimum path cost $c_R^{rs,k}(\mathbf{d})$ defined in Eq. (17).

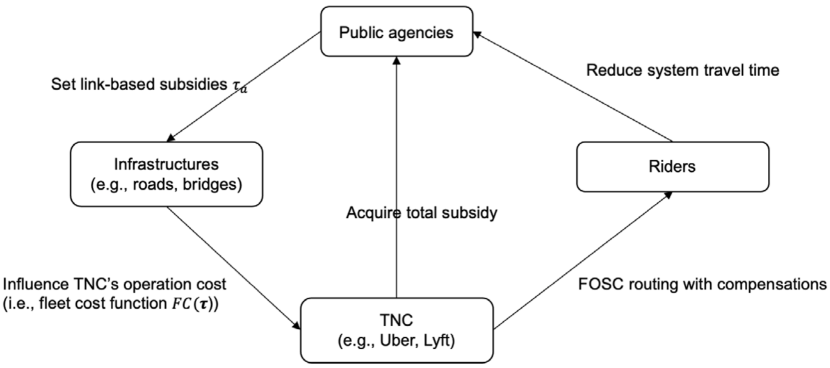


Fig. 1. The framework of ORHP.

Finally, the entire process and algorithm for solving ME-FOSC is summarized in Algorithm 3.

Algorithm 3: A heuristic solution algorithm for ME-FOSC

Initialization: Iteration $\upsilon = 0$, given demands $\mathbf{q}_D$ and $\mathbf{q}_R$, initialize path sets, compensations $\mathbf{d}^0$, path flows $\mathbf{f}^0$, the tolerance ϵ, maximum iteration V

do

1. Fix compensations $\mathbf{d}^\upsilon$, solve the mixed equilibrium $\mathbf{f}^*(\mathbf{d}^\upsilon)$ using Algorithm 2, and $\mathbf{f}^{\upsilon+1} = \mathbf{f}^*(\mathbf{d}^\upsilon)$
2. Fix path flows $\mathbf{f}^{\upsilon+1}$, and calculate auxiliary compensations $\mathbf{d}(\mathbf{f}^{\upsilon+1})$ using Eq. (11)
3. $\mathbf{d}^{\upsilon+1} = \frac{\upsilon}{\upsilon+1}\mathbf{d}^\upsilon + \frac{1}{\upsilon+1}\mathbf{d}(\mathbf{f}^{\upsilon+1})$, and $\upsilon = \upsilon + 1$

while $\|\mathbf{f}_R^\upsilon \odot (\mathbf{d}^\upsilon - \mathbf{d}^{\upsilon-1})\|_F^2 > \epsilon$ **and** $\upsilon \leq V$

Output $\mathbf{f}^\upsilon$ and $\mathbf{d}^\upsilon$

---

4. The optimal ride-hailing pricing scheme (ORHP)

Section 3 shows three types of fleet behaviors: (1) UE fleet behavior, (2) FO fleet behavior, and (3) FOSC fleet behavior. We would reasonably expect TNCs to choose a routing strategy close to FOSC among the three since FOSC balances both service quality and service cost. Depending on the travel demand and ride-hailing vehicles' penetration, FOSC may coincide with UE or FO without compensations. FOSC, in general, may or may not lead to the desired system performance for the general public. Next, we discuss, from the public agencies' stand point, how to set up regulations, e.g. subsidies and policies, to incentivize TNCs pushing the network performance towards SO. In other words, we would anticipate the public agencies to intervene ME-FOSC in such a way to benefit both TNC and the public.

We introduce the optimal ride-hailing pricing scheme (ORHP), where the public agency first set up the value of a subsidy on each link provided to TNCs (but not directly to riders). This means that for any rider served by a TNC, TNC will be subsidized for a mount that is the sum of link-based subsidies across all links along the rider's path. This subsidy scheme would affect ride-hailing vehicles' routing behaviors. For public agencies, the goal is to improve network performance. Fig. 1 depicts how ORHP works among public agencies, TNCs, riders and ride-hailing vehicles. Analogous to a surcharge to use airport curb spaces for TNCs, a monetary subsidy to TNCs can be issued for specific links, if a ride-hailing vehicle uses any of those links, most likely deviating routes. This is equivalent, in theory, to issuing a monetary surcharge (or other forms of a premium fee) to a ride-hailing vehicle on specific links. In the case of surcharge, this vehicle is willing to take the most "expensive" but efficient routes passing through those links. Note that this is fundamentally different from link-based congestion pricing, because riders have options to personally drive or choose from multiple routes from ride-hailing services priced differently, namely voluntary participation.

In the simplistic network equilibrium theory, subsidy and surcharge can be equivalent, as both are economic instruments to equalize generalized travel costs among travelers. However, we acknowledge their fundamental distinction when it comes to real-world implementation. In this study, we only use subsidy, instead of surcharge, because subsidy is more likely to attract TNCs to participate in this program.

We assume TNCs centrally provide routing options for ride-hailing vehicles, and the riders are willing to choose one of the assigned routes based on three considerations: (1) With FOSC, riders taking deviating routes will receive compensations set by TNCs; (2) TNCs can make the deviating routes even more appealing with higher compensations, if they receive subsidies from public agencies that encourage routing going through links with high subsidies; and (3) Not every rider has to deviate. TNCs can provide each rider with different route options, including the "shortest" route without compensations and deviated routes with compensations, and it is likely there are a sufficient number of riders deviating for some compensations when the compensations are attractive.

Practically, ORHP would be hard to game. TNCs receive subsidies from public agencies. They need to provide vehicle traces as an evidence, along with measurements on desired total travel time/delay (or time/delay reductions), in order to receive subsidies from links. Routing riders only to receive subsidies would not make sense, as TNCs' service is a lot more costly then the subsidy and riders would not appreciate any unnecessary deviation without the right compensations. Riders cannot game the system as well, because the compensation they receive is merely a very small fraction of their fare.

4.1. Bi-level formulation

The goal is for public agency to optimally distribute vehicles among all links/paths to minimize the total generalized system cost, including the system travel time cost and the subsidy cost. ORHP can be regarded as a Stackelberg game. The public agency is the leader who sets subsidies for ride-hailing services. Driving vehicles and ride-hailing vehicles are the followers. The ride-hailing subsidies affect the route choices of ride-hailing vehicles (on top of their desired FOSC behavior), which further affects the route choices of personal driving vehicles. By analyzing how all vehicles respond to ride-hailing subsidies, the public agency is able to set and leverage those ride-hailing subsidies to minimize the generalized system cost.

In particular, we design that a public agency would provide a subsidy of τ_a to TNC platform for any rider passing through the link a once. Note that this subsidy must be provided as part of the service fare, presumably substantially less than the original total fare for each rider. With the ORHP in place, a TNC platform would route their ride-hailing vehicles by incorporating the added benefits into FOSC. The ride-hailing routing problem (12) is modified as,

$$\min_{\mathbf{f}_R \in \Omega_R} \quad FC(\boldsymbol{\tau}) = \sum_{a\in A} x_a^R(\mu_t t_a(x_a) - \tau_a) + \sum_r \sum_s \sum_k f_R^{rs,k} d^{rs,k} \tag{23a}$$

$$\text{s.t. Eqs (3) and (11)} \tag{23b}$$

where τ_a is the subsidy of link a provided by the public agency for each ride-hailing trip using link a, provided with driving flow $\mathbf{f}_D$. Then, the ORHP is given by the following bi-level formulation.

$$\min_{\boldsymbol{\tau}} \quad z = \sum_a x_a t_a(x_a) + \gamma \sum_a x_a^R \tau_a \tag{24a}$$

$$\text{s.t.} \quad \tau_a \geq 0, \forall a \in A \tag{24b}$$

$$\text{Eqs. (3)} \tag{24c}$$

$$\mathbf{f}(\boldsymbol{\tau}) \text{ is a solution of problems (4) and (23) given } \boldsymbol{\tau} \tag{24d}$$

where γ is a parameter for trade-off between system travel time and subsidy cost, and $\boldsymbol{\tau}$ is the vectorized τ_a. γ is a flexible parameter to be set by the public agency. When γ is large, the goal is to use a small set of total subsidies in exchange for a small system improvement comparing to FOSC. However, setting γ to small implying an aggressive strategy: a best system performance close to SO is aimed, possibly at the price of high subsidies for the public agency.

Note $\boldsymbol{\tau}$ is given and fixed when solving for $\mathbf{f}(\boldsymbol{\tau})$. Thus, $\mathbf{f}(\boldsymbol{\tau})$ can still be solved using Algorithm 3 by replacing $c_R^{rs,k}(\mathbf{d})$ with $c_R^{rs,k}(\mathbf{d},\boldsymbol{\tau})$ which is given by,

$$c_R^{rs,k}(\mathbf{d},\boldsymbol{\tau}) = \sum_a \left(\mu_t(t_a(x_a) + x_a^R t_a'(x_a)) - \tau_a\right)\delta_a^{rs,k} + d^{rs,k} \tag{25}$$

To summarize, ORHP involves one leader player, that is, the public agency, and two follower players, TNC and the personal driving vehicles. The public agency aims to minimize the social cost of the whole network by setting up a link-based subsidy, formulated in the set of Eqs. (24). TNC aims to maximize its profit by strategically routing its fleet to minimize the entire fleet cost, formulated in Eq. (23). Personal driving vehicles follow the UE routing principle according to Eq. (4).

This paper considers two types of scenarios, including (1) system travel time minimization; and (2) generalized system cost minimization. In the system travel time minimization scenario, the public agency aims to minimize system travel time regardless the subsidy cost (i.e. $\gamma = 0$ in Eq. (24a)). This scenario gives an lower bound for the system travel time with ORHP, namely what is the best system performance ORHP can achieve, provided with the current TNC penetration. For the generalized system cost minimization scenario, the public agency makes a trade-off between a substantial benefit in system travel time and a reasonable subsidy cost (i.e. $\gamma > 0$ in Eq. (24a)).

4.2. Solution algorithm for ORHP

The ORHP problem (24) is non-convex because the equilibrium constraint (24d) is nonlinear. It is difficult to guarantee a global optimum or obtain a closed-form solution. Therefore, we developed a heuristic solution algorithm based on sensitivity analysis with respect to flow and link-based subsidy, in which $\nabla_{\boldsymbol{\tau}} z$ is calculated to update $\boldsymbol{\tau}$ is iteratively.

To calculate $\nabla_{\boldsymbol{\tau}} z$, we firstly calculate $\nabla_{\boldsymbol{\tau}}\mathbf{x}$ by extending the sensitivity analysis method in Yang and Huang (2005) to the multi-class traffic assignment problems, where $\mathbf{x} = (\mathbf{x}_D^T, \mathbf{x}_R^T)^T$. Sensitivity analysis of equilibrium evaluates the change in flow patterns with a small perturbation in decision variables (i.e., subsidies). We conduct such sensitivity analysis on a small perturbation of subsidies $\boldsymbol{\tau}$, with a given set of compensations $\mathbf{d}$, as part of the iterative solving process.

Let $\Delta = \begin{pmatrix} \Delta_D & 0 \\ 0 & \Delta_R \end{pmatrix}$ and $\Lambda = \begin{pmatrix} \Lambda_D & 0 \\ 0 & \Lambda_R \end{pmatrix}$ be the link/path and O–D/path incidence matrices. Note that only links with positive flows and equilibrated paths (i.e., paths with minimum costs) are considered in the incidence matrices. Now, $\mathbf{f}$ and $\mathbf{x}$ denote an equilibrium solution, and we have:

$$\begin{pmatrix} \Delta \\ \Lambda \end{pmatrix} \mathbf{f} = \begin{pmatrix} \mathbf{x} \\ \mathbf{q} \end{pmatrix} \tag{26}$$

Then, we only consider a maximum set of linearly independent paths in $(\Delta^T, \Lambda^T)^T$, and the reduced vectors and matrices are represented by $\tilde{\cdot}$. By the equilibrium condition and O–D demand conservation,

$$\tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau}) - \tilde{\Lambda}^T \boldsymbol{\pi} = 0 \tag{27}$$

$$\tilde{\Lambda}\tilde{\mathbf{f}} - \mathbf{q} = 0 \tag{28}$$

where $\boldsymbol{\pi}$ is the vector of the minimum path costs for two classes vehicles of all the O–D pairs. Take the derivative of above equations respect with to $\boldsymbol{\tau}$,

$$\begin{pmatrix} \nabla_{\tilde{\mathbf{f}}} \tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau}) & -\tilde{\Lambda}^T \\ \tilde{\Lambda} & 0 \end{pmatrix} \begin{pmatrix} \nabla_{\boldsymbol{\tau}} \tilde{\mathbf{f}} \\ \nabla_{\boldsymbol{\tau}} \boldsymbol{\pi} \end{pmatrix} = \begin{pmatrix} -\nabla_{\boldsymbol{\tau}} \tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau}) \\ 0 \end{pmatrix} \tag{29}$$

where $\nabla_{\tilde{\mathbf{f}}} \tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau})$ and $\nabla_{\boldsymbol{\tau}} \tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau})$ are calculated given $\mathbf{d}$ is fixed.

$$\nabla_{\tilde{\mathbf{f}}} \tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau}) = \begin{pmatrix} \tilde{\Delta}_D^T \mathrm{diag}(\cdots t_a'(x_a) \cdots) \tilde{\Delta}_D & \tilde{\Delta}_D^T \mathrm{diag}(\cdots t_a'(x_a) \cdots) \tilde{\Delta}_R \\ \tilde{\Delta}_R^T \mathrm{diag}\left(\cdots \mu_t(t_a'(x_a) + x_a^R t_a''(x_a)) \cdots\right) \tilde{\Delta}_D & \tilde{\Delta}_R^T \mathrm{diag}\left(\cdots \mu_t(2t_a'(x_a) + x_a^R t_a''(x_a)) \cdots\right) \tilde{\Delta}_R \end{pmatrix} \tag{30}$$

$$\nabla_{\boldsymbol{\tau}} \tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau}) = \begin{pmatrix} Z \\ -\tilde{\Delta}_R^T \end{pmatrix} \tag{31}$$

where Z denotes a zero matrix whose shape is same as $\tilde{\Delta}_D^T$. Let the inverse of the Jacobian $\mathbf{J}$ in Eq. (29) be

$$\mathbf{J}^{-1} = \begin{pmatrix} \nabla_{\tilde{\mathbf{f}}} \tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau}) & -\tilde{\Lambda}^T \\ \tilde{\Lambda} & 0 \end{pmatrix}^{-1} = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} \tag{32}$$

Then, from Eq. (29), we have

$$\nabla_{\boldsymbol{\tau}} \mathbf{x} = \tilde{\Delta} \nabla_{\boldsymbol{\tau}} \tilde{\mathbf{f}} = -\tilde{\Delta} B_{11} \nabla_{\boldsymbol{\tau}} \tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau}) \tag{33}$$

Finally, we obtain the gradient of z with regard to $\boldsymbol{\tau}$,

$$\nabla_{\boldsymbol{\tau}} z = \left(\frac{\partial z}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\tau}} + \frac{\partial z}{\partial \boldsymbol{\tau}} \right)^T \tag{34}$$

where

$$\frac{\partial z}{\partial \mathbf{x}} = \left(\cdots t_a(x_a) + x_a t_a'(x_a) \cdots, \cdots, t_a(x_a) + x_a t_a'(x_a) + \gamma \tau_a \cdots\right) \tag{35}$$

$$\frac{\partial \mathbf{x}}{\partial \boldsymbol{\tau}} = \nabla_{\boldsymbol{\tau}}^T \mathbf{x} \tag{36}$$

$$\frac{\partial z}{\partial \boldsymbol{\tau}} = \left(\cdots \gamma x_a^R \cdots\right) \tag{37}$$

After calculation of the gradients, the set of subsidies $\boldsymbol{\tau}$ is updated using the AdaGrad method, which is one of the most widely-used stochastic gradient descent method and often improves convergence speed by updating parameters with adaptive step sizes (Duchi et al., 2011). The proposed heuristic algorithm to solve the ORHP problem (24) is summarized in Algorithm 4.

Algorithm 4: Sensitivity analysis-based heuristic algorithm for ORHP

Initialization: Iteration $n = 0$, given demands $\mathbf{q}_D$ and $\mathbf{q}_R$, initialize subsidies $\boldsymbol{\tau}^0$, a tolerance ϵ, a maximum number of iterations N

do

1. Given subsidies $\boldsymbol{\tau}^n$, solve for FOSC equilibrium path flows $\mathbf{f}^n$ using Algorithm 3
2. Given $\mathbf{f}^n$, calculate gradient $\nabla_{\boldsymbol{\tau}} z$ based on the sensitivity analysis
3. Update $\boldsymbol{\tau}^n$ using AdaGrad
4. $n = n + 1$

while $n \le N$ and $\|\mathbf{x}_R \odot (\boldsymbol{\tau}^n - \boldsymbol{\tau}^{n-1})\|^2 > \epsilon$

Output $\boldsymbol{\tau}^n$ and $\mathbf{f}^n$

For a single-class static traffic assignment problem, Yang and Huang (2005) prove the Jacobian $\mathbf{J}$ is guaranteed to be invertible with the assumption that the link travel time function is positive, first-order differentiable and strictly monotone increasing with regard to link flows. For a multi-class traffic assignment, this no longer holds, namely the Jacobian $\mathbf{J}$ is no longer necessarily invertible under the same assumption. If $\mathbf{J}$ is not invertible, the system of linear Eqs. (29) has no or infinite solutions for $\begin{pmatrix} \nabla_{\tau}\tilde{\mathbf{f}} \\ \nabla_{\tau}\boldsymbol{\pi} \end{pmatrix}$. If the system of linear equations in Eq. (29) is feasible, we can still obtain one feasible gradient using Moore–Penrose pseudo-inverse of $\mathbf{J}$.

Nevertheless, we propose the following assumption to guarantee that the Jacobian $\mathbf{J}$ is invertible, which is mathematically more compelling while ensuring link travel time functions that capture traffic flow basics.

Assumption 3. Link travel time function $t_a(x_a)$ is a strictly monotone increasing and piece-wise linear function, $\forall a \in A$.

Note a piece-wise linear function can approximate any linear or nonlinear functions to any accuracy by adding intervals (Imamoto and Tang, 2008). Thus, Assumption 3 can be satisfied by approximating the empirical link travel time function with a piece-wise linear function. With Assumption 3, we propose the following proposition.

Proposition 4. *Given Assumption 3 and the columns in $(\tilde{\Delta}^T, \tilde{\Lambda}^T)^T$ are linearly independent, the Jacobian,* $\mathbf{J}$ *in Eq. (29) is invertible.*

Proof. To prove this theorem, it suffices to prove all the columns of $\mathbf{J}$ are linearly independent. Suppose the columns of $\mathbf{J}$ are linear dependent, so we can find a nonzero vector $\lambda = (\lambda_1^T, \lambda_2^T)^T$ such that $\mathbf{J}\lambda = 0$, where the length of λ_1^T equals to the number of paths in $\tilde{\Lambda}$, and the length of λ_2^T equals to the number of OD pairs in $\tilde{\Lambda}$. Then, we have

$$\nabla_{\tilde{\mathbf{f}}}\tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau})\lambda_1 - \tilde{\Lambda}^T\lambda_2 = 0 \tag{38}$$

$$\tilde{\Lambda}\lambda_1 = 0 \tag{39}$$

Multiply Eq. (38) by λ_1^T, and we have

$$\lambda_1^T\nabla_{\tilde{\mathbf{f}}}\tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau})\lambda_1 - \lambda_1^T\tilde{\Lambda}^T\lambda_2 = 0 \tag{40}$$

Substitute Eq. (39) into above equation, and we obtain

$$\lambda_1^T\nabla_{\tilde{\mathbf{f}}}\tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau})\lambda_1 = 0 \tag{41}$$

Under Assumption 3, within each interval, the first derivative and the second derivative of $t_a(x_a)$ are the slope and 0 respectively. When differentiating $t_a(x_a)$ at endpoints of intervals, it is reasonable to replace the sub-gradients with either slope of two sides. Thus, $\nabla_{\tilde{\mathbf{f}}}\tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau})$ is given by

$$\nabla_{\tilde{\mathbf{f}}}\tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau}) = \begin{pmatrix} \tilde{\Delta}_D^T\mathrm{diag}(\cdots t_a' \cdots)\tilde{\Delta}_D & \tilde{\Delta}_D^T\mathrm{diag}(\cdots t_a' \cdots)\tilde{\Delta}_R \\ \tilde{\Delta}_R^T\mathrm{diag}\left(\cdots \mu_t t_a' \cdots\right)\tilde{\Delta}_D & \tilde{\Delta}_R^T\mathrm{diag}\left(\cdots 2\mu_t t_a' \cdots\right)\tilde{\Delta}_R \end{pmatrix} \tag{42}$$

One can multiply the driving path cost function in Eq. (7) by the constant μ_t, which makes no difference with the behaviors of driving vehicles as driving demands are fixed. Then, $\nabla_{\tilde{\mathbf{f}}}\tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau})$ becomes

$$\nabla_{\tilde{\mathbf{f}}}\tilde{\mathbf{c}}(\mathbf{f}, \boldsymbol{\tau}) = \begin{pmatrix} \tilde{\Delta}_D^T\mathrm{diag}(\cdots \mu_t t_a' \cdots)\tilde{\Delta}_D & \tilde{\Delta}_D^T\mathrm{diag}(\cdots \mu_t t_a' \cdots)\tilde{\Delta}_R \\ \tilde{\Delta}_R^T\mathrm{diag}\left(\cdots \mu_t t_a' \cdots\right)\tilde{\Delta}_D & \tilde{\Delta}_R^T\mathrm{diag}\left(\cdots 2\mu_t t_a' \cdots\right)\tilde{\Delta}_R \end{pmatrix} \tag{43}$$

Then, Eq. (41) implies $\tilde{\Delta}\lambda_1 = 0$, which, combined with Eq. (39), implies $(\tilde{\Delta}^T, \tilde{\Lambda}^T)^T\lambda_1 = 0$. Since the columns in $(\tilde{\Delta}^T, \tilde{\Lambda}^T)^T$ are linearly independent, $\lambda_1 = 0$. Substituting $\lambda_1 = 0$ into Eq. (38) implies $\tilde{\Lambda}^T\lambda_2 = 0$. The rows of $\tilde{\Lambda}$ are linearly independent, so $\lambda_2 = 0$. Therefore, $\lambda = 0$, which contradicts with the fact that λ is a nonzero vector. □

5. Numerical examples

In this section, we conduct numerical experiments of two networks including Sioux Falls network and Pittsburgh network shown in Fig. 2. Sioux Falls network is obtained from the GitHub repository "Transportation Networks for Research" https://github.com/bstabler/TransportationNetworks. It has 24 nodes, 76 links and 58 OD pairs. Pittsburgh network is obtained from the planning model from Southwestern Pennsylvania Commission, which includes the proximity of downtown area of Pittsburgh. It has 247 nodes, 573 links and 247 OD pairs.

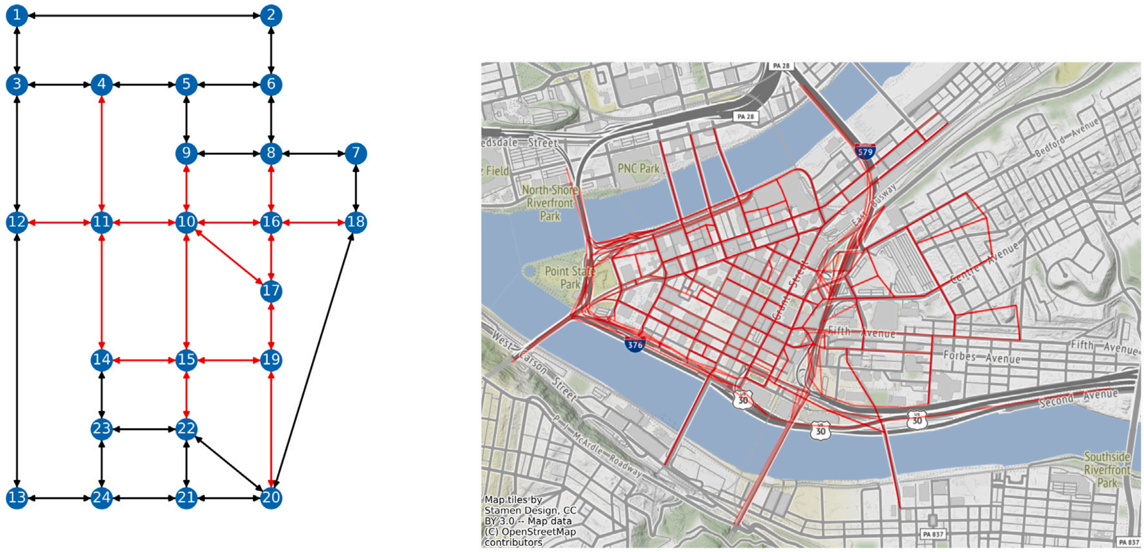
As Proposition 4 shows, the Jacobian of sensitivity analysis is guaranteed to be invertible when the link travel time function is a piece-wise linear function that monotonically increases w.r.t. link flows. For simplicity, we use the following linear cost function for Sioux Falls network for rigorous tests of the theory.

$$t_a = t_{a,0}\left(1 + \beta_a\left(\frac{x_a}{s_a}\right)\right) \tag{44}$$

where $t_{a,0}$ and s_a are the free flow travel time and the capacity of link a, and β_a is the parameter of the linear cost function of link a.

In Pittsburgh network, we consider a more practical link travel time function, namely the BPR functions.

$$t_a = t_{a,0}\left(1 + 0.15\left(\frac{x_a}{s_a}\right)^4\right) \tag{45}$$



(a) Sioux Falls network (b) Pittsburgh network

Fig. 2. Networks for numerical experiments.

Besides ride-hailing trips providing services to riders, idle and pickup flows generated by ride-haling vehicles also contribute to network congestion. We assume idle drivers are more likely to cruise around areas with high ride-hailing demands. Therefore, idle flows are generated proportional to the ride-hailing service trips. One can set the OD locations of the idle flows and the volume of the idle flows to reflect various searching strategies. In this study, the full list of OD locations of the idle flows is set to be same as the ride-hailing service trips, and the volume of idle flows is proportional to the ride-hailing trips for simplicity. The pickup flows are regarded as flows from the current location to the pickup location (i.e., the origin of the ride-hailing flows). The trip matching strategies are assumed to be constant and exogenous. Then, for each ride-hailing flow, we generate a pickup flow whose volume is proportional to the ride-hailing service trips. The destination of the pickup flow is the origin of the ride-hailing flow, while the origin of the pickup flow is sampled randomly from other nodes.

When solving the mixed equilibrium, we assume the idle flows and pickup flows, for their respective O–D pairs, follow the UE principle to choose respective routes. For idle cruising flows, we divide the whole cruising tours into several small trips that can be modeled by UE route choices. For the pickup flows, we expect that the TNCs aim to minimize the passenger waiting time, so the pickup vehicles are assigned with shortest paths (i.e., follow the UE). Those idle and pick up flow do not carry passengers, and thus are not impacted by fleet optimal routing or ORHP.

Ride-hailing services may induce more travel demands or may substitutes private vehicle trips in some cases. Though the effect may vary by locations, in this study, we assume the total travel demands are constant among various ride-hailing penetration rates, since the primary consideration of our study is to understand the effectiveness of fleet-optimum behavior and our proposed new management scheme ORHP.

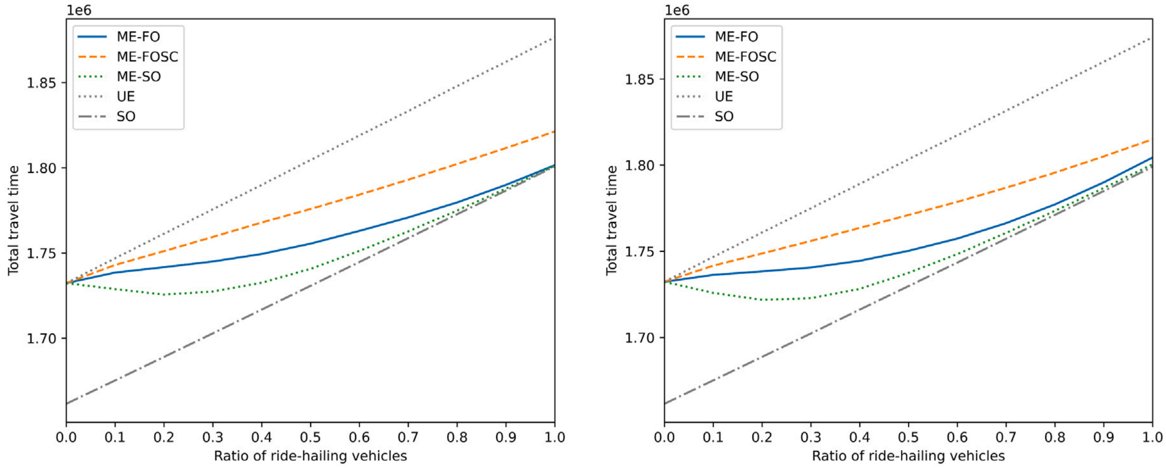
5.1. Impacts of ride-hailing vehicles on Sioux Falls network

To study the impacts of ride-hailing vehicles, we first analyze the mixed equilibrium state of the network without any interventions from the public agency. The driving vehicles always adapt to the UE routing behavior after day-to-day experience, while ride-hailing vehicles adopt UE or FO fleet behavior. To benchmark, we also include an ideal case where all ride-hailing vehicles follow SO fleet behavior.

Empirical studies indicate ride-hailing services are more prevalent in the downtown area (Marquet, 2020), so we compare a scenario with a homogeneous penetration rate and a scenario with a heterogeneous penetration rate. The homogeneous penetration rate means the ride-hailing demand ratio increases uniformly across all OD pairs, while the heterogeneous penetration rate means the penetration rates among OD pairs vary. In this case, we assume the penetration rate of the OD pairs that start or end in the downtown area is twice as high as the penetration rate of the other OD pairs in the heterogeneous penetration rate scenario.

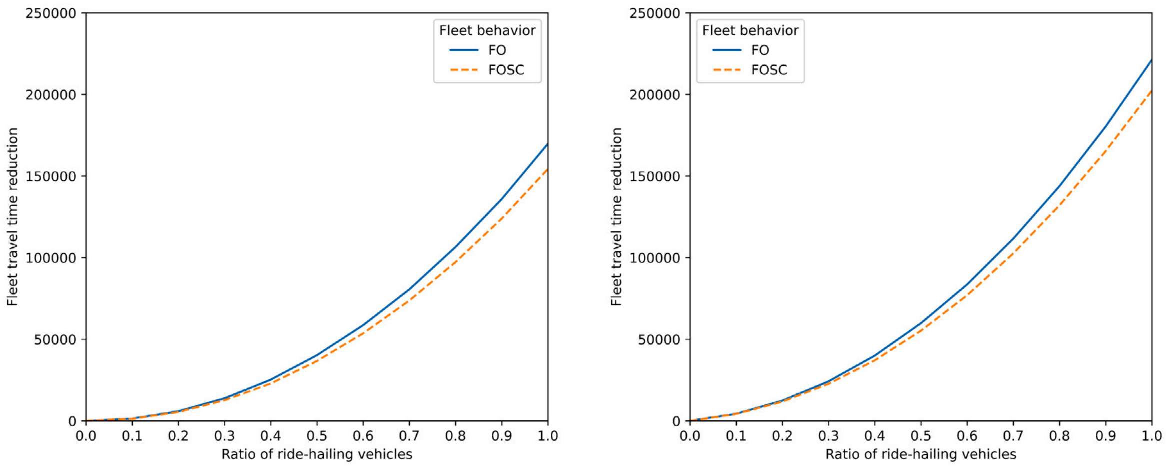
5.1.1. Network congestion

We first show the impacts of ride-hailing vehicles on system total travel time with different ride-hailing fleet behaviors and ride-hailing demand ratios. Conceivably, Fig. 3 shows that, for every penetration rate, ME-SO leads to the minimum system total travel time given the premise of all driving vehicles choosing the SO paths that benefit the system the most. If ride-hailing vehicles



(a) Homogeneous penetration rate (b) Heterogeneous penetration rate

Fig. 3. System total travel time of Sioux Falls network with respect to demand ratios.



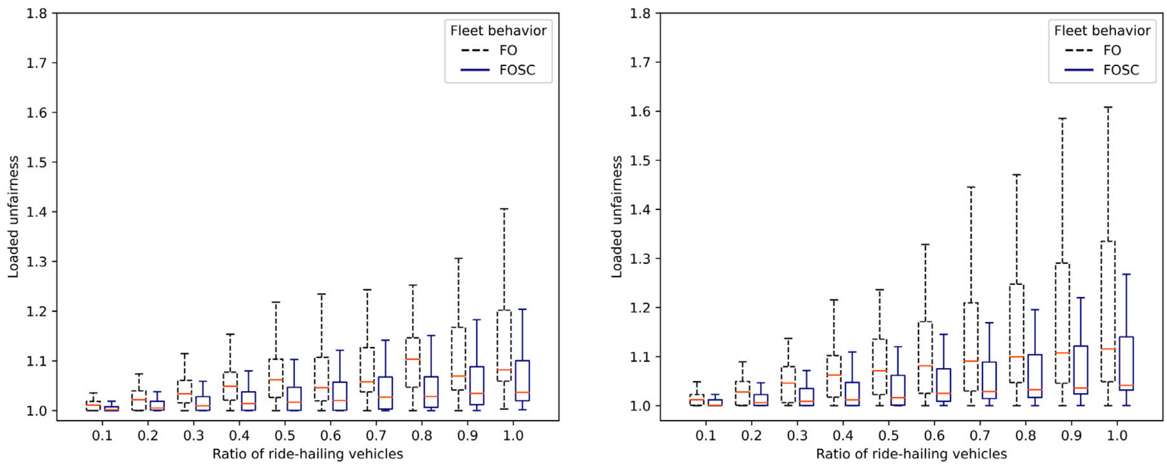
(a) Homogeneous penetration rate (b) Heterogeneous penetration rate

Fig. 4. Fleet travel time reduction from UE to ME-FO or ME-FOSC of Sioux Falls network with different demand ratios.

follow FO or FOSC fleet behavior, the system total travel time is smaller than that under UE (i.e. when all the vehicles adopt the UE travel behavior), and the reduction of total travel time is more pronounced as ride-hailing vehicles penetrate the demand more. Generally, ME-FO leads to lower system total travel time than ME-FOSC, implying the system benefits of having a significant portion of vehicles being TNC fleet would be discounted if TNC considers equity and service quality among all its riders. For each fleet behavior, the total travel time generally increases as the penetration rate increases because, besides the ride-hailing trips that service riders, the ride-hailing vehicles generate idle flows and pickup flows that also contribute to network congestion to some degree. Note these findings are consistent with the homogeneous penetration rate scenario and the heterogeneous penetration rate scenario.

5.1.2. Fleet travel time of ride-hailing vehicles

Fleet total travel time of all ride-haling vehicles largely determines the operating costs and quality of TNCs, and thus we examine the this performance metric for TNCs. Theoretically, among UE, FO and FOSC fleet behaviors, the UE fleet behavior leads to the maximum fleet travel time because there is no cooperation among ride-hailing vehicles, while the fleet travel time is minimized when a FO fleet behavior is adopted. Fig. 4 shows, in both penetration scenarios, the fleet travel time of FO fleet behavior and FOSC fleet behavior when compared with the UE fleet behavior. Similar to FO behavior, FOSC behavior also significantly reduces the fleet total travel time because the ride-hailing vehicles behave cooperatively. In addition, the reduction of fleet total travel time



(a) Homogeneous penetration rate (b) Heterogeneous penetration rate

Fig. 5. Travel time variability in Sioux Falls network with respect to demand ratios (for better visualization, only vehicles with loaded unfairness larger than 1 are included).

with FO behavior or FOSC behavior become more pronounced as the total number of ride-hailing vehicles grows. Note that a lower fleet travel time implies a lower average ride-hailing travel time, which is also desirable for ride-hailing passengers.

Comparing the results with the homogeneous penetration rate and the heterogeneous penetration rate, TNC can save more fleet travel time in the heterogeneous penetration rate scenario. This finding indicates OD pairs have heterogeneous impacts on fleet travel time reduction. Thus, TNC may further reduce fleet travel time by strategically allocating fleet vehicles to certain O–D pairs that attain more benefits than others.

5.1.3. Fairness: Travel time variability among ride-hailing passengers

In addition to the average ride-hailing travel time, the travel time variability among ride-hailing passengers within the same OD pair also affects the mobility service quality, as well as fairness among riders. To measure the travel time variability among ride-hailing vehicles, we use "loaded unfairness" defined by Jahn et al. (2005). Loaded unfairness for a vehicle is the ratio of its experienced travel time to the experienced travel time of the fastest vehicle for the same OD pair. Fig. 5 plots the loaded unfairness of ride-hailing vehicles with FO fleet behavior or with FOSC fleet behavior. With FO fleet behavior, when TNC considers solely the total fleet time, the loaded unfairness seems quite large, and it expands as the ride-hailing fleet size grows. On the contrary, FOSC behavior, if adopted by TNC, leads to a much smaller loaded unfairness value than the FO behavior, and the loaded unfairness is well upper bounded even if TNC accounts for all vehicles in the network. Besides, with FOSC fleet behavior, ride-hailing vehicles with loaded unfairness greater than 1 will receive compensation for their deviation. Note the compensation amount is relatively low compared to the total fare for most ride-hailing trips. The portion of compensation included in the ride-haling fare is calculated as (loaded unfairness − 1) $*$ 100%. Even with the ride-hailing penetration rate of 100%, the compensation makes up less than 20% of the ride-hailing fare for all trips, and less than 10% for 75% of all trips. Therefore, FOSC leads to a less variable travel time and overall higher service quality among all ride-hailing riders than FO. Overall, the loaded unfairness in the heterogeneous penetration rate scenario is higher than in the homogeneous penetration rate scenario. This finding indicates TNC deviates riders more to attain the minimum fleet travel time in the heterogeneous penetration rate.

5.2. ORHP on the Sioux Falls network

The proposed ORHP is tested on Sioux Falls network with the ride-hailing demand ratio at 50%. As for the penetration rate, we assume a homogeneous penetration rate for ORHP to avoid being lengthy. The system travel time minimization and the generalized system cost minimization are implemented.

5.2.1. System travel time minimization

The convergence of system total travel time and the subsidy cost is shown in Fig. 6. The system total travel time converges quickly, and the ultimate value is very close to the optimal system state in which all ride-hailing vehicles follow SO fleet behavior. This implies that with a substantial subsidy from the public agency, the TNC service can be regulated effectively towards the most desired system optimum state.

Fig. 7 shows how ORHP affect the link volume-to-capacity (V/C) ratios by imposing subsidies on links. When no subsidy is applied (i.e. ME-FOSC), links 17–19 and 10–16 are the most congested as indicated by Fig. 7(a). ORHP sets the highest subsidies

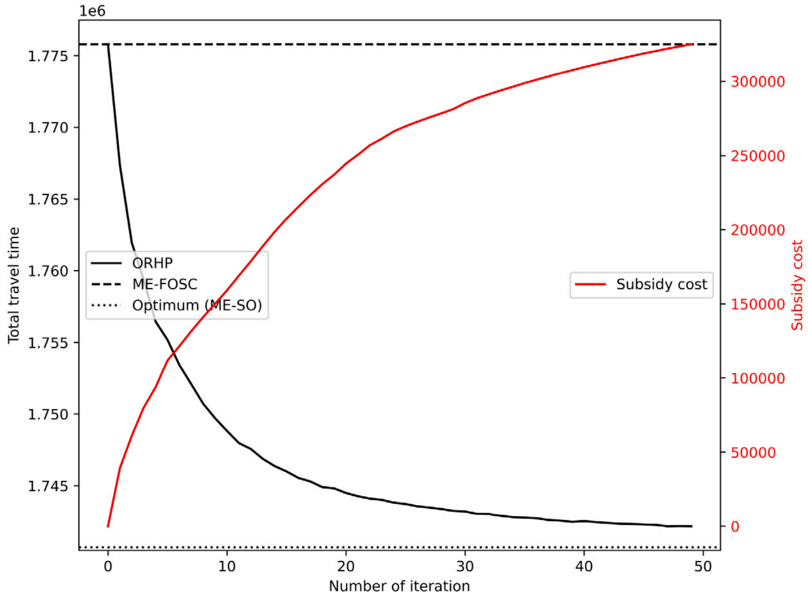


Fig. 6. System total travel time and subsidy cost of Sioux Falls network with ORHP in the system travel time minimization scenario ($\gamma = 0$).

on links 10–9, 9–8, and 8–16, which are not congested and geographically close to those most congested links (see Fig. 7(b)). As a result, V/C ratios of the most congested links are substantially reduced, and V/C ratios of the those less congested links climb (see Fig. 7(c)). Overall, ORHP improves the network efficiency by leveraging link-based subsidies to balance the flow between congested and no-so-congested links. Comparing ORHP to the ME-SO (see Figs. 7(c) and 7(d)), the V/C ratio changes by ORHP are quite similar to the V/C ratio changes by ME-SO, which indicates ORHP can attain almost optimal network performance on Sioux Falls network, and the proposed Algorithm 4 is able to solve ORHP efficiently.

Due to the non-convexity of the ORHP, the solution may lead to multiple local optima, we examine solutions of the ORHP with different starting points. At the initialization, for each link, a subsidy is generated through random sampling from [0, 4], which provides ORHP an initial solution to iterate. Fig. 8 plots the results of ORHP with different initial values of subsidies. Most of these converged solutions are similar. All three converged solutions are consistent in the result that links 10–9, 9–8, and 8–16 are the most subsidized links. Though differences can be observed on several links (e.g., 11–4), they can be attributed to the slow convergence sourcing from the low V/C ratios and a very small impact on the total travel time. Reasonably, we can conclude that even with totally different initial values, ORHP in this test network converges to a small neighborhood of a local optimum and possibly a global optimum. Practically, the observed differences with different initial values seem insignificant.

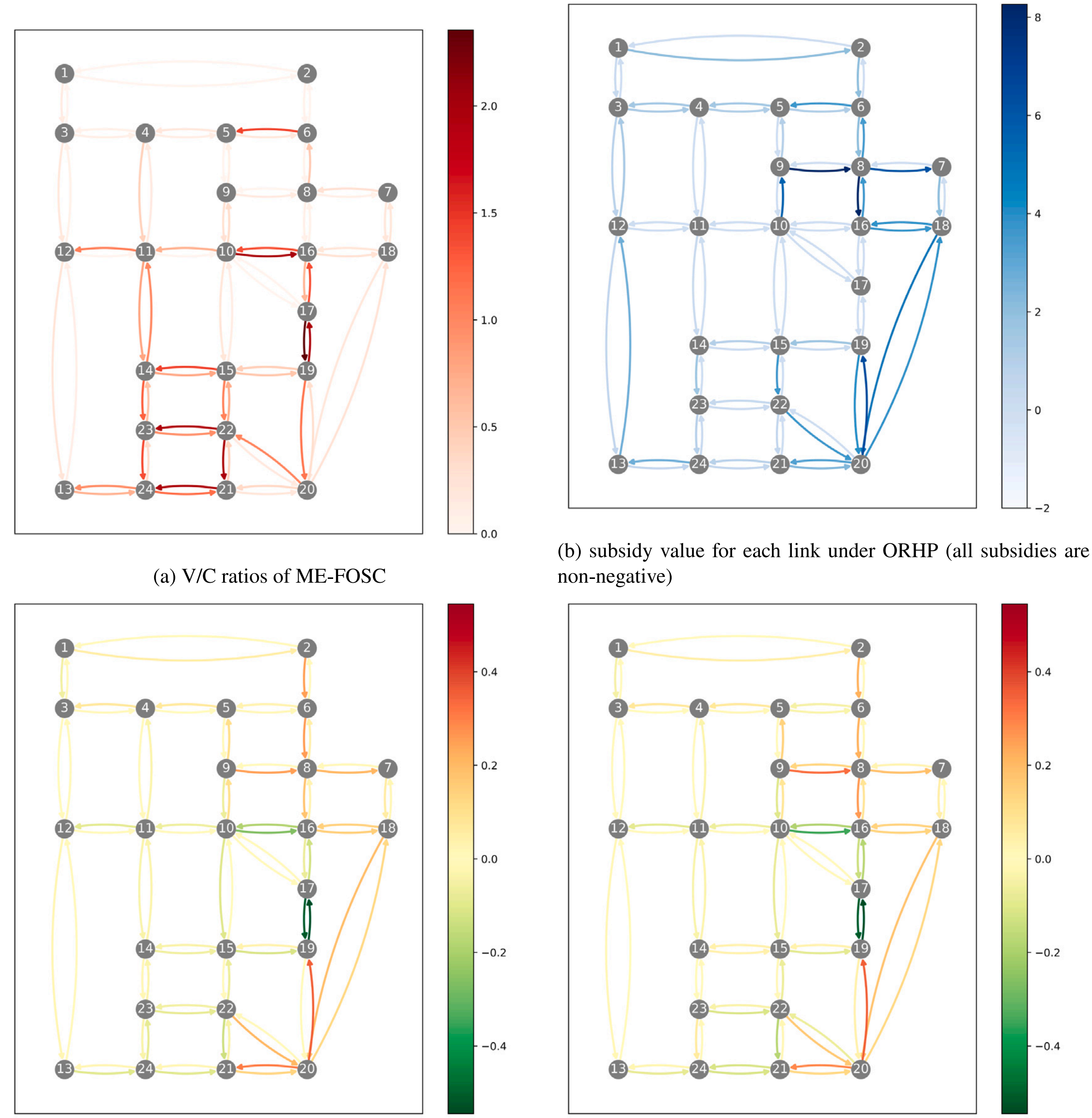
5.2.2. Generalized system cost minimization

When minimizing the generalized system cost, the parameter γ quantifies the trade-off between the system travel time cost and the subsidy cost. We investigate how the value of γ leads to a desired subsidy in exchange for a guaranteed system performance improvement.

The generalized cost consists of the system travel time and the subsidy cost multiplied by γ. As Fig. 9(a) shows, the generalized cost drops and converges quickly with various values of γ. A less γ implies an increased investment from the public agency, and leads to an improved system performance. Fig. 10 plots the Pareto frontier for the system travel time and the subsidy cost. If the public agency plans to reduce the system travel time to a greater degree, more substantial subsidy cost is needed. As the reduction of system travel time is pushed towards the maximum point ($\gamma = 0$), the marginal efficiency of subsidies drops. A most cost-effective subsidy, in this case, falls between \$38,701 to \$136,584 when γ is between 0.1–0.4, and the resultant total travel time reduction is from 16,813 h to 30,299 h.

The subsidy values on all links with selected values of γ are compared in Fig. 11. Generally, the spatial patterns of the subsidy look fairly consistent with respect to the values of γ. With a greater γ, the values of the subsidies are smaller, and the number of those subsidized links is less too. The overall subsidy is proportional among all links.

The path-level subsidies (i.e., subsidies for ride-hailing trips) under different values of γ are plotted in Fig. 12. As γ grows, the average path subsidy for each OD pair decreases substantially and the path-level subsidy can be limited to be under \$5 with $\gamma \geq 0.4$. In this case, the number of subsidized OD pairs is also small. With $\gamma \geq 0.4$, for every ride-hailing trip, the ratio of the path subsidy to the ride-hailing fare is lower than 0.3, and for over 80% of ride-hailing trips, the ratio is below 0.1. These findings suggest that if the budget of the public agency is limited, relatively small subsidies imposed on a small fraction of links or OD pairs may suffice to minimize the system generalized cost effectively.



(a) V/C ratios of ME-FOSC

(b) subsidy value for each link under ORHP (all subsidies are non-negative)

(c) V/C ratio changes from ME-FOSC to ORHP (negative values mean V/C ratios are reduced by ORHP)

(d) V/C ratio changes from ME-FOSC to ME-SO (negative values mean lower V/C ratios under ME-SO)

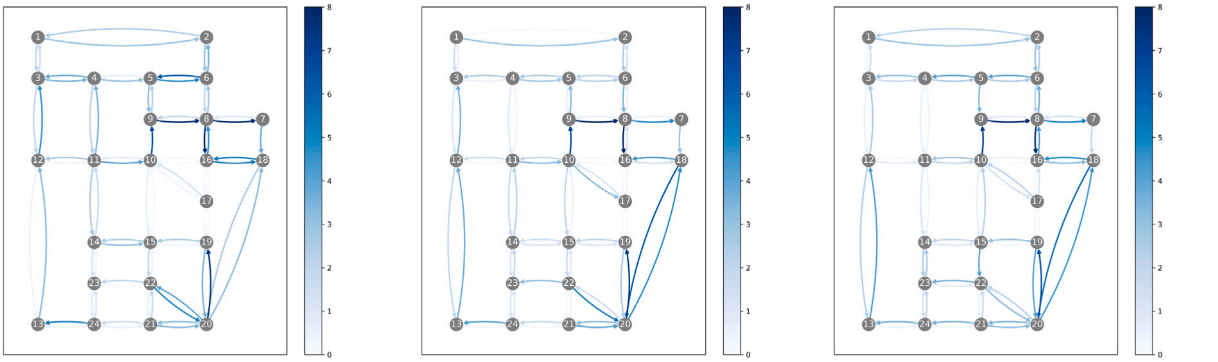
Fig. 7. V/C ratios and subsidies for all links in the Sioux Falls network, when minimizing solely the system travel time ($\gamma = 0$).

5.3. Pittsburgh network

In this subsection, we analyze the impacts of the ride-hailing vehicles and the proposed ORHP on a larger network (i.e. the Pittsburgh network in Fig. 2(b)). The penetration rate is assumed to be homogeneous among OD pairs for simplicity.

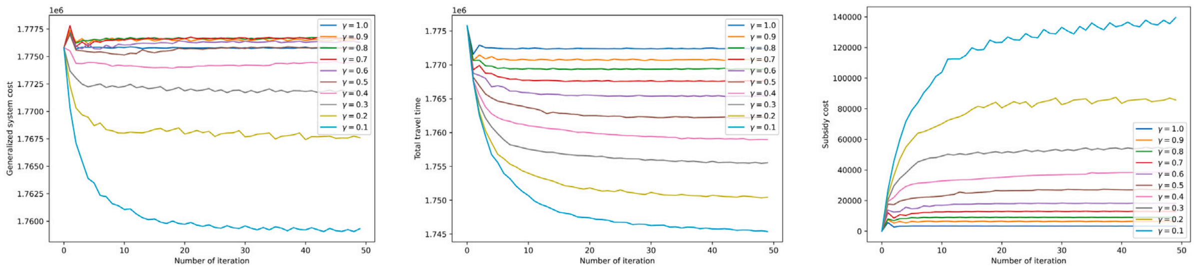
5.3.1. Impacts of ride-hailing vehicles on Pittsburgh network

The equilibrium states of Pittsburgh network with the ride-hailing vehicles following different fleet behaviors are compared in Fig. 13. As shown in Fig. 13(a), when the ratio of ride-hailing vehicles is greater than or equal to 0.8, the system travel time of ME-FO is identical to ME-SO. This implies that 80% penetration (uniformly applied to all O–D pairs in this example) of TNC fleet following the fleet optimum in the Pittsburgh network is able to attain the SO network state for everyone, as if every vehicle intends to route for SO. Among ME-FO, ME-FOSC and ME-SO, the system travel time of ME-FOSC is the greatest consistently under various values of penetration rates, implying TNC would have to trade some marginal system efficiency for rider fairness (ultimately the



(a) Solved subsidies with initial subsidies 1 (b) Solved subsidies with initial subsidies 2 (c) Solved subsidies with initial subsidies 3

Fig. 8. Results of the ORHP with different initial subsidies on the Sioux Falls network. (The initial subsidies are generated randomly with certain random seed.).



(a) generalized cost (b) system travel time (c) subsidy cost

Fig. 9. Costs in Sioux Falls network under ORHP, when minimizing total generalized cost ($\gamma > 0$).

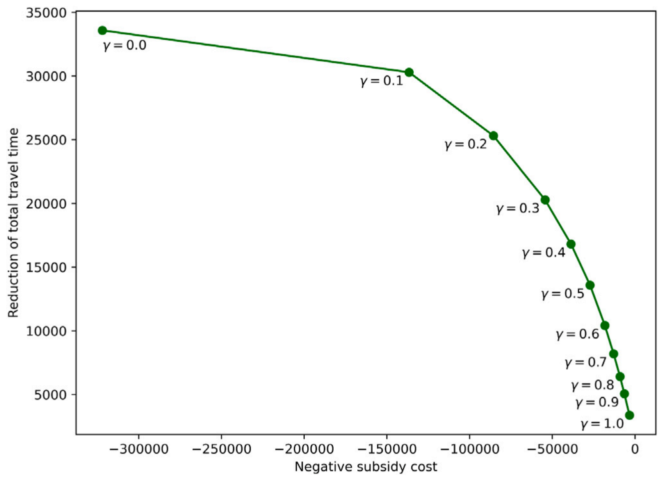
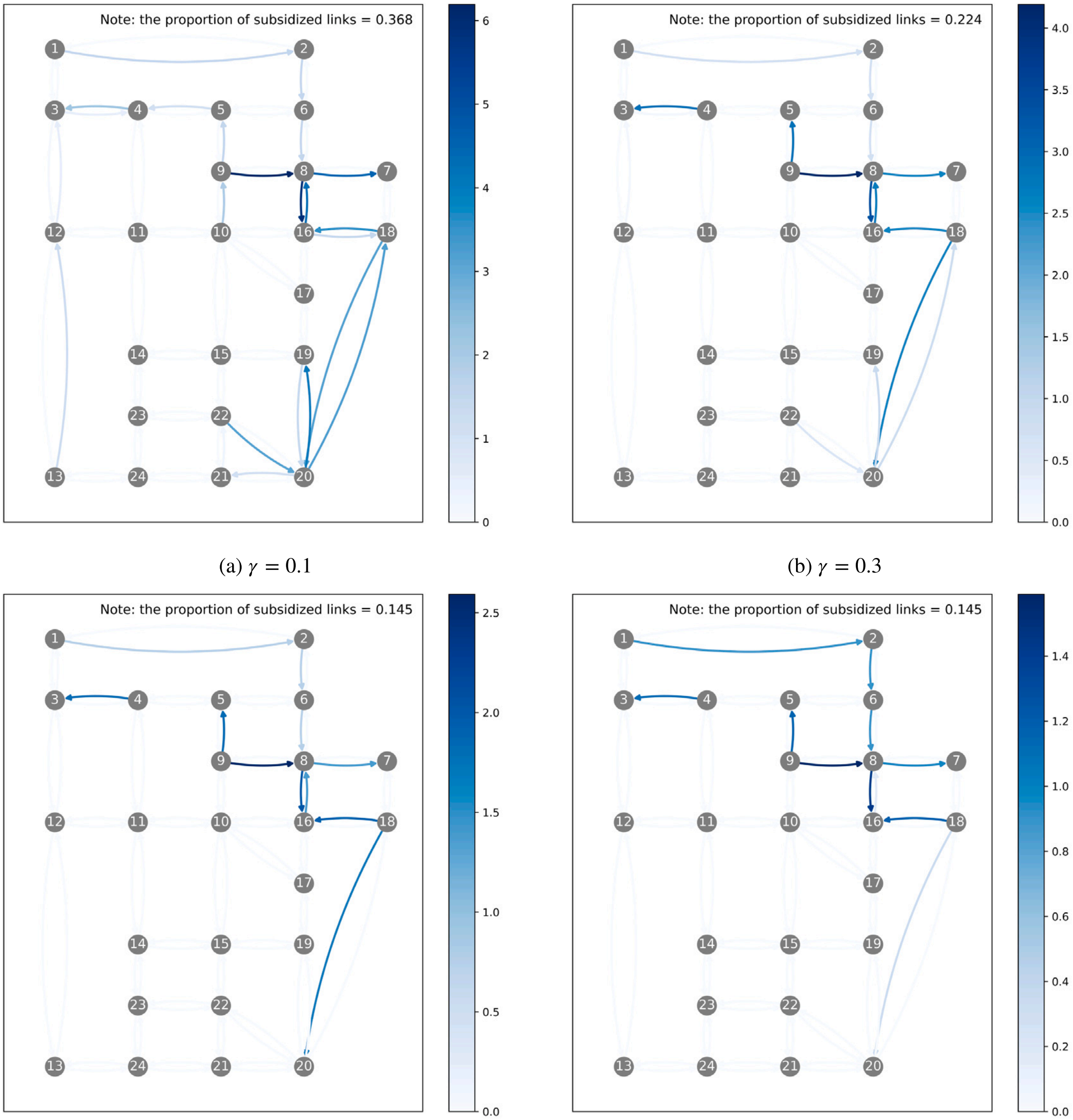


Fig. 10. Pareto frontier for system travel time and subsidy cost on Sioux Falls network.

overall service quality). For each fleet behavior, the total travel time increases as the penetration rate increases because of extra idle and pickup flows generated by ride-hailing fleet. FOSC fleet behavior is able to reduce the fleet travel time to the extent very close to the FO fleet behavior, indicated in Fig. 13(b). On the other hand, FO fleet behavior would have relatively higher loaded unfairness among ride-hailing vehicles, which can be effectively constrained when TNC adopts the FOSC fleet behavior. Generally, these results are consistent with the results from the Sioux Falls network.



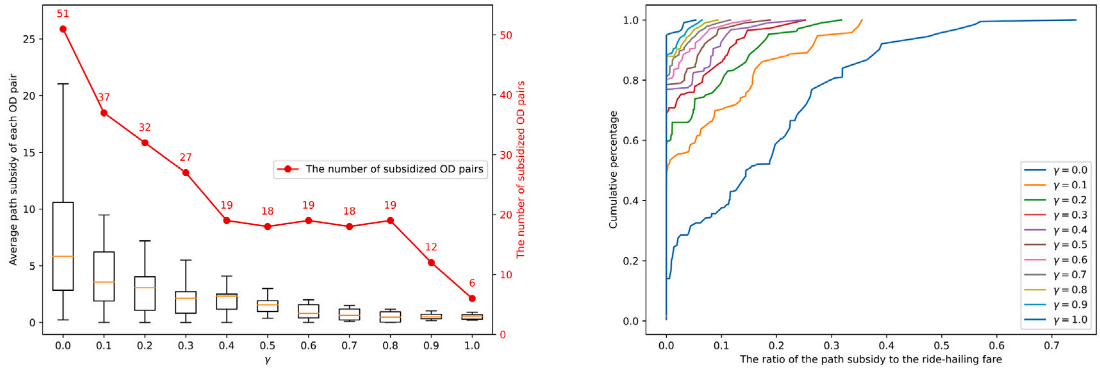
(a) $\gamma = 0.1$ (b) $\gamma = 0.3$ (c) $\gamma = 0.5$ (d) $\gamma = 0.7$

Fig. 11. Link-based subsidy in Sioux Falls network when minimizing the total generalized cost ($\gamma > 0$).

5.3.2. *ORHP on the Pittsburgh network*

The performances of ORHP on the Pittsburgh network are shown in Fig. 14. When minimizing solely the system travel time (i.e. $\gamma = 0$), though the total travel time does not reach the system optimal value, the total travel time declines quickly along those iterations (indicated by Fig. 14(a)). As shown in Fig. 14(b), the generalized cost decreases and converges quickly under various values of γ. The Pareto frontier for system travel time and subsidy cost on Pittsburgh network is plotted in Fig. 14(c), which depicts a trade-off between subsidy cost and system performance improvement, similar to that of the Sioux Falls network. In the Pittsburgh network, it seems a most cost-effective subsidy would be around \$1000 when γ is around 0.3. It can reduce over 50% of the maximum travel time savings from UE to SO with only a small amount of subsidy.

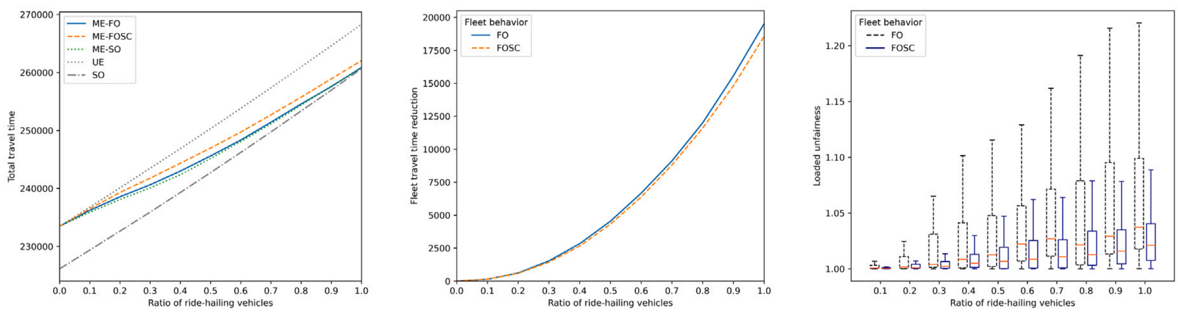
To investigate how ORHP affects the vehicle routing behavior, V/C ratios and subsidies for all links in the Pittsburgh network are shown in Fig. 15. The ORHP sets subsidies to those under-congested links that are adjacent to those over-congested links (shown in Figs. 15(a) and 15(b)). The V/C ratio changes from ME-FOSC to ORHP show a pattern similar to those from ME-FOSC to ME-SO. Therefore, we conclude that the proposed ORHP can effectively improve the efficiency of the Pittsburgh network by allocating vehicles between over-congested links and under-congested links, all through TNC vehicles only. Figs. 16 and 17 also indicate the system efficiency can be improved by imposing pretty small subsidies onto a small portion of links and OD pairs.



(a) Average path subsidy of each OD pair and the number subsidized OD pairs

(b) Cumulative percentage of the ratio of the path subsidy to the ride-hailing fare

Fig. 12. Path-level subsidies on Sioux Falls network.

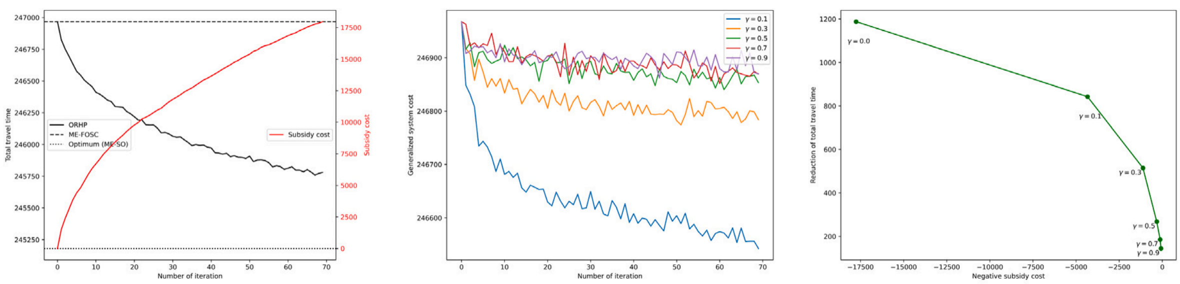


(a) System travel time

(b) Fleet travel time reduction

(c) Travel time variability

Fig. 13. Impacts of ride-hailing vehicles on Pittsburgh network.

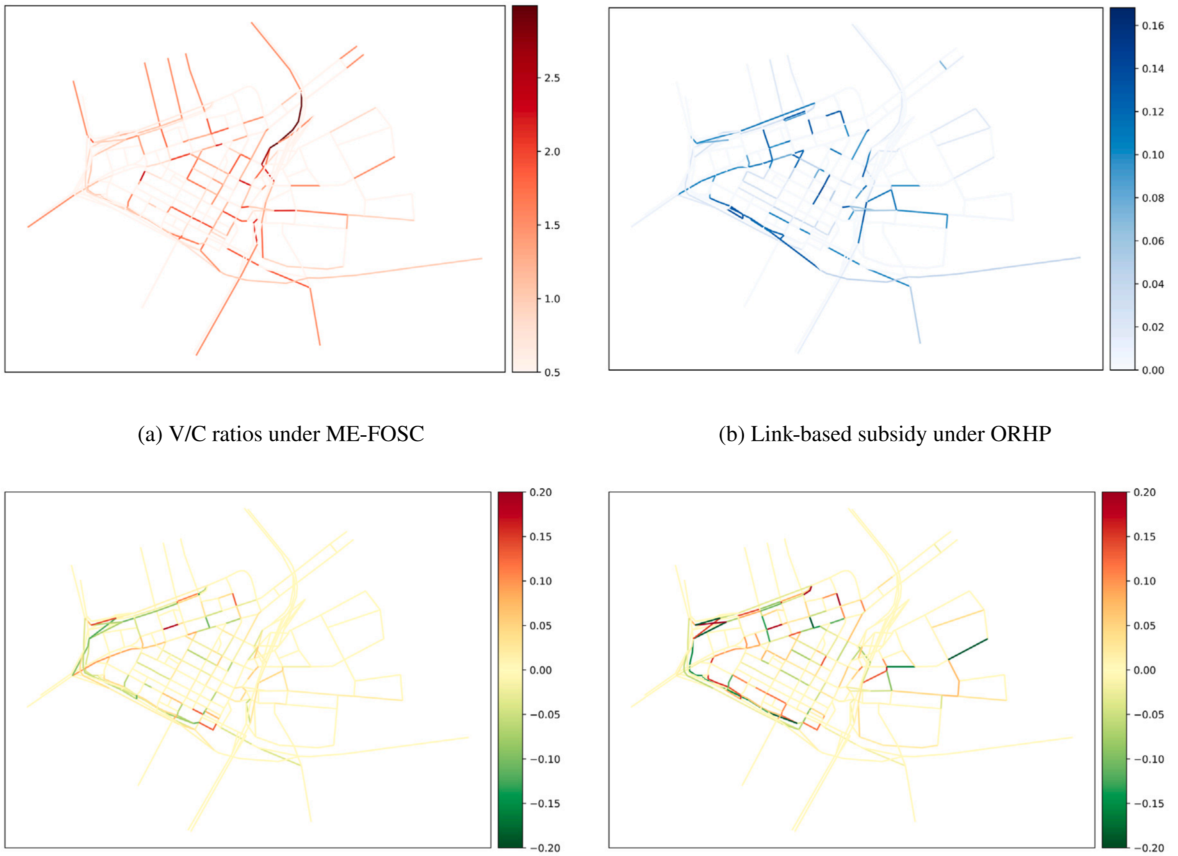


(a) Minimizing system travel time only($\gamma = 0$)

(b) Minimizing the total generalized cost ($\gamma > 0$)

(c) Pareto frontier for system travel time and subsidy cost

Fig. 14. Performances of ORHP on the Pittsburgh network.



(a) V/C ratios under ME-FOSC

(b) Link-based subsidy under ORHP

(c) V/C ratio changes from ME-FOSC to ORHP (negative values mean V/C ratios are reduced by ORHP)

(d) V/C ratio changes from ME-FOSC to ME-SO (negative values mean lower V/C ratios under ME-SO)

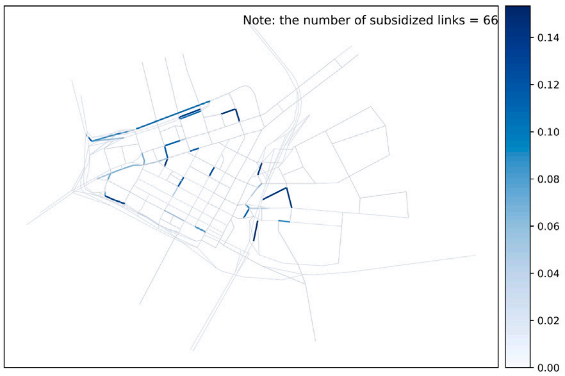
Fig. 15. Spatial distribution of V/C ratios/ratio changes and subsidies of Pittsburgh network in the system travel time minimization scenario ($\gamma = 0$).

Compared with the Sioux Fall network, the Pittsburgh network has more alternative paths whose travel time is slightly larger than the shortest path. As a result, a small amount of subsidies are able to induce ride-hailing vehicles to choose alternative paths. This implies that the subsidy can effectively encourage ride-hailing vehicle deviating when the network has a number of alternative paths with close travel time/costs.

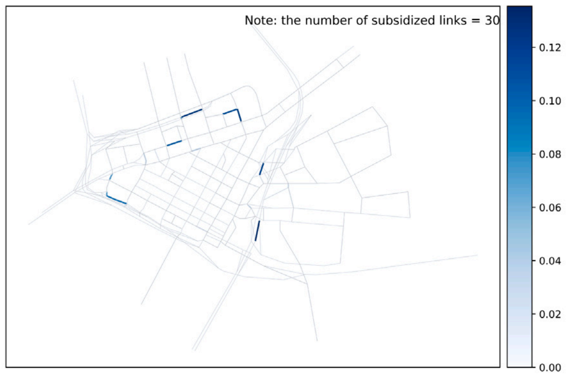
6. Conclusion and future research

This paper studies the impacts of a fleet of vehicles on network performance and proposes a novel strategy for public agencies to price/regulate those fleet vehicles, in exchange for system improvement. In particularly, we use TNCs as an example of fleet vehicles, but the concept works for any fleet as long as it has sufficient market penetrations. First, impacts of TNCs vehicle routing are estimated by which fleet behavior is modeled. Different fleet behaviors, include UE, fleet optimum (FO) and SO, are compared. A fleet-optimal behavior with service constraint (FOSC) is proposed for TNCs to ensure fairness among riders and service quality. With this FOSC fleet behavior, TNC assign routes to riders to minimize fleet cost. To maintain service quality and long-term profits, TNC pay compensations (oftentimes a small fraction of fares) to riders who are willing to deviate to some degree. We also propose a heuristic algorithm to solve the mixed equilibrium of UE and FOSC.

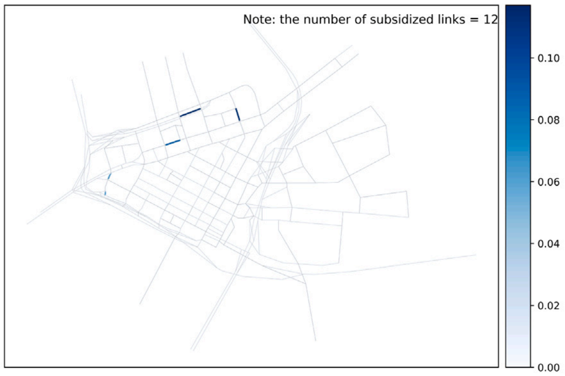
We propose a novel Optimal Ride-hailing Pricing (ORHP) scheme for public agencies as an efficient manner to intervene ride-hailing online platforms. The essential idea of ORHP is to regulate and subsidize TNCs, in exchange for guaranteed network performance improvement. In ORHP, public agencies set the value of a subsidy for each link for any TNC rider using this link. The subsidies are provided to TNCs, not directly to riders. TNC receives subsidies, and determines the best way to provide compensations for each rider who deviates from his/her “shortest” route. TNC’s ultimate goal is to reduce their total cost, including total fleet vehicle travel time and total compensations, subtracted by the total subsidy received from public agencies. We argue that such a system can be hard to game by TNCs or riders. The ORHP is likely less controversial than other pricing schemes, since it calls for voluntary



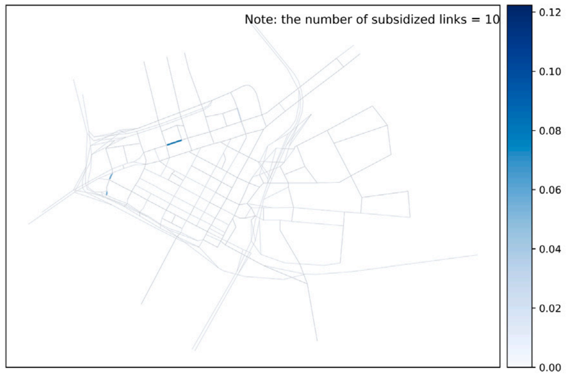
(a) $\gamma = 0.1$



(b) $\gamma = 0.3$

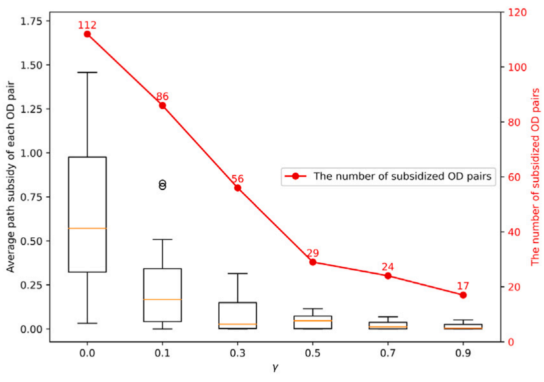


(c) $\gamma = 0.5$

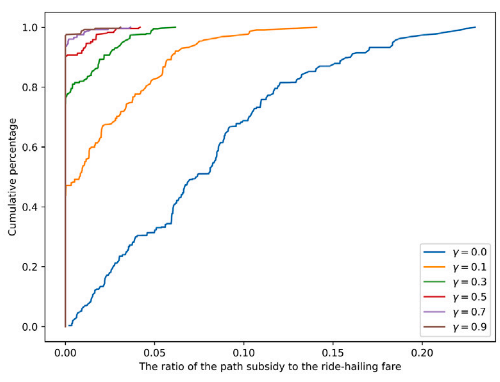


(d) $\gamma = 0.7$

Fig. 16. Link-based subsidy in Pittsburgh network when minimizing the total generalized cost ($\gamma > 0$).



(a) Average path subsidy of each OD pair and the number subsidized OD pairs



(b) Cumulative percentage of the ratio of the path subsidy to the ride-hailing fare

Fig. 17. Path-level subsidies on Pittsburgh network.

participation of travelers, provided with multiple route options. Because it is built into the TNCs' fare system, it is cost effective to implement and may eliminate scamming behavior that may occur in other credit schemes. The ORHP benefits both the public agencies for effectively reducing system cost and TNC for being more profitable.

ORHP is formulated as a bi-level optimization problem. The upper level decides optimal link-based subsidies for ride-hailing vehicles, and the lower level solves the mixed equilibrium of fleet vehicles and driving vehicles given link-based subsidies. A solution algorithm for ORHP is proposed based on sensitivity analysis for multi-class vehicles.

Findings from numerical examples of Sioux Falls network and Pittsburgh network are encouraging:

- If ride-hailing vehicles follow FO or FOSC fleet behavior, the total travel time is lower than UE but higher than ME-SO. The savings of total travel time from UE to ME-FO or ME-FOSC are more pronounced when the penetration of the ride-hailing vehicles is higher.
- Compared with FO, FOSC leads to higher total travel time and total fleet travel time. However, FOSC ensures fairness among ride-hailing riders. Compared with UE fleet behavior, FOSC can effectively reduce fleet travel time through cooperative routing.
- ORHP reduces the total travel time by balancing traffic on both over-congested links and under-congested links. With ORHP, under-congested links close to over-congested links are set subsidies, such that traffic on the over-congested links would take those under-congested links alternatively. Consequently, links with high V/C ratios in ME-FOSC have lower V/C ratios under ORHP, whereas some links with low V/C ratios in ME-FOSC have higher V/C ratios under ORHP.
- With small subsidies implemented onto a small portion of links, ORHP suffices to efficiently reduce the system travel time. This implies that ORHP can effectively incentivize TNCs to improve total system performance at a small cost.
- The proposed solution algorithm based on sensitivity analysis for ORHP is efficient. On Sioux Falls network, the system travel time almost reaches the optimal value in the system travel time minimization scenario (i.e. $\gamma = 0$), and the generalized cost converges quickly in the generalized cost minimization scenario (i.e. $\gamma > 0$). On Pittsburgh network, the cost still declines and converges quickly.

This paper is only a start to explore how fleet vehicles with significant market power can be leveraged to improve transportation system. In order to encapsulate the main effect of traffic routing by pricing ride-hailing vehicles, we simplified the demand model to assume exogenous trips induced by ride-hailing vehicles' idling and pickup flow. It is evident that those idling and pickup flow would also be dependent on the pricing that would in turn alter travelers' demand and drivers' choices. We fully acknowledge this simplification is a limitation of the current work. What is potentially promising is that, if those idling and pickup flow are proportional to service trips, the novel ORHP can effectively improve the system performance with the presence of a small fraction of ride-hailing vehicles. In practice, some regulations or pricing strategies can be imposed to limit those ride-hailing deadweight loss trips. In addition, our findings indicate that under arbitrary mixture of exogenous demand of private and ride-hailing vehicles, ORHP can effectively improve system performance with a small amount of subsidy paid to TNC platforms, which leads to a win–win strategy for all players: TNC, non-TNC private drivers and TNC riders. This finding is based on the simplification of the ride-hailing trip and demand model, but it is reasonably expected that the effect would remain, with some variations, even if the deadweight loss trips are present to some extent. For future study, we suggest taking into consideration passengers' modal choices between private driving and ride-hailing, and TNC choices in matching and pickup strategies (Ban et al., 2019). Modeling such endogenous ride-hailing trips and pickup strategies can further improve model realism and potentially provide a more practical direction to enhance the ORHP framework.

In particular, the fleet dispatch policies in Xu et al. (2021) can be considered. Another immediate extension would be to consider heterogeneous values of time (VOT) of passengers in future research, as riders may value compensations differently depending on their departure time, origins and destinations. TNC can estimate passenger VOTs with historical trip data. A better estimate of VOTs would help a more accurate setting of subsidies and VOT-varying compensations. In addition, additional networks with various demand scenarios would be tested for ORHP.

CRediT authorship contribution statement

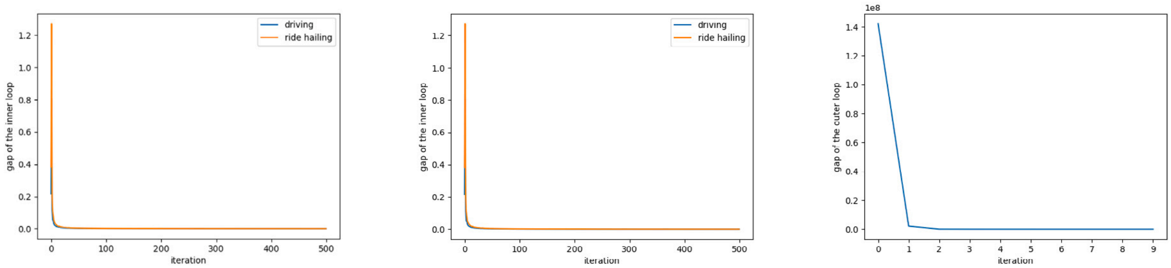
Zemian Ke: Literature review, Model design, Data analytics, Programming and implementation, Analysis and interpretation of results, Manuscript preparation. **Sean Qian:** Study conception, Model design, Data acquisition, Analysis and interpretation of results, Manuscript preparation.

Data availability

Data and source codes are available on https://github.com/Jimmy-Ke404/ORHP.

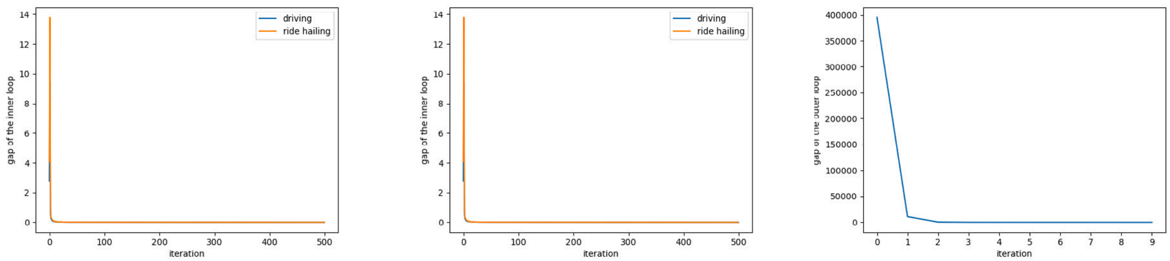
Acknowledgments

This research is supported by a National Science Foundation, United States grant CMMI-1931827. The contents of this article reflect the views of the authors, who are responsible for the facts and accuracy of the information presented herein. The U.S. Government assumes no liability for the contents or use thereof.



(a) Algorithm 1　　(b) Algorithm 2　　(c) Algorithm 3

Fig. 18. Convergence results of Algorithms 1, 2, and 3 for Sioux Falls network with $\gamma = 0.5$.



(a) Algorithm 1　　(b) Algorithm 2　　(c) Algorithm 3

Fig. 19. Convergence results of Algorithms 1, 2, and 3 for Pittsburgh network with $\gamma = 0.5$.

Appendix

In the appendix, we provide the convergence results of Algorithms 1, 2, and 3 with the homogeneous penetration rate and $\gamma = 5$. Note the convergence results with other values of γ are almost identical with $\gamma = 0.5$ (See Figs. 18 and 19).

References

Anderson, D.N., 2014. “Not just a taxi”? For-profit ridesharing, driver strategies, and VMT. Transportation 41, 1099–1117. http://dx.doi.org/10.1007/s11116-014-9531-8.

Babar, Y., Burtch, G., 2017. Examining the impact of ridehailing services on public transit use. SSRN Electron. J. http://dx.doi.org/10.2139/ssrn.3042805.

Ban, X.J., Dessouky, M., Pang, J.S., Fan, R., 2019. A general equilibrium model for transportation systems with e-hailing services and flow congestion. Transp. Res. B 129, 273–304.

Battifarano, M., Qian, S., 2023. The impact of optimized fleets in transportation networks. Transportation Science.

Chen, Y., Florian, M., 1995. The nonlinear bilevel programming problem: Formulations, regularity and optimality conditions. Optimization 32, 193–209.

Chen, Z., Lin, X., Yin, Y., Li, M., 2020. Path controlling of automated vehicles for system optimum on transportation networks with heterogeneous traffic stream. Transp. Res. C 110, 312–329. http://dx.doi.org/10.1016/j.trc.2019.11.017.

Dafermos, S.C., 1972. Traffic assignment problem for multiclass- user transportation networks. Transp. Sci. 6, 73–87, URL: http://dx.doi.org/10.1287/trsc.6.1.73%0AFull.

Delle Site, P., 2021. Pricing of connected and autonomous vehicles in mixed-traffic networks. Transp. Res. Rec. http://dx.doi.org/10.1177/0361198120985850.

DMR, 2021. Uber statistics, user counts and facts (2021). URL: https://expandedramblings.com/index.php/uber-statistics/.

Duchi, J., Hazan, E., Singer, Y., 2011. Adaptive subgradient methods for online learning and stochastic optimization.. J. Mach. Learn. Res. 12.

Florian, M., 1977. A traffic equilibrium model of travel by car and public transit modes. Transp. Sci. 11, 166–179.

Geisberger, R., Luxen, D., Neubauer, S., Sanders, P., Volker, L., 2009. Fast detour computation for ride sharing. arXiv:0907.5269.

Harding, S., Kandlikar, M., Gulati, S., 2016. Taxi apps, regulation, and the market for taxi journeys. Transp. Res. A 88, 15–25. http://dx.doi.org/10.1016/j.tra.2016.03.009.

Harker, P.T., 1988. Multiple equilibrium behaviors on networks.. Transp. Sci. 22, 39–46. http://dx.doi.org/10.1287/trsc.22.1.39.

Harker, P.T., Pang, J.S., 1988. Existence of optimal solutions to mathematical programs with equilibrium constraints. Oper. Res. Lett. 7, 61–64. http://dx.doi.org/10.1016/0167-6377(88)90066-1, URL: https://www.sciencedirect.com/science/article/pii/0167637788900661.

He, F., Yin, Y., Shirmohammadi, N., Nie, Y.M., 2013. Tradable credit schemes on networks with mixed equilibrium behaviors. Transp. Res. B 57, 47–65. http://dx.doi.org/10.1016/j.trb.2013.08.016.

Imamoto, A., Tang, B., 2008. Optimal piecewise linear approximation of convex functions. In: Proceedings of the World Congress on Engineering and Computer Science 2008.

Jahn, O., Möhring, R.H., Schulz, A.S., Stier-Moses, N.E., 2005. System-optimal routing of traffic flows with user constraints in networks with congestion. Oper. Res. 53, 600–616. http://dx.doi.org/10.1287/opre.1040.0197.

Jin, S.T., Kong, H., Wu, R., Sui, D.Z., 2018. Ridesourcing, the sharing economy, and the future of cities. Cities 76, 96–104. http://dx.doi.org/10.1016/j.cities.2018.01.012.

Liu, L.N., McDonald, J.F., 1999. Economic efficiency of second-best congestion pricing schemes in urban highway systems. Transp. Res. B 33, 157–188.
Lo, H.K., Luo, X.W., Siu, B.W.Y., 2006. Degradable transport network: Travel time budget of travelers with heterogeneous risk aversion. Transp. Res. B 40, 792–806. http://dx.doi.org/10.1016/j.trb.2005.10.003, URL: https://www.sciencedirect.com/science/article/pii/S0191261505001177.
Marcotte, P., Zhu, D.L., 1996. Exact and inexact penalty methods for the generalized bilevel programming problem. Math. Program. 74, 141–157.
Marquet, O., 2020. Spatial distribution of ride-hailing trip demand and its association with walkability and neighborhood characteristics. Cities 106, 102926.
May, A.D., Milne, D.S., 2000. Effects of alternative road pricing systems on network performance. Transp. Res. A 34, 407–436.
May, A.D., Milne, D.S., Shepherd, S.P., Sumalee, A., 2002. Specification of optimal cordon pricing locations and charges. Transp. Res. Rec. 1812, 60–68.
Meng, Q., Yang, H., Bell, M.G.H., 2001. An equivalent continuously differentiable model and a locally convergent algorithm for the continuous network design problem. Transp. Res. B 35, 83–105.
Mun, S.i., Konishi, K.j., Yoshikawa, K., 2003. Optimal cordon pricing. J. Urban Econ. 54, 21–38.
Nagurney, A., 1998. Network Economics: A Variational Inequality Approach, vol. 10. Springer Science & Business Media.
Nagurney, A., Dong, J., 2002. A multiclass, multicriteria traffic network equilibrium model with elastic demand. Transp. Res. B 36, 445–469. http://dx.doi.org/10.1016/S0191-2615(01)00013-3, URL: https://www.sciencedirect.com/science/article/pii/S0191261501000133.
Newell, G.F., 1987. The morning commute for nonidentical travelers. Transp. Sci. 21, 74–88.
Ni, W., Cassidy, M., 2020. City-wide traffic control: Modeling impacts of cordon queues. Transp. Res. C 113, 164–175.
Nie, Y.M., 2017. How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China. Transp. Res. C 79, 242–256. http://dx.doi.org/10.1016/j.trc.2017.03.017.
Pham, K.H., Ipeirotis, P., Sundararajan, A., 2020. Ridesharing and the use of public transportation. In: 40th International Conference on Information Systems. ICIS 2019, pp. 1–11.
Pigou, A.C., 1920. The Economics of Welfare. MacMillan, London.
Qian, Z.S., Xiao, F.E., Zhang, H.M., 2012. Managing morning commute traffic with parking. Transp. Res. B 46, 894–916.
Shao, H., Lam, W.H.K., Tam, M.L., 2006. A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand. Netw. Spat. Econ. 6, 173–204. http://dx.doi.org/10.1007/s11067-006-9279-6.
Sheffi, Y., Powell, W.B., 1982. An algorithm for the equilibrium assignment problem with random link times. Networks 12, 191–207.
Tobin, R.L., Friesz, T.L., 1988. Sensitivity analysis for equilibrium network flow. Transp. Sci. 22, 242–250. http://dx.doi.org/10.1287/trsc.22.4.242.
Van Vuren, T., Van Vliet, D., Smith, M.J., 1990. Combined equalibrium in a network with partial route guidance. In: Traffic Control Methods. Proceedings of the Fifth NG Foundation Conference. Santa Barbara, California.
Van Vuren, T., Watling, D., 1991. A multiple user class assignment model for route guidance. Transp. Res. Rec. 22–32, National Academy of Sciences.
Verhoef, E., Nijkamp, P., Rietveld, P., 1996. Second-best congestion pricing: The case of an untolled alternative. J. Urban Econ. 40, 279–302.
Vickrey, W.S., 1963. Pricing in urban and suburban transport. Am. Econ. Rev. 53, 452–465.
Wallsten, S., 2015. The competitive effects of uber_preview. Technol. Policy Inst. 1–22, URL: www.researchgate.net/publication/279514652_The_Competitive_Effects_of_the_Sharing_Econo.
Wang, J., Peeta, S., He, X., 2019. Multiclass traffic assignment model for mixed traffic flow of human-driven vehicles and connected and autonomous vehicles. Transp. Res. B 126, 139–168. http://dx.doi.org/10.1016/j.trb.2019.05.022.
Wang, H., Yang, H., 2019. Ridesourcing systems: A framework and review. Transp. Res. B 129, 122–155. http://dx.doi.org/10.1016/j.trb.2019.07.009.
Xiao, F., Qian, Z.S., Zhang, H.M., 2013. Managing bottleneck congestion with tradable credits. Transp. Res. B 56, 1–14.
Xu, T., Barman, S., Levin, M.W., Chen, R., Li, T., 2022. Integrating public transit signal priority into max-pressure signal control: Methodology and simulation study on a downtown network. Transp. Res. C 138, 103614.
Xu, T., Cieniawski, M., Levin, M.W., 2023. FMS-dispatch: A fast maximum stability dispatch policy for shared autonomous vehicles including exiting passengers under stochastic travel demand. Transp. A Transp. Sci. 1–39.
Xu, T., Levin, M.W., Cieniawski, M., 2021. A zone-based dynamic queueing model and maximum-stability dispatch policy for shared autonomous vehicles. In: 2021 IEEE International Intelligent Transportation Systems Conference. ITSC, IEEE, pp. 3827–3832.
Yang, H., 1997. Sensitivity analysis for the elastic-demand network equilibrium problem with applications. Transp. Res. B 31, 55–70. http://dx.doi.org/10.1016/S0191-2615(96)00015-X.
Yang, H., 1998. Multiple equilibrium behaviors and advanced traveler information systems with endogenous market penetration. Transp. Res. B 32, 205–218. http://dx.doi.org/10.1016/S0191-2615(97)00025-8.
Yang, H., Huang, H.J., 2004. The multi-class, multi-criteria traffic network equilibrium and systems optimum problem. Transp. Res. B 38, 1–15. http://dx.doi.org/10.1016/S0191-2615(02)00074-7.
Yang, H., Huang, H.J., 2005. Mathematical and Economic Theory of Road Pricing. pp. 68–70.
Yang, H., Zhang, X., 2008. Existence of anonymous link tolls for system optimum on networks with mixed equilibrium behaviors. Transp. Res. B 42, 99–112. http://dx.doi.org/10.1016/j.trb.2007.07.001.
Yang, H., Zhang, X., Meng, Q., 2007. Stackelberg games and multiple equilibrium behaviors on networks. Transp. Res. B 41, 841–861. http://dx.doi.org/10.1016/j.trb.2007.03.002.
Yu, B., Ma, Y., Xue, M., Tang, B., Wang, B., Yan, J., Wei, Y.M., 2017. Environmental benefits from ridesharing: A case of Beijing. Appl. Energy 191, 141–152. http://dx.doi.org/10.1016/j.apenergy.2017.01.052.
Zhang, K., Nie, Y.M., 2018. Mitigating the impact of selfish routing: An optimal-ratio control scheme (ORCS) inspired by autonomous driving. Transp. Res. C 87, 75–90. http://dx.doi.org/10.1016/j.trc.2017.12.011.
Zhang, Y., Zhang, Y., 2018. Exploring the relationship between ridesharing and public transit use in the united states. Int. J. Environ. Res. Public Health 15, http://dx.doi.org/10.3390/ijerph15081763.