Transportation Research Part B 169 (2023) 29–52

Contents lists available at ScienceDirect

Transportation Research Part B

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

On the utilization of dedicated bus lanes for pooled ride-hailing services

Lynn Fayed, Gustav Nilsson, Nikolas Geroliminis *

Urban Transport Systems Laboratory (LUTS) École Polytechnique Fédérale de Lausanne (EPFL), Switzerland

ARTICLE INFO

Keywords:
Multi-modal networks
Network delays
Public transportation
Regulations
Ride-splitting services
Space allocation

ABSTRACT

Ride-sourcing platforms, among other solution services, offer convenience and flexibility when it comes to pick-up/drop-off time and location. Similarly, ride-splitting renders itself as an extension of ride-sourcing where platform users agree to share their rides in return for a reduced fare yet possibly a longer travel time. Despite the numerous advantages that sharing introduced to the platform operator by reducing the fleet size necessary to serve demand levels, e-hailing is still overall negatively impacting traffic performance in urban spaces. This is partially due to the current tendency of users to favor solo over shared rides. This paper aims to use aggregate traffic flow models to put forward a network space redistribution policy that has the potential to reduce total delays of the different mode users in the network. Accordingly, we investigate the solo-pool demand split that minimizes the total Passenger Hours Traveled for all network commuters in the event where shared rides are allowed to use underutilized bus lanes. As a result, the choice to share is associated with an inevitable additional detour distance but with a lower-than-expected trip time compared to standard scenarios where the whole fleet utilizes the same network space. In this paper, we present an analytical macroscopic modeling approach to evaluate equilibrium solutions between network supply and multi-modal demand. By resorting to a numerical example of the model, we show that the optimal strategy that minimizes delays for multi-modal transport users occurs when only a fraction of the pooling vehicles uses the dedicated bus lanes. This fraction is dependent on the distribution of space and the value of demand for private vehicles, buses, and ride-hailing services.

1. Introduction

The rapid growth of ride-sourcing systems evinces an increasing interest in personalized trips where users input their origins and destinations, and service providers assign them to convenient rides. These services, also interchangeably called e-hailing, are characterized by a single platform connecting both riders and drivers. It collects the totality of requested trip details for which it provides a feasible driver match from the operating fleet of available online vehicles. Another type of service fundamentally similar to ride-sourcing is referred to as ride-splitting. In the latter, passengers grant their approval *a priori* to share their rides with other users of the system in return for an upfront decrease in fares to compensate for any inconvenience that sharing incurs. Ride-splitting has a myriad of advantages since it allows the same fleet of vehicles to accommodate a larger number of arriving requests while reducing the total travel distance (Ma et al., 2015).

Despite their success, the ubiquitous character of these platforms raises particular concerns because they provide the same service structure in heterogeneous urban areas with diverse spatial configurations, various mode availabilities, mismatched public transport

* Corresponding author.
E-mail addresses: lynn.fayed@epfl.ch (L. Fayed), gustav.nilsson@epfl.ch (G. Nilsson), nikolas.geroliminis@epfl.ch (N. Geroliminis).

https://doi.org/10.1016/j.trb.2023.01.005
Received 30 May 2022; Received in revised form 24 January 2023; Accepted 31 January 2023
Available online 6 February 2023

service levels, and different traffic intensities. These factors justify the continuous effort to (i) set forth a framework that models the operation of ride-sourcing services, and (ii) unravel their influences on traffic externalities, Vehicle Kilometers Traveled (VKT), modal substitution and complimentary, long-term car ownership, and lastly social welfare of drivers and riders (Tirachini, 2020; Henao and Marshall, 2017; Erhardt et al., 2019). Accordingly, a comprehensive understanding of the interactions between the previous factors helps decision-makers to adopt informed policies and actions to regulate the operation of e-hailing platforms. The type and extent of regulations are largely dependent on the expected outcomes that the government aims to achieve. The multi-purpose intervention usually targets the fleet of vehicles by capping the number of registered drivers (Yu et al., 2020) or the cumulative time of driving empty (Schaller, 2018). Beojone and Geroliminis (2021) showed that prompting Transport Network Companies (TNCs) to enact parking management strategies is efficient in reducing empty VKT and subsequently mitigating congestion, as demonstrated by a simulation study from a Chinese megacity.

To our knowledge, however, the assessment of incentive-based regulatory approaches rather than enforcement strategies is still deficient in the literature. Moreover, the majority of the previous models developed to study the efficiency of any regulation policy only accounts for stakeholders in direct affiliation with the platform without consideration of other transport network users. A review of available work on pooling in the context of e-hailing and carsharing validates the potential of trip sharing in mitigating congestion and reducing VKT (Shaheen and Cohen, 2019; Tirachini, 2020). Accordingly, one such regulatory approach is to allow pool ride-hailing vehicles to travel in dedicated, high-occupancy lanes. Practically, this policy is mainly characterized by its ease of implementation. For example, it is commonly known based on previous experience that a High-Occupancy Vehicle (HOV) lane is not easily enforceable, while ride-sharing for modes that use online platforms (e.g. like Uber, Didi, Lyft) is more straightforwardly enforced. Nevertheless, if such a policy is implemented without proper knowledge of the interactions between vehicles, passengers, and infrastructure, it can lead to solutions that deteriorate the overall performance of the system.

Accordingly, the contributions of this paper are to present a new occupancy-dependent and modal-dependent space allocation strategy for multi-modal transportation networks where each mode of transport exploits the network space differently: private vehicles and solo ride-hailing users utilize the vehicle network, while pooled vehicles and buses utilize the bus network. We therefore provide a macroscopic modeling framework of the proposed allocation strategy using an aggregate formulation of the network supply based on the theory of Macroscopic Fundamental Diagrams (MFDs), and a high-level modeling of ride-hailing, bus, and private vehicle demand. To analyze this strategy, we design an analytical framework to demonstrate the solution existence, and to evaluate the change in the different network accumulations as function of the demand for solo and pool. This equilibrium is assessed according to the allocation strategy. The model we present is versatile because it can be easily exploited to assess the influence of different spatial split and modal distribution on user delays. Our intention is to avoid the integration of tedious efforts related to traffic assignment at the network level with multi-modal transport and instead to make use of an approximate aggregated network model (the MFD) that has been utilized in the literature mainly for car traffic, but also for bus–car interactions.

The remainder of this paper is organized as follows. The next section, Section 2, provides an overview of the existing literature tackling different aspects of e-hailing operation, performance, and external influence. Section 3 elaborates on the transportation mode-based network allocation model we propose in this work, and demonstrates the main analytical findings. In Section 4, we investigate the effectiveness of the proposed model by closely studying a numerical example. We additionally look into the results for one variant of this model where only a fraction of the pooled vehicles is allowed in the bus lanes. Finally, Section 5 concludes with the results' summary and advances future research considerations.

2. Literature review

Efforts are converging in the research community to properly model, understand, and possibly improve vehicle-passenger matching operations, spatial and temporal fare differentiation, demand forecasting, and idle vehicle rebalancing in ride-sourcing systems. By analogy with aggregate equilibrium in traditional taxi services, some characteristics of the ride-hailing supply markets are straightforwardly inferred. However, probably the most obvious distinction to point out is a direct outcome of the existence of complete information on the location of drivers and users. The platform dispatches an available driver to perform a pick-up operation based on prior knowledge of the user's origin. Despite its apparent usefulness, the dispatch distance is a source of inefficiency in the event of a supply shortage as pointed out by Castillo et al. (2017). This behavior is otherwise not observable in street-hailing markets where supply is a monotonic function of the waiting time. Xu et al. (2020) reproduced the backward bending supply curve and showed that by limiting the search area for an empty vehicle, it is possible to lessen the bend.

The formulation of bilateral meeting functions or the behavioral modeling of drivers and passengers are also some of the fundamental aspects of the taxi/e-hailing markets. Researchers on one hand assessed the stationary equilibrium state for street-hailing taxi services (Yang et al., 2010), simultaneous e-hailing and street-hailing services (He and Shen, 2015), or exclusive e-hailing services (Zha et al., 2016; Zhang and Nie, 2021). On the other hand, they studied e-hailing markets under dynamic non-equilibrium settings for different pricing schemes (Nourinejad and Ramezani, 2020). Pricing itself is also extensively addressed in the literature as a course of action to eliminate market inefficiencies. Cachon et al. (2017) examined the importance of surge pricing in ensuring an optimal service in terms of capacity utilization and demand satisfaction. Zha et al. (2018a) showed that passenger welfare is negatively impacted by abrupt price increase, and the attenuation of this effect is possible through a platform commission cap. Contrary to the previous three studies which focused on temporal price variation, Zha et al. (2018b) and Bimpikis et al. (2019) assessed spatial price discrimination to balance the heterogeneous distribution of demand.

Another research line within this scope focuses on the precise quantification of the ambiguous effect of ride-sourcing on VKT and congestion, mainly because of the numerous intertwining factors governing this relationship (Rayle et al., 2016). Efficient real-time

management of the vehicle fleet entails an improved utilization of available capacity and a lower vacant to occupied distance ratio compared to conventional taxi services (Cramer and Krueger, 2016). Nevertheless, ride-sourcing is sometimes validly regarded as a transit substitute, particularly in areas where public transportation services are deficient. Using a simulation-based approach and survey data from Santiago, Tirachini and Gómez-Lobo (2019) demonstrated that in scenarios where e-hailing replaces transit, an increase in VKT is inevitable if no pooling is involved. This increase does not necessarily hold when e-hailing rather complements transit by reinforcing its usage (Hall et al., 2018). Using static multi-modal choice models, Ke et al. (2021b) and Zhu et al. (2020) provided insights on how to enforce complementarity between public transport from one side and ride-sourcing/ride-splitting from the other. Similarly, Feng et al. (2022) proposed an algorithm to improve the coordination between ride-hailing and public transit, and therefore to achieve better utilization of transit modes, and higher platform revenues.

From a modeling point of view, understanding the impact of ride-sourcing or ride-splitting on congestion requires the revisit of traffic assignment or aggregate traffic flow models. Xu et al. (2019) proposed a framework to incorporate the behavior of idling and dispatching vehicles into current assignment models and investigated network equilibrium under different spatial matching strategies. Vignon et al. (2021) proposed a model for ride-splitting markets in the presence of background congestion to evaluate platform profit maximization for different congestion levels. Similarly, Mo et al. (2022) captured traffic externalities in their modeling of ride-hailing markets with a mixed fleet of autonomous and human-driven vehicles. Ke et al. (2020b) used a linear speed-density relationship to compute the maximum achievable ride-sourcing and ride-splitting service rates and showed that under specific conditions, pooling decreases the travel time in the network. Their model accounts for private and ride-hailing vehicles without consideration of mass transit or the potential utilization of bus lanes for pooled trips. By assuming that the origins and destinations of passengers are within walkable distance, they did not factor for the additional detour time which, along with the price discount, are the major determinants of the willingness to share (Alonso Gonzalez et al., 2020; Lo and Morseman, 2018). The detour distance itself is also critical in evaluating the pooling potential to reduce VKT. Empirical evidence exhibits an obvious trend in terms of the relation between passenger detour and pool demand level where the former decreases with the latter (Ke et al., 2021a). Through the development of a detour-constrained analytical model, Daganzo et al. (2020) validated the observation that ride-splitting becomes more beneficial when trip requests increase. The sole consideration of the detour distance, however, is not sufficient in formulating choice models between solo and pooled rides.

To incentivize ride-splitting and ameliorate sharing opportunities, researchers focused on assessing price regulation rather than pooled vehicle spatial prioritization. Generally speaking though, dedicated bus lanes or special Autonomous Vehicles (AV) lanes are expected to improve network delays or social costs if space allocation policies are carefully chosen whether on the link level for AVs (Lamotte et al., 2017) or for buses (Eichler and Daganzo, 2006), or the network level (Mesbah et al., 2011). In the former case, Bus Lanes with Intermittent Priority (BLIP) are useful to decrease the delays of highly occupied buses but this strategy is time and space dependent. Moreover, it creates a platoon of vehicles moving in the bus lanes in between bus arrivals while disabling the possibility of simultaneous usage by buses and vehicles. This strategy is therefore not able to capture the interference of vehicles in bus flow. In the latter case, the existence of an urban-scale MFD (Geroliminis and Daganzo, 2008) or a three-dimensional MFD (Geroliminis et al., 2014) provides increased convenience and flexibility when solving optimal multi-modal space allocation problems as it circumvents the need for a lower-level traffic assignment models. Zheng and Geroliminis (2013) for instance used a macroscopic approach to find a static and dynamic optimal allocation policy between buses and private vehicles. They showed that a dynamic distribution of available space brings off a lower cumulative delay because of improved utilization of network capacities. Building on this work, Tsitsokas et al. (2021) developed an optimization framework based on a large neighborhood search algorithm to find the optimal bus lane allocation scheme that minimizes the total travel time for all passengers in a network. With the advent of ride-splitting, it becomes hence particularly interesting to observe how such policies are adapted to integrate new modes into optimal allocation strategies.

3. Model formulation

The following section elaborates on the model that enables us to examine the redistribution of ride-splitting demand in an urban space in the existence of other modes of transport. We start by presenting a high-level overview of the model in Section 3.1. Subsequently, in Section 3.2, we delve into the dynamics of the vehicle and bus networks, and we derive some theoretical properties for each network model. Under the assumption that the demand for each transportation mode is fixed, it is possible to compute the number of private vehicles, buses, and ride-hailing drivers that are present in the network. In Section 3.3, we present our models for the computation of the vehicle and bus network accumulation together with their analytical properties. In Section 3.4 we exhibit the optimization problem for obtaining a system optimum value for delays, and we furthermore discuss the properties of this optimum. At the end of this modeling section, we suggest how to obtain this system optimum when ride-hailing vehicles are not allowed to use the bus network. This modal allocation will serve as a benchmark scenario for the numerical studies displayed in the section after.

3.1. Network model

In the network under consideration, travelers perform their trips by one of the set $\mathcal{M}$ of available options: private vehicles pv, buses b, or ride-hailing services rh such that $\mathcal{M} = \{pv, b, rh\}$. Commuters who opt for the ride-hailing alternative pick out from the choice set $\mathcal{M}_{rh}$ to either travel solo or to pool their trips with other users of the service. We refer to the latter two trips as s and p respectively such that $\mathcal{M}_{rh} = \{s, p\}$. In the first case, passengers do not share their rides with any other passenger all throughout

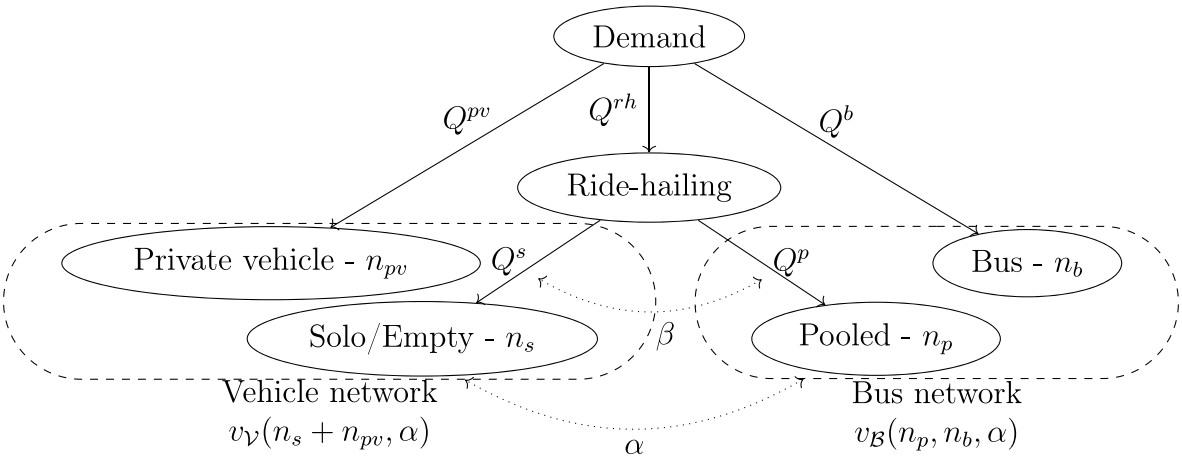


Fig. 1. A schematic sketch of the suggested network configuration and the modal-dependent space allocation that we assess in this paper.

their trips. In the second case, passengers share their rides with exactly one other passenger for at least a subpart of their total trips. Throughout the paper, we assume a spatial distribution between modes such that buses exclusively utilize dedicated bus lanes to transport passengers, and private vehicles share the remaining of the network. For this purpose, we let $\alpha \in [0, 1]$ denote the spatial fraction of the whole network allocated to private vehicles. Thereafter, the spatial fraction devoted to dedicated bus lanes is $\bar{\alpha} = 1 - \alpha$. When $\alpha = 0$, the full network space is devoted to buses whereas when $\alpha = 1$, the network becomes an exclusive vehicle network. We note that α is an intrinsic property of the network infrastructure, and changing its value is often very costly. With respect to the fleet of ride-hailing vehicles, in the event where drivers are matched to a pooled trip, their entire rides are performed in the bus network. Elsewise, all idle, dispatched, and solo-ride drivers travel in the vehicle network in conjunction with private vehicles. For the rest of this study, we will refer to the vehicle network as $\mathcal{V}$, and to the bus network as $\mathcal{B}$.

Let Q^m be the travel demand for mode $m \in \mathcal{M}$ expressed in passengers per hour, its value remains unchanged in this study. The split of ride-hailing demand between solo and pool however is assumed to be variable. Therefore, we let $\beta \in [0, 1]$ denote the fraction of Q^{rh} that selects a solo trip. In particular, we investigate in this work how the value of β affects the total delay for all network commuters. Mainly, when ride-hailing users choose to pool such that $Q^p = (1 - \beta)Q^{rh}$ is the pooling demand, this portion of ride-hailing trips interferes with the bus speed, causing it to decrease. If instead ride-hailing users choose to travel solo such that $Q^s = \beta Q^{rh}$ is the demand for solo trips, they affect the speed in the vehicle network. Another element that has a significant influence on speed is the factor α dictating the spatial size of the bus and vehicle networks. The larger the value of α, the higher the capacity of the vehicle network, and the lower that of the bus network. Within this context, we additionally investigate in this study how the optimal demand split between solo and pool trips varies with α. The ultimate purpose is to assess how the two variables α and β affect the quality of the mobility in a network in terms of total user delays. Fig. 1 provides a summary of our model, including the different modes of transport we consider and how they differently load each of the networks. Additionally, it exhibits the various variables we investigate in this work.

To evaluate network delays, a proper estimation of the speeds in the vehicle and bus networks is required. We denote the speed in these networks by $v_{\mathcal{V}}$ and $v_{\mathcal{B}}$ respectively. According to the model in Fig. 1, $v_{\mathcal{V}}$ is a function of the accumulation of private vehicles n_{pv} and that of empty/solo trip e-hailing vehicles n_s. Conversely, the speed in the bus network is dependent on the number of pooled drivers that we denote by n_p and the fleet of buses that we denote by n_b. We point out here that, because buses have to repeatedly dwell at bus stops, their influence on speed is not equivalent to that of the pooled vehicles. For this reason, we split both vehicle quantities instead of aggregating them to accurately model speed in the bus network. Moreover, as we previously described, both speed functions are linked to the fraction α which dictates the space division between the two networks. For us to examine the values of travel times for each individual mode, we resort first to aggregate traffic flow models to define the relationship between speed and accumulation.

A summary of all the different variables used in this work, along with their descriptions is displayed in Table 1.

3.2. Traffic dynamics

In the model we presented, the space-mean speed is a crucial element to evaluate the average trip time for all mode users. Let n be the vehicle accumulation in the full network, and v its speed. We know that as the accumulation n increases, the network becomes more and more congested, and hence its speed decreases. If Q is the total effective trip demand and $\bar{l}$ is the average trip length, then the vehicle production P expressed in vehicle kilometers per unit time is defined by $P = nv$. Hence, when the system is at steady-state, it must hold that $P = Q\bar{l}$. Next, we formalize the properties of the network speed function.

Table 1
List of notations.

Variable name	Description
$\mathcal{M}$	Set of transportation modes where $\mathcal{M} = \{pv, rh, b\}$
$\mathcal{M}_{rh}$	Set of ride-hailing options where $\mathcal{M}_{rh} = \{s, p\}$
n	Total vehicle accumulation
P	Total production in the network
v	Network speed
α	Fraction of space allocated to the vehicle network $\mathcal{V}$
$n_{\mathcal{V}}$, $n_{\mathcal{B}}$	Accumulation in the vehicle and bus network $\mathcal{V}$ and $\mathcal{B}$ respectively
n_i	Number of vehicles of category $i \in \mathcal{M} \cup \mathcal{M}_{rh}$
$P_{\mathcal{V}}$, $P_{\mathcal{B}}$	Production in the vehicle and bus network $\mathcal{V}$ and $\mathcal{B}$ respectively
P_p, P_b	Production of pool vehicles and buses respectively in the bus network $\mathcal{B}$
$v_{\mathcal{V}}$, $v_{\mathcal{B}}$	Speed in the vehicle and bus network $\mathcal{V}$ and $\mathcal{B}$ respectively
v_p, v_b	Speed of pool vehicles and buses respectively in the bus network $\mathcal{B}$
k	Function of influence of buses on speed in the bus network $\mathcal{B}$
Q^i	Demand for mode $i \in \mathcal{M} \cup \mathcal{M}_{rh}$
β	Fraction of ride-hailing requests opting for solo trips
N	Ride-hailing fleet size where $N = n_{rh}$
I	Number of idle ride-hailing vehicles in the network
d	Dispatched distance for ride-hailing vehicles
w	Waiting time of ride-hailing requests
τ	Ride-hailing platform target waiting time
T	Total driver pooled trip length
$\bar{o}_p$, $\bar{o}_b$	Average pool vehicles and bus occupancy respectively
$\bar{v}_b$	Design bus speed
$\bar{l}$, $\bar{l}_b$	Average vehicle and bus trip length respectively
Δl_d, Δl_p	Driver and passenger pooled trip detour distance respectively
Δt_d	Driver pooled trip detour time
μ	Monetary value of time
κ, κ_p	Logit scale and nest scale parameter respectively
γ	Fraction of Q^p utilizing the bus lane
F_s	Solo ride fare
ϕ	Pool trip discount factor
t^s, t^p	Solo and pool passenger trip length respectively
$t^s_{\mathcal{V}}$, $t^p_{\mathcal{B}}$	Pool passenger trip length in the vehicle and bus network $\mathcal{V}$ and $\mathcal{B}$ respectively
$\bar{t}_d$	Average bus dwell time at stops
$\bar{s}$	Average spacing between bus stops

Definition 1 (*Network Speed Function*). Let $n_{cr} > 0$, $n_{jam} > 0$ and $v_{max} > 0$ be the critical accumulation, jam accumulation, and maximum speed in the network respectively such that $n_{cr} < n_{jam}$, then a network speed function $v : [0, n_{jam}] \to [0, v_{max}]$ has the following properties:

(i) $v(0) = v_{max}$
(ii) $v(n_{jam}) = 0$
(iii) $v' \leq 0$ for all $n \in [0, n_{jam}]$
(iv) $v(n_{cr}) + n_{cr}v'(n_{cr}) = 0$
(v) $2v' + nv'' < 0$ for all $n \in [0, n_{jam}]$

Part (i) of the definition ensures that the network speed reaches its maximum when there are no vehicles in the network. Part (ii) indicates that when the network is jammed, the speed is zero. Part (iii) states that the network speed is non-increasing with its accumulation, i.e., the speed can never increase when more vehicles join the network. Finally, parts (iv) and (v) guarantee that the production function $P = nv$ is strictly concave. Although there are decreasing speed functions that yield a non-concave MFD, the concavity assumption is valid for most of the MFD functions used in the literature (Cheng et al., 2021), especially in the uncongested and lightly congested regime. This concave production function has a unique maximum, and this maximum is achieved when the accumulation in the network attains its critical point n_{cr}. It follows that when $n < n_{cr}$, the network is in the *uncongested* regime and $\frac{\partial P}{\partial n} = v + nv' > 0$. On the contrary, when $n > n_{cr}$, the network enters the *congested* regime where $v + nv'$ becomes negative.

3.2.1. Vehicle network

Under the proposed space allocation strategy, the network infrastructure is segmented into two distinct subnetworks, each having its own production function. Consequently, having a well-defined production MFD function for the full network makes it possible to derive that of the vehicle network using the spatial fractional split factor α. This corresponds to a rescaling of the full network MFD such that if $P_{\mathcal{V}}$ is the production function for the vehicle network only, then $\alpha P(n) = P_{\mathcal{V}}(\alpha n)$. This results in the following speed relation (Geroliminis et al., 2013; Ni and Cassidy, 2019)

$$v(n) = v_{\mathcal{V}}(\alpha n)\,. \tag{1}$$

Accordingly, we obtain a relationship between speed and accumulation that is exclusive to the vehicle network, irrespective of the state of the bus network. It follows that the production $P_{\mathcal{V}}$ and speed $v_{\mathcal{V}}$ in the vehicle network have the same properties introduced in Definition 1 except for replacing n_{cr} and n_{jam} by αn_{cr} and αn_{jam} respectively. As a result, if $n_{\mathcal{V}}$ is the accumulation in the vehicle network according to our model, then the speed in the vehicle network is $v_{\mathcal{V}}(n_{\mathcal{V}}, \alpha)$. To simplify the notation, unless the dependency of α is explicitly needed, we will write the vehicle network speed function as $v_{\mathcal{V}}(n_{\mathcal{V}})$ and let $v'_{\mathcal{V}}(n_{\mathcal{V}}) = \frac{\partial}{\partial n_{\mathcal{V}}} v_{\mathcal{V}}(n_{\mathcal{V}}, \alpha)$.

Lemma 1. *If the overall network speed function satisfies Definition 1, then the production function for the vehicle network $P_{\mathcal{V}}(n_{\mathcal{V}}) = n_{\mathcal{V}} v_{\mathcal{V}}(n_{\mathcal{V}})$ is concave with respect to $n_{\mathcal{V}}$.*

Proof. Finding the first and second derivative with respect to n of (1), we obtain that $v'(n) = \alpha v'_{\mathcal{V}}(\alpha n)$ and $v''(n) = \alpha^2 v''_{\mathcal{V}}(\alpha n)$. It follows from part (v) of Definition 1 that

$$0 > 2v'(n) + nv''(n) = 2\alpha v'_{\mathcal{V}}(\alpha n) + n\alpha^2 v''_{\mathcal{V}}(\alpha n) = \alpha\left(2v'_{\mathcal{V}}(\alpha n) + n\alpha v''_{\mathcal{V}}(\alpha n)\right). \tag{2}$$

Replacing αn by $n_{\mathcal{V}}$, we obtain that $2v'_{\mathcal{V}}(n_{\mathcal{V}}) + n_{\mathcal{V}} v''_{\mathcal{V}}(n_{\mathcal{V}}) < 0$ for all $n_{\mathcal{V}} \in [0, \alpha n_{jam}]$ since $\alpha > 0$. □

3.2.2. Bus network

In a similar manner, the production in the bus network is related to the full network production by knowing that $\bar{\alpha} P(n) = P_{\mathcal{B}}(\bar{\alpha} n)$. In contrast however to the vehicle network, simply adding the bus and vehicle accumulation does not yield a reliable input for the estimation of the running speed from the full network production function. This is because an adjustment is required to account for the frequent stops of buses at stations and the subsequent hindering of vehicle movements. This effect is observed when translating the bus network MFD to the three-dimensional space where the accumulation of buses and vehicles are separated (Geroliminis et al., 2014; Loder et al., 2017; Fu et al., 2020). This is because this function returns the values of production with respect to the accumulation of buses n_b and the accumulation of pooled vehicles n_p and not as the sum of the two accumulations $n_p + n_b$. In the baseline scenario where $n_b = 0$, we know that Definition 1 also applies to the speed in the bus network $v_{\mathcal{B}}$ except that the jam accumulation is achieved for a value equal to $\bar{\alpha} n_{jam}$. However, as the number of buses increases, the MFD definition is expanded to englobe the distinctive influence that buses have on speed. Hence, the functional form of production in the bus network is $P_{\mathcal{B}}(n_p, n_b, \alpha) = n_p v_{\mathcal{B}}(n_p, n_b, \alpha) + n_p v_p(n_p, n_b, \alpha)$ where v_p is the operating speed of buses in network $\mathcal{B}$. Just as for the vehicle network we will, for simplicity of the notation, drop the dependency of α unless it is explicitly needed, such that the speed in the bus network will be denoted as $v_{\mathcal{B}}(n_p, n_b)$. Given this expression of production in the bus network, we define a bus network speed function as follows:

Definition 2 (*Bus Network Speed Function*). A bus network speed function $v_{\mathcal{B}}$ where $v_{\mathcal{B}}(n_p, n_b, \alpha) : \{(n_p, n_b) \mid n_p \geq 0, n_b \geq 0, n_b + n_p \leq \bar{\alpha} n_{jam}\} \times [0,1] \to [0, v_{max}]$ is given by $v_{\mathcal{B}}(n_p, n_b, \alpha) = v\left(\frac{1}{\bar{\alpha}}(n_p + n_b)\right) k(n_b)$ where $k : \mathbb{R}_{\geq 0} \to (0,1]$ is a decreasing function and v is the full network speed function.

The following three properties of the bus network are immediate consequences of Definition 2:

(i) if $n_p^*(n_b) = \arg\max_{n_p}(P_{\mathcal{B}})$, then $\frac{dn_p^*}{dn_b} < 0$
(ii) $\frac{\partial v_{\mathcal{B}}}{\partial n_p} \leq 0$ and $\frac{\partial v_{\mathcal{B}}}{\partial n_b} \leq 0$
(iii) $\frac{\partial v_{\mathcal{B}}}{\partial n_b} \leq \frac{\partial v_{\mathcal{B}}}{\partial n_p}$

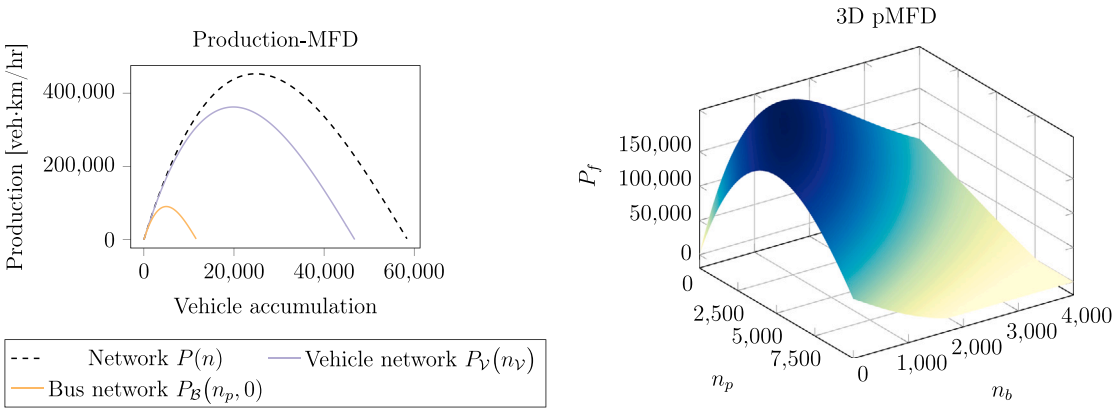
The first element guarantees that the vehicle accumulation at capacity reduces as the number of buses in the network increases. With respect to the vehicle speed function, part (ii) ensures that the running speed in the bus network decreases with both bus and vehicle accumulations. Part (iii) shows that the speed decrease due to the addition of one bus is greater in absolute terms than that due to an addition of a pooling vehicle with $\frac{\partial v}{\partial n_b} = \frac{\partial v}{\partial n_p}$.

It is noteworthy that the speed obtained from the bus network MFD is the vehicle speed, i.e., the running speed $v_{\mathcal{B}}$ of ride-hailing drivers performing a pooled trip. Since buses make repetitive stops at stations to board and alight passengers, we compute the operating bus speed v_b by reducing $v_{\mathcal{B}}$ by a factor less than or equal to one. This factor depends on the average spacing between stops $\bar{s}$ and on the dwell time $\bar{t}_d$ such that

$$v_b = \left(\frac{1}{1 + v_{\mathcal{B}} \frac{\bar{t}_d}{\bar{s}}}\right) v_{\mathcal{B}}. \tag{3}$$

We conclude first that $v_b < v_{\mathcal{B}}$ since the factor $\frac{1}{1+v_{\mathcal{B}}\frac{\bar{t}_d}{\bar{s}}} < 1$. Moreover, by writing v_b in terms of v, it is possible to show that $\frac{\partial v_b}{\partial v} > 0$ and $\frac{\partial^2 v_b}{\partial v^2} < 0$. This also applies when we write v_b in terms of $v_{\mathcal{B}}$, and find the first and second order derivatives with respect to $v_{\mathcal{B}}$. Furthermore, by finding $\frac{\partial v_b}{\partial v_{\mathcal{B}}}$, it is straightforward to show that this derivative is always less than 1.

For the purpose of this analysis, we consider that the bus network is in the congested regime when $\frac{\partial P_{\mathcal{B}}}{\partial n_p} = v_{\mathcal{B}} + (n_p + n_b \frac{\partial v_b}{\partial v_{\mathcal{B}}}) \frac{\partial v_{\mathcal{B}}}{\partial n_p}$ is negative. It follows from part (iii) and from (3) that $\frac{\partial P_{\mathcal{B}}}{\partial n_b}$ is also negative because $\frac{\partial P_{\mathcal{B}}}{\partial n_b} \leq \frac{\partial P_{\mathcal{B}}}{\partial n_p}$.



(a) Production MFD (b) Bus network 3D-pMFD

Fig. 2. Vehicle and passenger production MFDs.

Lemma 2. *If the overall network speed function satisfies Definition 1, then the production function for the bus network* $P_{\mathcal{B}}(n_p, n_b) = n_p v_{\mathcal{B}} + n_b v_b$ *is concave in* n_p.

Proof. The second order partial derivative of $P_{\mathcal{B}}$ with respect to n_p is

$$\frac{\partial^2 P_{\mathcal{B}}}{\partial n_p^2} = \frac{1}{\bar{\alpha}} k \left(\frac{1}{k\bar{\alpha}} n_b (v')^2 \frac{\partial^2 v_b}{\partial v^2} + 2v' + \frac{1}{\bar{\alpha}} \left(n_p + \frac{n_b}{(1 + kv\frac{\bar{l}_d}{\bar{s}})^2} \right) v'' \right), \tag{4}$$

which is always negative because the first element is negative since $\frac{\partial^2 v_b}{\partial v^2} < 0$. Moreover, according to Definition 1, $2v' + nv'' < 0$, and for $n = \frac{1}{\bar{\alpha}}(n_p + n_p)$, $n \in [0, n_{jam}]$, $2v' + \frac{1}{\bar{\alpha}}(n_p + n_b)v'' < 0$. Given $\frac{1}{(1+kv\frac{\bar{l}_d}{\bar{s}})^2} < 1$, it follows that $n_p + \frac{n_b}{(1+kv\frac{\bar{l}_d}{\bar{s}})^2} < n_p + n_b$. Hence $P_{\mathcal{B}}$ is concave in n_p. □

As buses belong to the mass transit transportation mode category, it is important to verify that the disturbance to the bus network by pooled vehicles is minimized. For this purpose, we consider passenger production instead of the production when capturing the influence of pooled vehicles on the flow of bus passengers. We let P_f be the passenger production value in the bus network expressed in passenger kilometers per hour, then P_f is defined by

$$P_f(n_p, n_b) = \bar{o}_p n_p v_{\mathcal{B}} + \bar{o}_b n_b v_b \,, \tag{5}$$

where $\bar{o}_p$ and $\bar{o}_b$ are the average occupancies of pooled vehicles and buses respectively. The first term of (5) computes the production of passengers performing a pooled trip, and the second term computes the passenger flow for buses.

To exemplify the network MFDs, Fig. 2 shows the MFD functions we will use in the numerical section of this work to compute the relationship between speed and accumulation. If the value of α is set to 0.8, then Fig. 2(a) reflects how the production MFDs for the bus and vehicle networks are expressed as a partition of the full network MFD. Every line shows the dependency of the production on the actual category of only vehicles that are utilizing the network under consideration, i.e. n for the full network, $n_{\mathcal{V}}$ for the vehicle network, and n_p for the bus network because n_b is set to 0. Furthermore, when the bus network MFD is translated to the three-dimensional space where both the accumulation of buses and vehicles are considered, then looking at the passenger flow instead of the vehicle one is substantiated by the relatively larger occupancy of buses. Fig. 2(b) displays an example of the 3D passenger MFD that we denote by 3D pMFD where the values of production P_f are expressed in terms of passenger rather than vehicle production. Once all these relationships are established, it becomes possible to convert all demand values into vehicle accumulation for the purpose of estimating delays for all commuters in the network.

3.3. Network accumulation

Assuming that the demand values Q^{pv}, Q^{rh}, and Q^b are known, we explore next how to use the network speed definitions presented in Section 3.2 to estimate the ride-hailing fleet size but also the accumulations in the vehicle and bus network. We particularly assess how this speed dependency affects the capacity and the service level of ride-hailing markets.

3.3.1. Ride-hailing services

Given that ride-hailing services are central in our analysis, we start first by characterizing the fleet size necessary for their service provision. For simplicity, we assume a spatially homogeneous distribution of empty ride-hailing vehicles and passengers across the

network. A monopoly platform receives all requests, and its functionality is to arrange a proper vehicle-request match and dispatch but also to seize appropriate pooling opportunities when possible. If we discard the pooling alternative, the fleet size N required to serve the totality of the ride-hailing demand Q^{rh} consists of (i) idling vehicles, (ii) dispatched vehicles on their way to pick up a passenger, and (iii) occupied vehicles. Hence the fleet size N is computed as

$$N = I + Q^{rh}\frac{d(I)}{v_v} + Q^{rh}\frac{\bar{l}}{v_v}, \tag{6}$$

where, I is the number of idle vehicles, d is the dispatched distance from a request's origin location to the nearest idle vehicle. The dispatched distance d is itself dependent on the number of unoccupied vehicles such that $d' < 0$ because the higher this number, the greater the chance of matching the passenger to a neighboring empty vehicle. Consequently, when idling vehicles are abundant, the pick-up occurs almost instantaneously where $\lim_{I\to+\infty} d(I) = 0$. When the number of idling vehicles is too low however, drivers are expected to travel longer to pick up a passenger and as a consequence, we consider that the $\lim_{I\to 0+} d(I) = \infty$. We additionally assume that d is convex and hence $d'' > 0$. Without loss of generality, we adopt for the following analysis that the number of private vehicles is constant in the vehicle network and hence if $n_v = n_{pv} + N$, we know that $v'_v = \frac{\partial v_v}{\partial n_v} = \frac{\partial v_v}{\partial N}$ and $v''_v = \frac{\partial^2 v_v}{\partial n_v^2} = \frac{\partial^2 v_v}{\partial N^2}$.

Proposition 1. *For a fixed number of idle vehicles $I \in [0, N]$, the function Q^{rh} is concave in N and hence admits a unique maximum value.*

Proof. From (6), we know that the supply rate Q^{rh} is

$$Q^{rh} = \frac{(N - I)v_v}{\bar{l} + d}. \tag{7}$$

The second order partial derivative of Q^{rh} with respect to N is given by

$$\frac{\partial^2 Q^{rh}}{\partial N^2} = \frac{2v'_v + (N - I)v''_v}{\bar{l} + d}. \tag{8}$$

From Lemma 1, we know that $2v'_v + n_v v''_v$ is always negative. Since $N \le n_v$, it follows that $2v'_v + (N - I)v''_v \le 0$. We conclude that for a fixed $I \in [0, N]$, the Q^{rh} admits a unique maximum value because its function is concave in N since $\bar{l} + d > 0$ and therefore $\frac{\partial^2 Q^{rh}}{\partial N^2}$ is negative. □

To further understand the variation of the ride-hailing supply with the fleet size, we take the derivative of Q^{rh} with respect to N

$$\frac{\partial Q^{rh}}{\partial N} = \frac{v_v + (N - I)v'_v}{\bar{l} + d}, \tag{9}$$

where the sign of the partial derivative is dependent on the sign of the numerator. We infer from Proposition 1 hence that when the influence of fleet size on network speed is accounted for in the ride-hailing supply function, the supply rate achieves a maximum, after which increasing the fleet size negatively impacts Q^{rh}. This is because when $v_v > -(N - I)v'_v$, the supply rate Q^{rh} strictly increases with the fleet size N and when $v_v < -(N - I)v'_v$, the supply strictly decreases with the fleet size.

Remark 1. It should be noted that under the assumption that the speed is constant as in, e.g., (Zha et al., 2018b), the supply rate will always increase with the fleet size. Namely, if the influence of the fleet size on speed is ignored such that v_v remains constant independently of N, the partial derivative of Q^{rh} with respect to N then becomes equal to

$$\frac{\partial Q^{rh}}{\partial N} = \frac{v_v}{\bar{l} + d}, \tag{10}$$

which is always greater or equal to zero because $v_v \ge 0$ and $\bar{l} + d > 0$. This implies that $\frac{\partial Q^{rh}}{\partial N} \ge 0$, and therefore it is always possible to increase the capacity of the ride-hailing market by providing more vehicles.

This remark highlights the need to integrate the speed variant – in our case with an MFD representation – to better capture interactions between different modes within our space allocation framework.

In a similar manner, it is possible to show that the dispatching time from an idling vehicle active position to the pick-up location of a request is influenced by the vehicle network speed. This quantity, which also designates the waiting time of requests, is therefore a function of the fleet size supplied by the ride-hailing operator and the number of idling vehicles in the network. First, the variations in the availability of idle vehicles are associated with different ride-hailing market states as demonstrated by Castillo et al. (2017). In fact, when the density of idle vehicles is high enough, the ride-hailing market exists in an *efficient* state where vehicles do not spend much of their time dispatching. On the contrary, and in an *inefficient* supply state, the dispatching distance is very high such that too few idle vehicles remain available for pick-ups. Second, the fleet size alters the network speed, thus implicitly dictating the waiting time of requests or alternatively the dispatching time of idle vehicles. We denote by w this waiting time of a request before being picked up. Its expression is given by the ratio of the dispatched distance d to the vehicle network speed v_v such that $w = \frac{d}{v_v}$.

Proposition 2. *For a fixed Q^{rh}, w admits a minimum with respect to N. This minimum occurs when the ride-hailing market is in the efficient state and when $v_v + (N - I)v'_v > 0$.*

Proof. First we observe that whether the ride-hailing market is in one of the efficient or inefficient regime is dependent on the sign of

$$\frac{\partial Q^{rh}}{\partial I} = -\frac{v_v + d' Q^{rh}}{\bar{l} + d}. \tag{11}$$

Rearranging the previous equation, we get

$$\frac{\partial Q^{rh}}{\partial I}\frac{1}{d'} = \frac{v_v}{\bar{l} + d}\left(-\frac{1}{d'} - \frac{Q^{rh}}{v_v}\right). \tag{12}$$

Assuming that the fleet size N remains constant, then when all vehicles are idling, i.e, when $I = N$, Q^{rh} must be equal to 0, and $\frac{\partial Q^{rh}}{\partial I}\frac{1}{d'} > 0$ which means that, since $d' < 0$, $\frac{\partial Q^{rh}}{\partial I} < 0$. As I decreases, $-\frac{1}{d'}$ decreases and $\frac{Q^{rh}}{v_v}$ increases. Hence, it follows from (11) that when $d'Q^{rh} > -v_v$, $\frac{\partial Q^{rh}}{\partial I}$ continues to be negative up until v_v and $-d'Q^{rh}$ becomes equal and therefore the supply rate achieves its maximum. As d' continues to decrease with I, $\frac{\partial Q^{rh}}{\partial I}\frac{1}{d'}$ becomes negative which implies that $\frac{\partial Q^{rh}}{\partial I} > 0$. This means that Q^{rh} starts increasing with I because $\frac{\partial Q^{rh}}{\partial I} > 0$. As I approaches zero, $-\frac{1}{d'}$ will always be strictly less than $\frac{Q^{rh}}{v_v}$ because if both functions intersect again, this happens at a simultaneously tangent and saddle/local optimum point for Q^{rh}. However, $\frac{\partial}{\partial I}(-\frac{1}{d'}) = \frac{d''}{d'^2} > 0$ which is strictly positive. This means that this point does not exist and hence $\frac{\partial Q^{rh}}{\partial I} > 0$ as I becomes lower than the optimum point where the maximum system capacity is achieved. In this region, we have that $d'Q^{rh} < -v_v$ and the ride-hailing market is in the inefficient state.

Since w is a function of both I and N but I is also an implicit function of N, then the partial derivative of w with respect to N is

$$\frac{\partial w}{\partial N} = \frac{dw}{dN} + \frac{\partial w}{\partial I}\frac{\partial I}{\partial N} \tag{13}$$

$$= -\frac{dv_v'}{v_v^2} + \frac{d'}{v_v}\frac{v_v + v_v'(N-I)}{Q^{rh}d' + v_v}. \tag{14}$$

First for a given N, $\frac{\partial w}{\partial I} = \frac{d'}{v_v} < 0$ which means that the waiting time always decreases with the number of idle vehicles. In other terms, if I_{eff} is the number of idle vehicles in the efficient regime and I_{ineff} is that in the inefficient regime for a fixed supply rate Q^{rh}, then $w(I_{\text{eff}}) < w(I_{\text{ineff}})$ irrespective of the value of N. Consequently, if a minimum of w with respect to N exists, this minimum occurs when the ride-hailing market is in the efficient state, i.e, when $d'Q^{rh} > -v_v$. Therefore, we limit the scope of this analysis to when the ride-hailing market is in this state. If the network is in the congested regime, $0 > v_v + n_v v_v'$ according to Definition 1. Nevertheless, $v_v + (N-I)v_v'$ is either positive or negative since $v_v + (N-I)v_v' > v_v + n_v v_v'$. We know however that when $v_v + (N-I)v_v' \leq 0$, $\frac{\partial w}{\partial N}$ is positive and hence the waiting time increases with N. However, when the network is in the uncongested regime, or when the network is congested and $v_v + (N-I)v_v' > 0$, $\frac{\partial w}{\partial N}$ happens to be either positive or negative depending on the values of N and I. Moreover, when the ride-hailing market just shifts to the efficient state, we know that $Q^{rh}d' + v_v$ goes to 0^+, and $\frac{\partial w}{\partial N}$ goes to $-\infty$. We conclude that the waiting time decreases with N at the start of the efficient state. This is because $\frac{\partial I}{\partial N}$ is positive in this case which means that the start of the efficient regime is achieved for the lowest N. This is true if $v_v + (N-I)v_v' > 0$ for the lowest N satisfying Q^{rh}. Accordingly, w admits a minimum with respect to N which occurs when the ride-hailing market is in the efficient state, and when $v_v + (N-I)v_v'$ is positive. □

So far we have assessed the fleet size required to exclusively serve solo trips. In our model however, we consider a ride-splitting market where users have the choice to travel solo or to pool. Consequently, following Ke et al. (2020a), we introduce two additional elements into (6) to accommodate the possibility of trip sharing. Vehicles are dispatched to perform either one of the solo or pooled trips. They then travel for a distance $\bar{l}$ to drop off a single passenger, or a distance $\bar{l} + \Delta l_d$ to pick up a second passenger and drop them both off. Δl_d is the driver detour defined as the additional distance traveled by drivers to perform a pooled trip relative to a solo one. It is a function of the pooling demand Q^p. Mainly, the higher the pooling demand, the better the trip match. This implies that the total length of a pooled trip becomes shorter when more passengers opt for pooling. Therefore, the driver detour decreases with Q^p which implies that $\Delta l_d' = \frac{\partial \Delta l_d}{\partial Q^p} \leq 0$. Moreover, we assume that $\lim_{Q^p \to 0} \Delta l_d = \infty$ and $\lim_{Q^p \to \infty} \Delta l_d = 0$. It follows that the fleet size in ride-splitting markets is given by:

$$N = I + \left(Q^s + \frac{1}{2}Q^p\right)\frac{d}{v_v} + Q^s\frac{\bar{l}}{v_v} + \frac{1}{2}Q^p\left(\frac{\bar{l} + \Delta l_d(Q^p)}{v_B}\right). \tag{15}$$

The four terms in (15) represent vehicles that are: (i) idling, (ii) dispatching, (iii) occupied with one passenger, and (iv) occupied with more than one passenger and assigned to a pooled trip. Notice that the speed function we use for the different terms in the equation directly emerges from the network allocation policy proposed. Besides idling, both dispatched and solo-trip vehicles travel at a speed v_v. In contrast, vehicles performing a pooled trip travel at a speed v_B. Clearly, ride-splitting markets have the ability to serve more trips with the same fleet size. Nevertheless, the detour distance is a crucial factor in determining the potential and efficiency of the pooling alternative. According to our proposed model, a portion of this fleet size contributes to the accumulation in the vehicle network whereas the remaining fraction enters the bus network and directly impacts the speed of buses.

3.3.2. Vehicle network accumulation

The accumulation in the vehicle network consists of (i) idle vehicles, (ii) dispatched vehicles, (iii) ride-hailing vehicles performing a solo trip, and (iv) private vehicles. Accordingly, the total network accumulation in the vehicle network is

$$n_{\mathcal{V}} = d^{-1}(\tau v_{\mathcal{V}}) + \left(Q^s + \frac{1}{2}Q^p\right)\tau + Q^s\frac{\bar{l}}{v_{\mathcal{V}}} + Q^{pv}\frac{\bar{l}}{v_{\mathcal{V}}}. \tag{16}$$

In (16), we assume that private vehicles share the same average trip length as solo rides. Moreover, $\tau = \frac{d}{v_{\mathcal{V}}}$ is the expected waiting time that the platform aims at achieving to guarantee an acceptable level of service. Subsequently, the number of idle vehicles required to maintain this service level is given by $d^{-1}(\tau v_{\mathcal{V}})$ where $\tau v_{\mathcal{V}}$ stands for the distance traveled by ride-hailing vehicles to pick up a passenger who waited for a time τ. We conclude that the number of idle vehicle I is a function of $v_{\mathcal{V}}$ and is therefore equal to $d^{-1}(\tau v_{\mathcal{V}})$ where $I' = \frac{\partial I}{\partial v_{\mathcal{V}}} < 0$.

Proposition 3. *The accumulation of the vehicle network $n_{\mathcal{V}}$ increases with Q^s when $-I'v_{\mathcal{V}} < I + \left(Q^s + \frac{1}{2}Q^p\right)\tau$ and the vehicle network is in the uncongested regime.*

Proof. Finding the implicit derivative of $n_{\mathcal{V}}$ with respect to Q^s from (16), we get that

$$\frac{\partial n_{\mathcal{V}}}{\partial Q^s} = \frac{\frac{1}{2}\tau v_{\mathcal{V}} + \bar{l}}{v_{\mathcal{V}} + (\frac{Q^{pv}\bar{l}}{v_{\mathcal{V}}} + \frac{Q^s\bar{l}}{v_{\mathcal{V}}} - I'v_{\mathcal{V}})v'_{\mathcal{V}}}. \tag{17}$$

The numerator of (17) is always positive and the sign of the derivative depends on the denominator. We know from Lemma 1 that the production function $P_{\mathcal{V}}$ is concave. Therefore, when the vehicle network is in the uncongested regime, $v_{\mathcal{V}} + n_{\mathcal{V}}v'_{\mathcal{V}}$ is positive. Given the assumption that $-I'v_{\mathcal{V}} < I + \left(Q^s + \frac{1}{2}Q^p\right)\tau$, we conclude that $v_{\mathcal{V}} + n_{\mathcal{V}}v'_{\mathcal{V}} < v_{\mathcal{V}} + (\frac{Q^{pv}\bar{l}}{v_{\mathcal{V}}} + \frac{Q^s\bar{l}}{v_{\mathcal{V}}} - I'v_{\mathcal{V}})v'_{\mathcal{V}}$. This leads to the denominator of (17) also being positive when the vehicle network is in the uncongested regime, and hence $\frac{\partial n_{\mathcal{V}}}{\partial Q^s} > 0$. □

3.3.3. Bus network accumulation

With respect to the bus network, estimating the accumulation requires to separate between buses and ride-hailing vehicles as they both have a non-identical influence on speed. For this purpose, we split the computation of the accumulation into two parts, where the first part, the one for pooled vehicles, is given by

$$n_p = \frac{1}{2}Q^p\left(\frac{\bar{l} + \Delta l_d(Q^p)}{v_{\mathcal{B}}}\right), \tag{18}$$

and the second part, the one for buses, is given by

$$n_b = \frac{Q^b\bar{l}_b}{\bar{o}_b\bar{v}_b}. \tag{19}$$

The number of vehicles required to serve all pooled trips utilizing the bus network is dependent on the demand for pooling Q^p, but also on the average driver detour which itself is a function of Q^p as shown by (18). Besides the vehicle accumulation, the number of buses in network $\mathcal{B}$ is computed using (19) where $\bar{l}_b$ is the average bus trip length generally greater than $\bar{l}$. Assuming that the average bus occupancy must be maintained at $\bar{o}_b$ and the average bus speed at $\bar{v}_b$, (19) returns the static number of buses required to serve Q^b.

In the following steps, we investigate how the pooled trip duration and the detour time vary with the demand for pooling, and then subsequently with the accumulation of the pooled drivers in the bus network. In order to accomplish this, we start primarily by characterizing the equilibria in the bus network. This narrows down to finding the range of values of Q^p such that the production of pooling vehicles that we denote by $P_p(n_p, n_b) = n_p v_{\mathcal{B}}(n_p, n_b)$ intersects with the trip demand F such that $F = \frac{1}{2}Q^p(\bar{l} + \Delta l_d)$. For the purpose of this analysis, it is logical to consider that $\frac{\partial F}{\partial Q^p} = \frac{1}{2}(\bar{l} + \Delta l_d + Q^p\Delta l'_d) > 0$ which means that the demand expressed in vehicle kilometers per hour will always increase with Q^p regardless of the reduction in the driver detour distance. Moreover, we assume that $\lim_{Q^p\to\infty} F = 0$ and $\lim_{Q^p\to 0} F = \infty$.

Lemma 3. *For all n_p and n_b such that $n_b + n_p \leq \bar{\alpha}n_{jam}$, (18) has at most two solutions depending on the value of Q^p. In case two solutions exist, one occurs when $\frac{\partial P_p}{\partial n_p} > 0$ and the other occurs when $\frac{\partial P_p}{\partial n_p} < 0$.*

Proof. We start first by showing that P_p is concave in n_p. To do so, we compute the second order partial derivative of P_p with respect to n_p

$$\frac{\partial^2 P_p}{\partial n_p^2} = \frac{1}{\bar{\alpha}}\left(2v' + \frac{1}{\bar{\alpha}}n_p v''\right), \tag{20}$$

which is always negative because if $v'' < 0$, then the sign of the derivative is straightforward. If, however, $v'' > 0$, $n_p < n_p + \frac{n_b}{(1+kv\frac{\bar{l}_d}{\bar{s}})^2}$ and since $2v' + \frac{1}{\bar{\alpha}}\left(n_p + \frac{n_b}{(1+kv\frac{\bar{l}_d}{\bar{s}})}\right) < 0$ following Lemma 2, the sign of (20) is always negative. Given that P_p is concave in n_p, it

admits a unique maximum. Moreover, when $n_p = 0$ or when $n_p = \bar{\alpha} n_{jam} - n_b$, $P_p = 0$ for a given n_b. If $\bar{n}_p^* = \arg\max(P_b)$ is the value of the pooled vehicle accumulation for which P_p is maximized, P_p is increasing when $n_p < \bar{n}_p^*$ and decreasing when $n_p > \bar{n}_p^*$. As a consequence, if the demand value F is less than this maximum, then there exist two intersection points between the network supply and pool vehicle demand. As F continues to increase with Q^p and becomes equal to that maximum, the bus network is operating at capacity such that if the demand for pooling increases further, we will have no solutions. □

For the remainder of this work, we denote the maximum pooling demand that the bus network can serve by Q^p_{max}.

Proposition 4. *The accumulation of pooling vehicles n_p increases with Q^p when the bus network is in the uncongested regime.*

Proof. Computing the derivative of (18) with respect to Q^p, we obtain

$$\frac{\partial n_p}{\partial Q^p} = \frac{\frac{\partial F}{\partial Q^p}}{v_B + n_p \frac{\partial v_B}{\partial n_p}}. \tag{21}$$

Since $\frac{\partial F}{\partial Q^p}$ is always positive, then the sign of $\frac{\partial n_p}{\partial Q^p}$ is dependent on the sign of the denominator. When the bus network is in the uncongested regime, we know that $\frac{\partial P_B}{\partial n_p} > 0$ following Definition 2. This implies that $v_B + (n_p + n_b \frac{\partial v_b}{\partial v_B}) \frac{\partial v_B}{\partial n_p} > 0$ and hence $v_B + n_p \frac{\partial v_B}{\partial n_p} > 0$. As a consequence, $\frac{\partial n_p}{\partial Q^p} > 0$ which means that as the demand for pooling increases, the accumulation increases when the bus network is uncongested. □

Assuming that the solution exists for a given Q^p, finding the pooling demand Q^p that will minimize the total trip length for drivers is equivalent to finding the point where, if passengers continued to pool, they will worsen the cost of the remaining pooling vehicles in the bus network. In fact, when Q^p increases, this causes the driver detour to decrease hence leading to a non-monotone change in the bus network accumulation depending on whether we are located in the uncongested or congested regime. Let T be the driver trip length defined by $\frac{\bar{l} + \Delta l_d}{v_B}$.

Proposition 5. *If $Q^p \Delta l'_d$ approaches zero for a sufficiently large value of Q^p, v_B is strictly positive for small values of Q^p, and $\lim_{Q^p \to \infty} \Delta l'_d = 0$, then T admits a minimum with respect to Q^p when $\frac{\partial P_B}{\partial n_p} > 0$.*

Proof. First, $\lim_{Q^p \to 0} T = \infty$ because $\lim_{Q^p \to 0} \Delta l_d = \infty$ and v_B is positive. Therefore, T is decreasing for small values of Q^p. Next, let us find $\frac{\partial T}{\partial Q^p}$, the partial derivative of T with respect to Q^p

$$\frac{\partial T}{\partial Q^p} = \frac{dT}{dQ^p} + \frac{\partial T}{\partial n_p} \frac{\partial n_p}{\partial Q^p} \tag{22}$$

$$= \frac{\Delta l'_d}{v_B} + \frac{-\frac{\partial v_B}{\partial n_p}(\bar{l} + \Delta l_d)}{v_B^2} \left(\frac{\frac{\partial F}{\partial Q^p}}{v_B + n_p \frac{\partial v_B}{\partial n_p}} \right). \tag{23}$$

Since $\Delta l'_d < 0$, the first term of (23) is negative. Moreover, when the production in the bus network strictly increases with n_p, $v_B + (n_p + n_b \frac{\partial v_b}{\partial v_B}) \frac{\partial v_B}{\partial n_p} > 0$ and hence $v_B + n_p \frac{\partial v_B}{\partial n_p} > 0$. This implies that the second term of the equation is positive because $\frac{\partial v_B}{\partial n_p} \leq 0$ and, as assumed earlier, $\frac{\partial F}{\partial Q^p} = \frac{1}{2}(\bar{l} + \Delta l + Q^p \Delta l'_d) > 0$. According to the definition of the driver detour function, Δl_d is strictly positive and decreasing with Q^p since $\Delta l'_d < 0$. Due to the assumption that $\lim_{Q^p \to \infty} \Delta l'_d = 0$, when the bus network is sufficiently large, the value of Q^p_{max} is high enough such that Δl_d is approaching 0 and so is $\Delta l'_d$ while Q^p has not yet attained Q^p_{max}. Given that the limit of $Q^p \Delta l'_d$ is zero for sufficiently large Q^p, the second term of (23) is bounded while the first term goes to 0. Consequently, when Q^p tends to a sufficiently large Q^p_{max}, $\frac{\partial T}{\partial Q^p}$ becomes positive. □

Remark 2. The previous analysis also applies for the pooled vehicle accumulation n_p since $\frac{\partial n_p}{\partial Q^p} > 0$ which means that as the demand for pooling increases, the bus network accumulation does too. As a result, the same behavior of the trip time is observed with respect to n_p. However, when $v_B + n_p \frac{\partial v_B}{\partial n_p} < 0$, the driver trip duration always increases with n_p since (23) is always negative and $\frac{\partial n_p}{\partial Q^p} < 0$.

Proposition 5 implies that there exists at least one value for which $\frac{\partial T}{\partial Q^p}$ is equal to 0. Moreover, because $\frac{\partial T}{\partial Q^p} > 0$ when $Q^p = Q^p_{max}$, we know that T admits a minimum over the range $(0, Q^p_{max})$ for a sufficiently large value of Q^p_{max}. Fig. 3 displays an example of the variation of the driver trip duration as function of Q^p. When Q^p is small, the decrease in the detour overtakes the decrease in speed due to the increase in accumulation, and we observe an improvement over the total pooled trip duration. However, after a certain point, it is the increase in accumulation that causes the bus network speed to deteriorate thus the observed rise in the trip duration.

Having shown the occurrence of the minimum for the total driver trip duration, it is also possible to show its occurrence for the driver detour time.

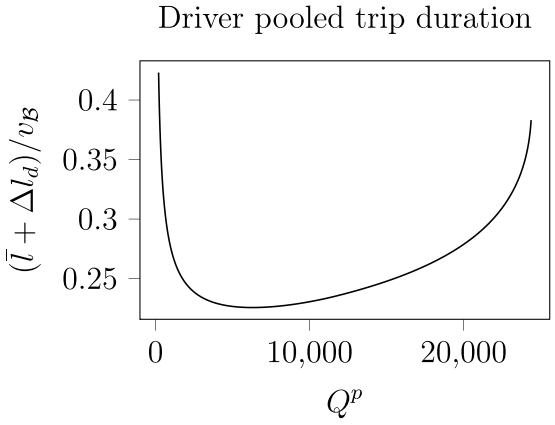


Fig. 3. Total pooled trip length as function of the pooling demand Q^p.

3.4. System optimum

In the following section, we investigate system optimum solutions under the condition that the fraction of travelers that makes a pooled trip is centrally chosen by the system operator. Because the purpose of allowing pooled vehicles to use bus lanes is to decrease the total delays for all commuters in the network, we investigate under a fixed network spatial split α, α being the fraction of network space allocated to the vehicle network, for what values of β we minimize the Passengers Hours Traveled (PHT) for all three mode users: buses, private vehicles, and ride-hailing users. The ultimate objective is to mitigate congestion in the vehicle network while simultaneously making sure that disturbances to buses are contained within acceptable ranges. Therefore, we model our objective as the sum of the individual PHT for every category of travelers with the total PHT function being expressed as

$$\underset{\beta\in[0,1]}{\text{minimize}} \quad Q^{pv}\frac{\bar{l}}{v_V} + \beta Q^{rh}\frac{\bar{l}}{v_V} + (1-\beta)Q^{rh}\left(\frac{\bar{l}+\Delta l_p}{v_B}\right) + Q^b\frac{\bar{l}_b}{v_b}, \tag{24}$$

where the first, second, and last terms correspond to delays for private vehicles, solo passengers, and bus users respectively. The third term reflects the PHT for pooled passengers using the bus network. Note that in (24), we substitute the driver detour Δl_d with the passenger detour Δl_p to consider the additional distance incurred by passengers compared to a direct trip between their origins and destinations. The behavior of the passenger detour however is assumed to be comparable to that of the driver detour such that they both monotonically decrease with Q^p. Minimizing over β is therefore equivalent to finding the optimal demand split between Q^s and Q^p such that $Q^s + Q^p$ is always equal to Q^{rh}. Therefore, knowing that $\frac{\partial Q^p}{\partial Q^s} = -1$, the first order partial derivative of the PHT with respect to Q^s – considering all the implicit dependencies of v_V, v_B, and v_b on β – represents the change in PHT due to a shift of one unit of demand from Q^p to Q^s, and is given by:

$$-\frac{1}{v_V}\frac{\partial v_V}{\partial n_V}\frac{\partial n_V}{\partial Q^s}\left(Q^s\frac{\bar{l}}{v_V} + Q^{pv}\frac{\bar{l}}{v_V}\right) + \frac{1}{v_B}\frac{\partial v_B}{\partial n_p}\frac{\partial n_p}{\partial Q^p}\left(Q^b\frac{\bar{l}_b}{v_B} + Q^p\frac{\bar{l}+\Delta l_p}{v_B}\right) + \frac{\bar{l}}{v_V} - \frac{\bar{l}+\Delta l_p}{v_B} - \frac{Q^p\Delta l'_p}{v_B}. \tag{25}$$

We note that in the previous equation, we replaced v_b by its function as shown in (3). Therefore, $\frac{\partial v_b}{\partial v_B} = \frac{1}{(1+v_B\frac{\bar{l}_d}{\bar{s}})^2}$. Following Proposition 3 that $\frac{\partial n_V}{\partial Q^s}$ is positive when the vehicle network is in the uncongested regime, the first term in (25) reflects the increase in PHT due to the addition of one unit of solo demand to the vehicle network. Next, following Proposition 4 that $\frac{\partial n_p}{\partial Q^p}$ is positive when the bus network is in the uncongested regime, the second term represents the decrease in network delays due to the shift of one demand unit from Q^p to Q^s. This shift also results in the addition of a solo trip duration and the reduction of a pooled trip duration as shown by the third and fourth terms of the equation respectively. Finally, the last term shows the increase in pooled trip detour duration due to the loss of one unit of pooled demand. We note here that further analysis of (25) is not straightforward because it requires having a well-defined shape of the PHT function, which in our problem is not the case. For this reason, the function is numerically assessed in Section 4.

3.5. Benchmark

To show under what conditions the proposed policy is efficient, we compare the delays obtained to the benchmark scenario where all ride-hailing users travel in the vehicle network. The consequence of this assumption is that the bus network becomes exclusively utilized by buses whereas pooling requests perform their trips on the vehicle network. They travel at a speed equal

Table 2
Main parameters.

Parameter	Symbol	Value	Unit
Demand for private vehicles	Q^{pv}	80 000	pax/h
Demand for buses	Q^{b}	30 000	pax/h
Demand for e-hailing	Q^{rh}	15 000	pax/h
Average vehicle trip length	$\bar{l}$	3.86	km
Average bus trip length	$\bar{l}_b$	5.40	km
Average spacing between bus stops	$\bar{s}$	0.8	km
Average bus dwell time at stops	$\bar{t}_d$	40	s
Ride-hailing platform target waiting time	τ	2	min
Average bus occupancy	$\bar{o}_b$	20	pax
Design bus speed	$\bar{v}_b$	18	km/h
Network area	A	107	km^2
Solo-ride fare	F^s	6	$
Monetary value of time	μ	30	$/h
Binary logit scale parameter	κ	1	–
Scale parameter of the nest	κ_p	2	–

to $v_{\mathcal{V}}$ whereas the bus speed v_b remains constant independently of Q^s. Minimizing network delays under this assumption consists thereafter of finding a solution of the objective function

$$\underset{\beta\in[0,1]}{\text{minimize}} \quad Q^{pv}\frac{\bar{l}}{v_{\mathcal{V}}} + \beta Q^{rh}\frac{\bar{l}}{v_{\mathcal{V}}} + (1-\beta)Q^{rh}\left(\frac{\bar{l}+\Delta l_p}{v_{\mathcal{V}}}\right) + Q^b\frac{\bar{l}_b}{v_b}. \tag{26}$$

The first order partial derivative of the PHT with respect to Q^s with $Q^s = \beta Q^{rh}$ under the benchmark scenario is

$$-\frac{1}{v_{\mathcal{V}}}\frac{\partial v_{\mathcal{V}}}{\partial n_{\mathcal{V}}}\frac{\partial n_{\mathcal{V}}}{\partial Q^s}\left(Q^s\frac{\bar{l}}{v_{\mathcal{V}}} + Q^{pv}\frac{\bar{l}}{v_{\mathcal{V}}} + Q^p\frac{\bar{l}+\Delta l_p}{v_{\mathcal{V}}}\right) + \frac{\bar{l}}{v_{\mathcal{V}}} - \frac{\bar{l}+\Delta l_p}{v_{\mathcal{V}}} - \frac{Q^p\Delta l'_p}{v_{\mathcal{V}}}. \tag{27}$$

The first element in (27) shows the effect of an increase in a unitary solo demand value on the users of the vehicle network. As opposed to the space allocation policy proposed, this shift has the same effect on delays for all private vehicles, solo trips, and pooled trips since they all use the vehicle network. The second and third terms are comparable to the trip duration terms in (25) except that the speed in this case is $v_{\mathcal{V}}$ for both solo and pooled trips. Similarly, the last term is the detour penalty due to the loss of one unit of pooled trip demand. Notice here that the number of buses in the bus network n_b is constant. Consequently, as no pooled vehicles are allowed on the bus network, the last term in (26) remains constant and is independent of Q^s. We note here that the reason why the solution does not always yield a value of $\beta = 0$ is that the detour comes into question in this particular setting. Consequently, when the total ride-hailing demand is lower than the critical boundary after which pooling becomes attractive, a scenario where $Q^p = Q^{rh}$ is not necessarily the optimum.

4. Numerical example

In the following section, we present a comprehensive numerical analysis of the influence of β on the network delays. Moreover, we carefully analyze how these delay values change with the property of the infrastructure by computing the PHT for different network configurations, i.e., for different α values. We assess the system optimum from a macroscopic approach for the main two scenarios: (i) the benchmark scenario that we refer to by $\{pv, s, p\}^{\mathcal{V}}|\{b\}^{\mathcal{B}}$, and (ii) the proposed strategy in this work that we denote by $\{pv, s\}^{\mathcal{V}}|\{p, b\}^{\mathcal{B}}$. To do so, we consider a numerical example using MFD, dispatched distance, and detour functions satisfying the assumptions we put forward throughout Section 3. The fixed parameters that we adopt in this example are presented in Table 2. We assume that the dispatched distance d is a function of the density of idle drivers, and is given by $d(I) = 0.63\sqrt{A/I}$ (Zha et al., 2018a). The production function of the entire network under consideration is $P(n) = A_0n^3 + B_0n^2 + C_0n$ where $A_0 = 5.74 \cdot 10^{-9}$, $B_0 = -1.02 \cdot 10^{-3}$, and $C_0 = 40$, and is concave over the interval of interest $n \in [0, 58536]$. The three-dimensional vehicle MFD for the bus network is obtained by multiplying the production MFD in the bus network with $k(n_b) = e^{-6.5\cdot10^{-4}n_b}$, and is used to compute the running speed in the bus network. With regard to the driver and passenger detour, their functional form is highly dependent on the OD distribution of Q^p. In this work, we use the demand pattern from Beojone and Geroliminis (2021), and we build the detour functions according to the matching optimization framework exhibited in the Appendix. The results are then fitted into the following functional forms where $\Delta l_d(Q^p)/\bar{l}$ or $\Delta l_p(Q^p)/\bar{l}$ is equal to $a_i/Q^{pb_i} + c_i$ for $i \in \{d, p\}$ depending on whether we are dealing with the driver or passenger detour. The values of the constants are equal to $a_d = 42.1$, $b_d = 0.570$, and $c_d = 0.395$ for the driver detour, and $a_p = 25.8$, $b_p = 0.674$, and $c_p = 0.0669$ for the passenger detour, assuming that $Q^p \in (0, 12000]$. When Q^p is greater than 12000, simulating the optimal matching becomes computationally exhaustive. We adopt therefore for this case a functional form that follows the same trend but goes to 0 for large enough Q^p such that the detour is equal to $a_ie^{-b_i(Q^p-11880)}$ and the values of the constants are $a_d = 0.596$ and $b_d = 1.61 \cdot 10^{-5}$ for the driver detour, and $a_p = 0.113$ and $b_p = 2.31 \cdot 10^{-5}$ for the passenger detour.

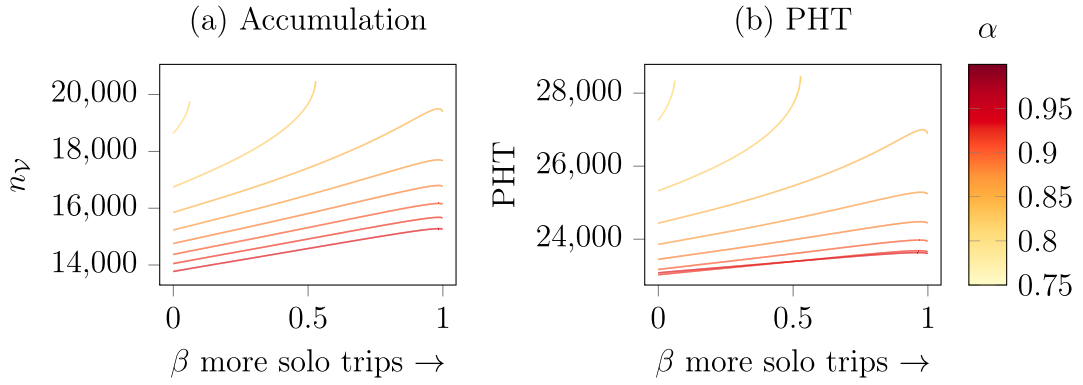


Fig. 4. Accumulation and PHT for $\{pv, s, p\}^{\mathcal{V}}|\{b\}^{B}$, i.e., when only buses are allowed to utilize the bus network.

4.1. System optimum and comparison to benchmark

For the scenario where all ride-hailing vehicles are using the vehicle network, the results for the accumulation and PHT as function of β for different spatial splits α are presented in Fig. 4. First, the network accumulation is naturally the lowest when the full ride-hailing demand is pooling and when α takes the largest possible value while still accommodating the bus demand. This means that a sufficiently high spatial capacity is granted for private vehicles and ride-hailing vehicles as can be seen from Fig. 4(a). However, this does not entail that the minimum delay is achieved for the same values of α because as α increases the space available for buses shrinks causing very high travel times to travelers in the bus network. This is inferred by looking at Fig. 4(b) where the PHT for different values of β starts increasing after reaching its lowest possible point. This occurs when decreasing the size of the bus network overtakes the gains from having a larger space dedicated to ride-hailing and private vehicles. This increase becomes more accentuated when the demand for buses Q^b is relatively high. The slight drop observed in the accumulation and PHT curves for high values of β is attributed to the high detour incurred when the demand for pooling is low. In fact, the detours for high values of β and therefore low values of Q^p become exorbitantly large such that it is always better for all passengers to travel solo instead of pooling their rides.

Fig. 5 displays the variation of accumulation and PHT as function of β for the strategy proposed in this paper. The results here are different from what is observed in Fig. 4. This distinction is significantly caused by two different trends: (i) as β increases, less vehicles use the bus networks creating fewer disturbances to bus users, ii) conversely, with the increase of β, more vehicles are needed to perform a solo ride in the vehicle network. This deteriorates the speed in the latter causing both the accumulation and delay to increase, particularly when the space allocated for the vehicle network is relatively small. What is interesting to note here is that, unlike the previous case, we obtain very similar values for the system optimum independently of the value of α. While changing α in reality requires significant infrastructure costs, influencing β is more of an operational decision. Therefore, given that demands can even change seasonally, having the ability to improve network conditions without changing α highlights the potential of the proposed approach. An example on when this happens is observed in Fig. 5(b) where the values of PHT are the same for different network splits α, and the reason behind this observation is the change in the fraction of pooling ride-hailing passengers.

To further understand the variation of the overall PHT with β, we plot the individual PHT values for every mode user for the benchmark scenario and for the proposed allocation strategy. The results are displayed in Figs. 6 and 7 respectively. First, when comparing the PHT for private vehicles for the two scenarios, the values are much higher for low values of α and hence the infrastructure capacity for this spatial split does not meet the demand levels for different modes even when $\beta = 0$ as shown in Fig. 6(a). Despite this common observation for the two scenarios, the delays of private vehicles are lower for the proposed strategy as seen in Fig. 7(a), mainly because less ride-hailing vehicles are utilizing network $\mathcal{V}$. Another intuitive difference we note is that, when looking at the benchmark scenario, bus delays remain unchanged as function of β (see Fig. 6(c)) as opposed to the delays in Fig. 7(c). This is because when the pooling demand is high, the outcome is a significant increase in the bus and pooling vehicle PHT values, whereas in the benchmark scenario, the bus PHT is constant with β as no pooling vehicles are allowed to switch to the bus network. On the contrary, as more passengers choose to travel solo, the solo passenger and private vehicle delays are impacted as shown in Figs. 7(a) and Fig. 7(b). With respect to the ride-hailing demand, the interpretation of the variation of PHT for solo rides in Fig. 6(b) is similar to that observed in Fig. 7(b). However, with respect to the variation of PHT for pooled rides in Fig. 6(d), the delays rise to values greater than the ones observed in Fig. 7(d) for small values of β indicating that when the number of drivers using the bus lanes is high, the deterioration in the bus network speed is very significant. To mitigate this influence, we assess in Section 4.4 the potential improvement we get by only sending a fraction of pooling vehicles to the bus network.

Next to understand how the variation of bus demand influences the system optimum, we plot the total PHT as function of β for different values of Q^b assuming that α is fixed, and we display the results in Fig. 8. It can be noted that for a predetermined network split α, as the bus demand increases, the system optimum occurs for higher values of β implying the need to restrict the delays for bus users especially when Q^b is relatively high.

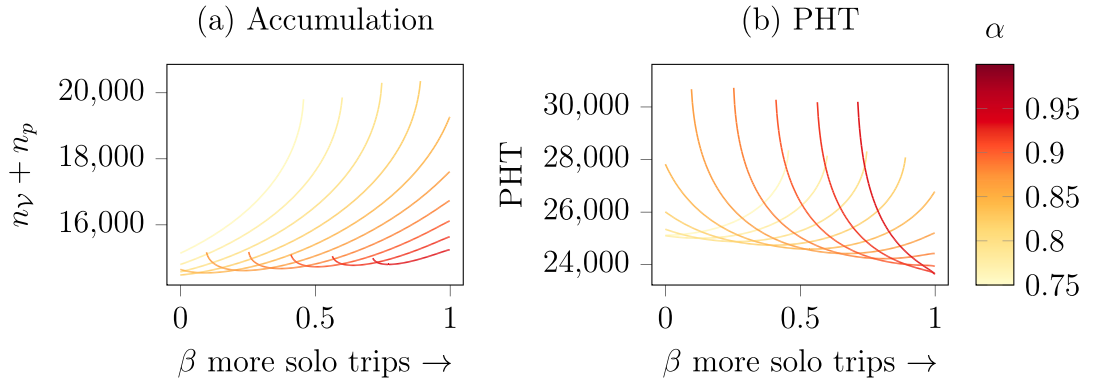


Fig. 5. Accumulation and PHT for $\{pv, s\}^V|\{p, b\}^B$, i.e., when pooling vehicles utilize the bus network.

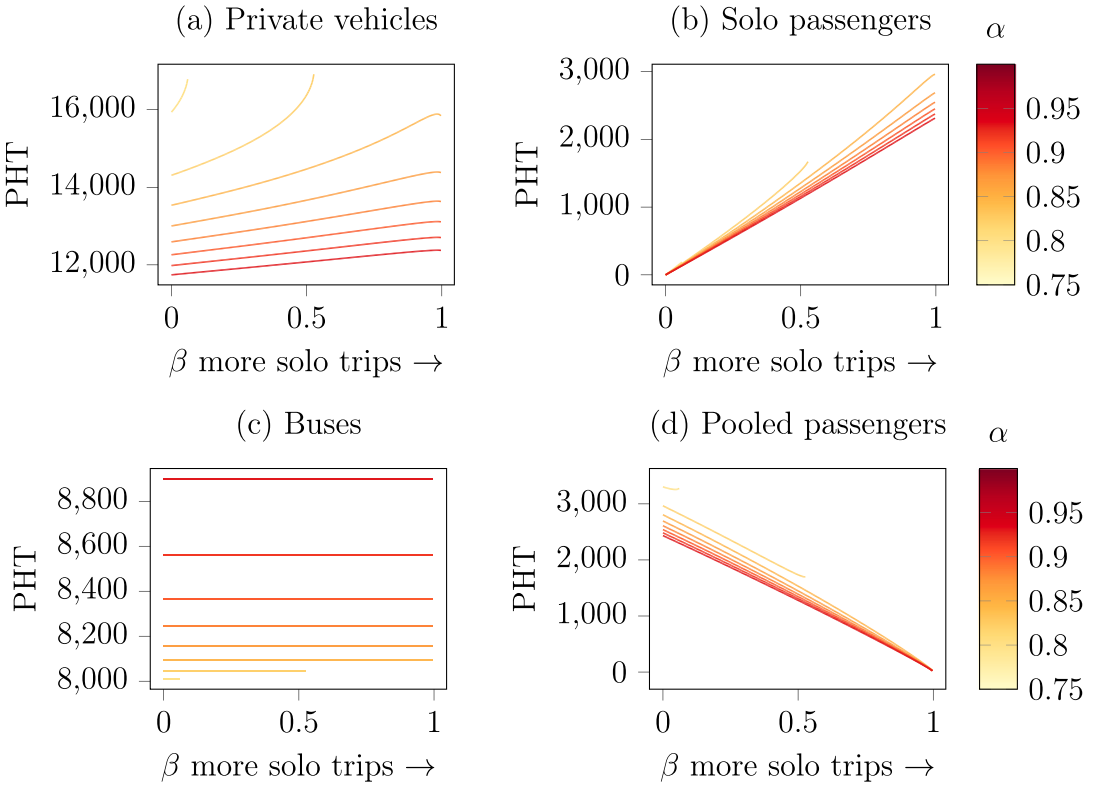


Fig. 6. PHT for every category of mode users in the network for the scenario $\{pv, s, p\}^V|\{b\}^B$, i.e., when only buses are allowed to utilize the bus network.

The results we display so far showed consistent observations irrespective of the values of α, β, or bus demand Q^b. In the next step, we assess the multi-modal delay functions for both the benchmark and the proposed allocation strategy, this time by varying the shape of the production function.

4.2. Results for different production functions

In practice, MFDs represent physical properties of the transport network, and they contain scatter due to various reasons (e.g. aggregating together non-steady states, weather conditions, different driving behavior, spatial heterogeneity of congestion). MFD functions can therefore vary significantly, and many efforts tend towards analyzing the accuracy of the different functional forms, but also advancing new parsimonious forms capable of fitting both uncongested and saturated regimes (Cheng et al., 2021). This is why it is crucial to assess the influence of different production functions on our results. For this purpose, we regenerate the results in Section 4.1 by considering an exponential production function $P_{\text{exp}}(n) = 200ne^{-\frac{1}{n_{cr}}n-1.4}$ and a logarithmic production

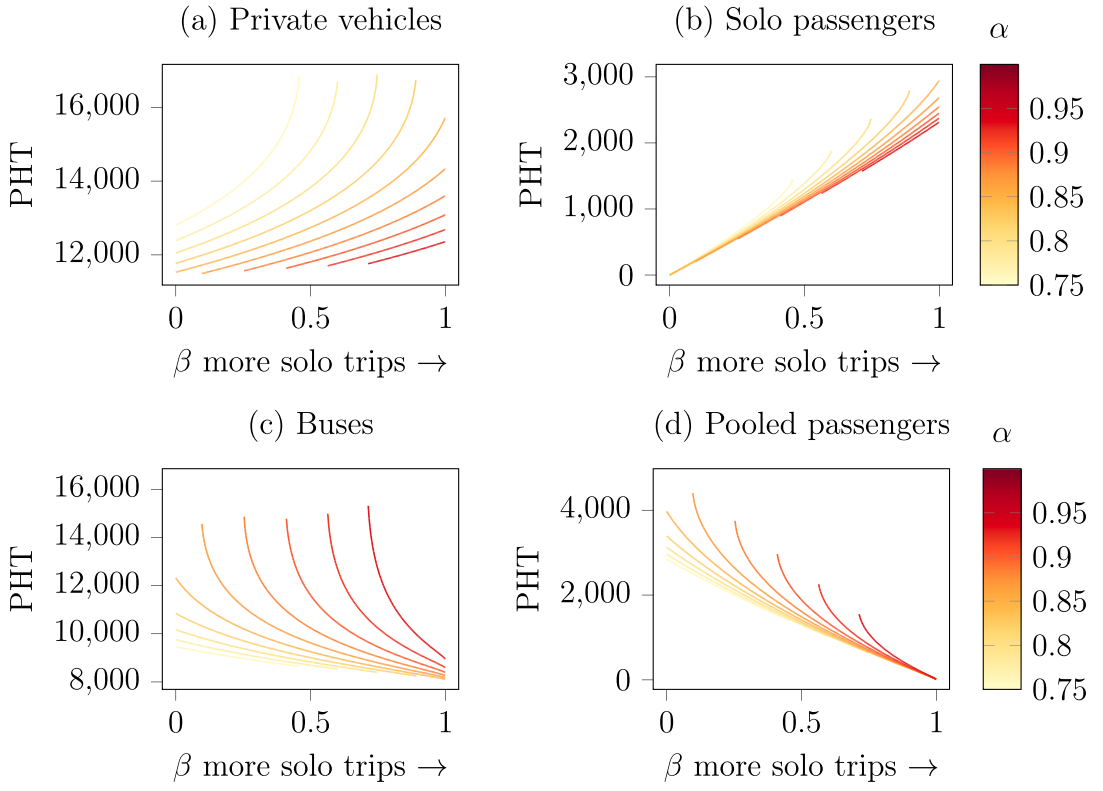


Fig. 7. PHT for every category of mode users in the network for the scenario $\{pv, s\}^V|\{p, b\}^B$, i.e., when pooling vehicles utilize the bus network.

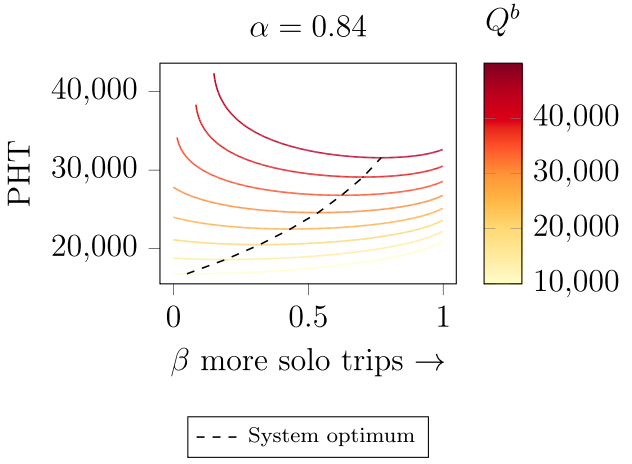


Fig. 8. PHT as function of β for different bus demand Q^b for the scenario $\{pv, s\}^V|\{p, b\}^B$, i.e., when pooling vehicles utilize the bus network.

function $P_{\log}(n) = -3.4n\log(\frac{n+40}{40}) + 40n$ in addition to the baseline polynomial function in Section 4 that we denote by P_{pol}. Because in this part we will restrict our analysis to the free-flow regime, we will only consider $n \in [0, n_{cr}]$, and can hence neglect the functions' behavior for larger accumulations. The choice of these functions is made such that they all have the same network capacity achieved for the same value of critical accumulation n_{cr}. Fig. 9 displays the speed and production curves for the three functions under assessment. We utilize these production curves to plot the graph of PHT as function of β for both the benchmark scenario and the proposed allocation strategy. The obtained results are displayed in Fig. 10. We note that, for the exponential and polynomial functions, the results are relatively identical given the similarity between the two production functions. Nevertheless, as the speed is generally higher when using an exponential function, the range of values of PHT is smaller compared to the baseline scenario where a polynomial function is used. A substantial difference can be nonetheless observed when comparing the baseline polynomial

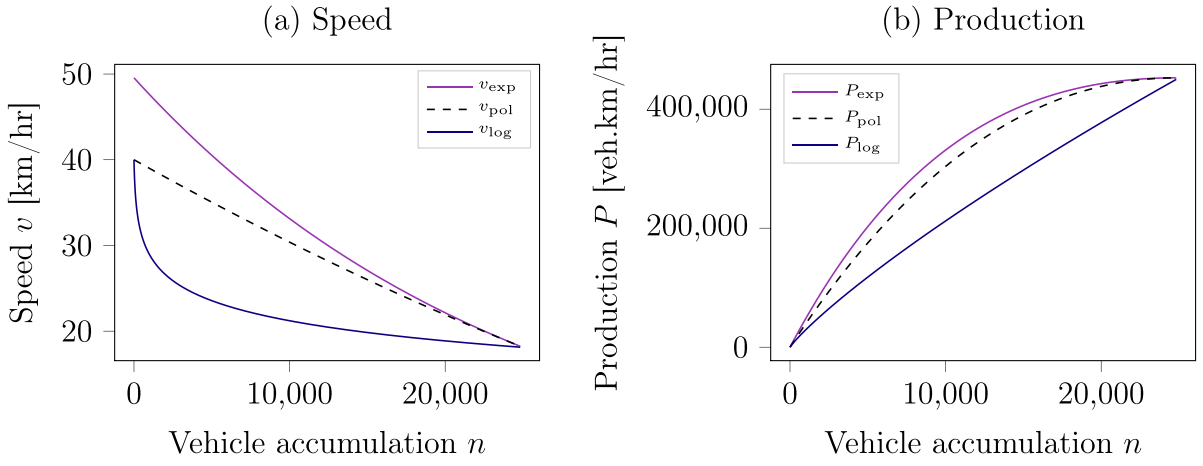


Fig. 9. Different speed and production functions.

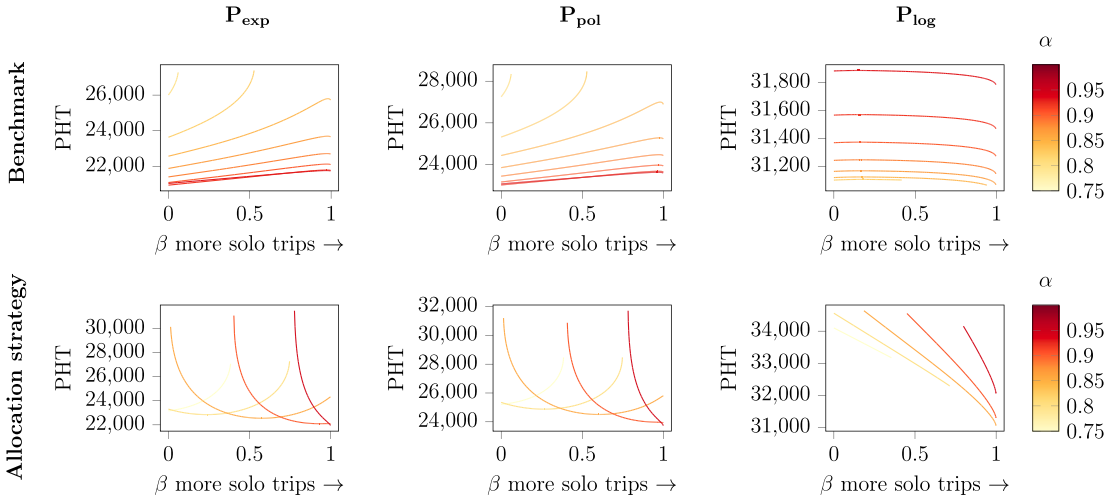


Fig. 10. PHT for the benchmark and allocation strategy for different production functions P.

function with the PHT graphs obtained for P_{log}. The first difference we note is that the PHT is relatively constant regardless of β, indicating that an increase in pooling has little influence on the network delays. This can be explained by looking at the logarithmic speed function v_{log} in Fig. 9(a) where the speed sharply decreases with accumulation before reaching a plateau. Given that the range of solutions occurs when the speed values are unvarying, little pooling benefit is derived. In other terms, having a lower fleet size does not necessarily cause a substantial improvement on speed. For the allocation strategy however, the differences are more accentuated. This is because adding more vehicles to the bus network causes a sharp decrease in speed, which explains why the minimum PHT always occurs when the entirety of the ride-hailing demand is using the vehicle network.

So far, we have assessed the total network delays by mainly assuming that the solo-pool demand split is an exogenously defined variable. Interestingly in almost all cases, the optimal value of β is far from the extreme values, i.e., $\beta = 0$ or $\beta = 1$, highlighting the need for carefully designed strategies or regulations to improve the mobility of the system. We know however that the pooling fare discount and the difference in pooled and solo trip time are the major factors determining the choice of ride-hailing users, and subsequently the value of this split. In the following subsection, we elaborate on how a pooling fare discount is a compelling tool that enables steering the split factor β towards better solutions.

4.3. Mode choice

To illustrate how the results of this work can potentially be utilized to drive the network into the system optimum, we replicate the choice of ride-hailing users for the following scenario $\{pv, s\}^{\mathcal{V}}|\{p, b\}^{\mathcal{B}}$, i.e., when pooling vehicles utilize the bus network. However, instead of β being a decision variable in our problem, we set it as the result of a mode choice model between solo and pool rides that we denote by β^{MC}. Therefore, β^{MC} itself is not only function of the travel time costs in the vehicle and bus network

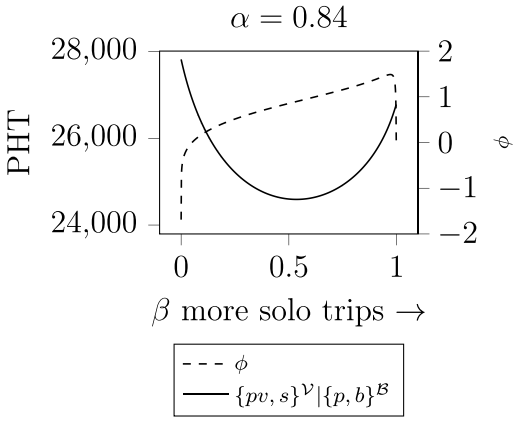


Fig. 11. PHT and pooling price discount for a fixed α.

but also of the pooling discount factor. The equation for β^{MC} is given by a standard logit model even if more complex models could also be introduced. However, given the lack of available data to calibrate these models, we consider that such a simplistic model could allow for a comparison of results between optimal β and users' choice to create some intuition. It follows that β^{MC} is computed using

$$\beta^{MC} = \frac{\exp\left(-\kappa(F^s + \mu t^s)\right)}{\exp\left(-\kappa(F^s + \mu t^s)\right) + \exp\left(-\kappa(\phi F^s + \mu t^p)\right)}, \tag{28}$$

with $t^s = \frac{\bar{l}}{v_{\mathcal{V}}}$ and $t^p = \frac{\bar{l}+\Delta l_p}{v_{\mathcal{B}}}$. In the above equation, t^s and t^p are the travel times for solo and pooled trips, both dependent on the value of β^{MC}. The constant $\kappa > 0$ is the binary mode choice scale parameter, μ is the monetary value of time, and F^s is the fare for solo rides. To encourage ride-hailing users to share their rides, the service operator introduces a discount factor to F^s that we denote here by ϕ. Combining the obtained results with the choice model, we resort to a numerical example to elucidate what should be this discount ϕ that drives the split towards β^* which is the point that minimizes the total PHT. The value ϕ^* in this case is the discount factor that yields a demand split naturally occurring at β^*.

To elaborate more on this approach, we take an example of $\alpha = 0.84$ and display in Fig. 11 the PHT when β is exogenous. We also show the pooling discount factor ϕ that we backcalculate to yield the same value of PHT given t^s and t^p using (28) for every value of β. This discount factor attains negative values when $\beta = 0$ because the speed in the bus network is too low when everyone is pooling. Therefore, despite the reduction in the travel distance due to the low detour incurred, the travel time remains very high. Consequently, passengers will favor traveling solo as the conditions in the vehicle network are relatively better, or else they should be awarded the appropriate monetary incentives to continue traveling in the bus network. As β increases, the discount factor becomes larger and its value exceeds one. At this point, the use of the bus network becomes privileged, and passengers who opt for a pooled trip are charged more compared to a solo trip. The drop in the discount factor that we observe when β approaches one is substantiated by the large decrease in the detour distance due to the low engagement in pooling, which results in a very high traveled distance even if the speed $v_{\mathcal{B}}$ is high.

The different solutions that we displayed so far assume that all the pooling vehicles utilize the bus network to perform their trips. However, this results in suboptimal solutions basically because we impose in advance the use of bus lanes by all vehicles irrespective of the speeds in the two networks. In the section below, we relax this assumption by only allowing a fraction of the pooling to utilize the bus lanes.

4.4. System optimum when restricting the number of pooled vehicles in the bus network

In this section, we investigate the total PHT for all users in the network in case where only a fraction of the total pooling vehicles utilizes the bus network. We refer to this scenario by the notation $\{pv, s, p\}^{\mathcal{V}}|\{p, b\}^{\mathcal{B}}$ to indicate that pooling vehicles drive on both networks. Also, we denote by γ the fraction of Q^p utilizing bus lanes. Therefore, finding the system optimum in the network is equivalent to finding β and γ that minimize

$$\underset{\beta\in[0,1],\gamma\in[0,1]}{\text{minimize}} \quad Q^{pv}\frac{\bar{l}}{v_{\mathcal{V}}} + \beta Q^{rh}\frac{\bar{l}}{v_{\mathcal{V}}} + (1-\gamma)(1-\beta)Q^{rh}\left(\frac{\bar{l}+\Delta l_p}{v_{\mathcal{V}}}\right) + \gamma(1-\beta)Q^{rh}\left(\frac{\bar{l}+\Delta l_p}{v_{\mathcal{B}}}\right) + Q^b\frac{\bar{l}_b}{v_b}. \tag{29}$$

When $\gamma = 1$, the solution becomes equivalent to that of finding the optimal β given that all pooled passengers use the bus network as in (24). In opposition, when $\gamma = 0$, we replicate the minimization problem given by (26) where the system behaves without any particular intervention from the network regulator. Finally, when $\gamma > 0$ and $\gamma < 1$, we emulate the scenario where only a fraction of pooling vehicles utilizes the bus network. We investigate these situations mainly due to two reasons: i) some users might still opt for

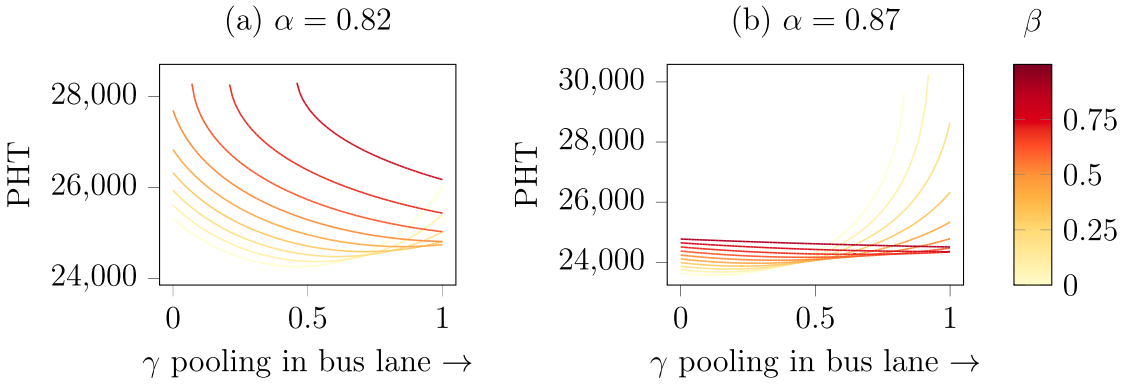


Fig. 12. Comparison of the $\{pv, s, p\}^{\mathcal{V}}|\{p, b\}^{\mathcal{B}}$ scenario for the two different values of α, i.e., when a fraction of pooling vehicles utilizes the bus network.

pooling without really having to use the bus network, and (ii) for the ride-splitting company to continue providing a door-to-door service, a portion of the pooled trip must be performed in the vehicle network especially when the value of α is relatively large, and the bus network does not cover the entirety of the network space. Pool vehicles therefore have to travel in the vehicle network for some part of their trips, particularly for the first and last mile.

Figs. 12(a) and 12(b) show the results of PHT as function of the fraction of pooling vehicles allowed to use the bus lanes for $\alpha = 0.82$ and $\alpha = 0.87$. For high values of β, i.e. when the demand for pooling is low, the value of γ that minimizes the PHT is equal to 1 because of the low engagement levels in pooling. If all pooled vehicles hence use the bus lanes, the disturbance to buses is naturally limited but the improvements to the vehicle network are significant. Nevertheless, when the demand for pooling is high, the system optimum is achieved for a value of $\gamma < 1$. This is because the delays for buses become the factor modulating the objective in this case. The same applies for $\alpha = 0.87$ except that for this case, the best solutions are achieved for high values of β yet lower γ since the potential of exploring the underutilized capacity in the bus lanes is restrained.

4.5. Mode choice with optional pooling in the bus network

In the following part, we replicate the same analysis as in Section 4.3 assuming nevertheless that only a fraction of the total pooling demand opts for traveling in the bus network. For this reason, we introduce two variables $\phi_{\mathcal{V}}$ and $\phi_{\mathcal{B}}$ to represent the two discount factors that the operator offers to pooling passengers traveling in the vehicle and bus network respectively. To model the choice of users between solo, pooling in the vehicle network, and pooling in the bus network, we resort to a multinomial nested logit. At the lower level, users who select to share their rides and travel in the bus network are represented by the fraction γ^{MC} where

$$\gamma^{MC} = \frac{\exp\left(-\kappa_p(\phi_{\mathcal{B}} F^s + \mu t^p_{\mathcal{B}})\right)}{\exp\left(-\kappa_p(\phi_{\mathcal{B}} F^s + \mu t^p_{\mathcal{B}})\right) + \exp\left(-\kappa_p(\phi_{\mathcal{V}} F^s + \mu t^p_{\mathcal{V}})\right)}. \tag{30}$$

At the upper level, the choice between solo and pool is given by

$$\beta^{MC} = \frac{\exp\left(-\kappa(F^s + \mu t^s)\right)}{\exp\left(-\kappa(F^s + \mu t^s)\right) + \exp\left(\frac{\kappa}{\kappa_p} \ln\left(\exp\left(-\kappa_b(\phi_{\mathcal{B}} F^s + \mu t^p_{\mathcal{B}})\right) + \exp\left(-\kappa_b(\phi_{\mathcal{V}} F^s + \mu t^p_{\mathcal{V}})\right)\right)\right)}, \tag{31}$$

where

$$t^p_{\mathcal{B}} = \frac{\bar{l} + \Delta l_p\left(\gamma^{MC}(1 - \beta^{MC})Q^{rh}\right)}{v_{\mathcal{B}}} \tag{32}$$

and

$$t^p_{\mathcal{V}} = \frac{\bar{l} + \Delta l_p\left((1 - \gamma^{MC})(1 - \beta^{MC})Q^{rh}\right)}{v_{\mathcal{V}}}. \tag{33}$$

The parameter κ_p is the scale parameter associated with the nest. Next, we follow the same approach as in Section 4.3 with the only difference that under the nested logit, we compute what should be the values of both $\phi_{\mathcal{V}}$ and $\phi_{\mathcal{B}}$ that regenerate all possible combinations for γ and β. The results for $\phi_{\mathcal{V}}$ and $\phi_{\mathcal{B}}$ as function of γ^{MC} for different values of β^{MC} are displayed in Fig. 13. The first point to note is that as speed in the bus network is higher in theory than that in the vehicle network, pooling passengers are granted a higher discount for them to accept the pooling option compared to the ones driving in the bus network. Moreover, in Fig. 13(a) as the value of γ^{MC} becomes close to 1, the discount factor becomes equal to 0 because the detour is naturally high for the pooling passengers driving in the vehicle network. The same logic applies to Fig. 13(b) for values of γ^{MC} approaching 0 as this means that no users are opting for pooling their rides in the bus network. The non-monotonicity observed in both graphs is simply the result of

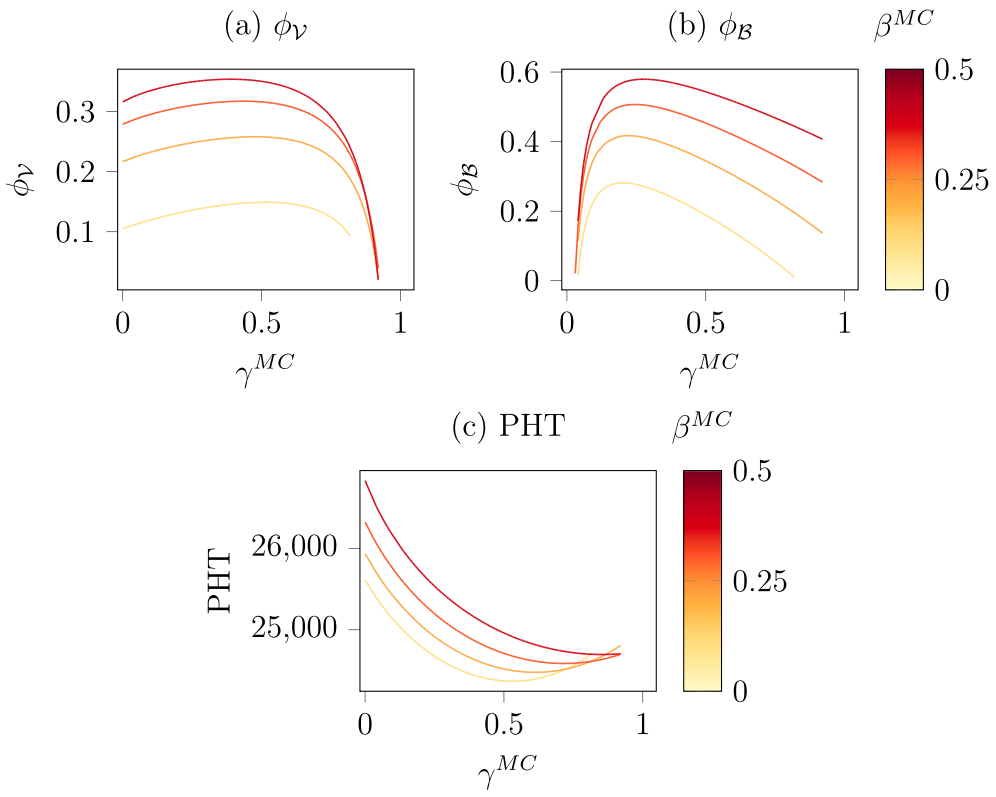


Fig. 13. The different discount factors ϕ_V and ϕ_B and PHT as function of γ^{MC} for $\alpha = 0.82$.

the interaction between two factors: the change in the network speed from one side, and the change in the detour distance from the other, both factors being the results of pooling passengers choosing between using the vehicle or bus network. These results however are highly dependent on the population under consideration. Moreover, the two discount factors that minimize the network delays are not necessarily that maximizing the profit of the ride-hailing operator. Consequently, further thoughts should be given to how to encourage ride-hailing platforms to offer appropriate service pricing and pooling discounts so that total network delays coincide with the system optimum observed in Fig. 13(c).

4.6. Comparison of the system optimum for different scenarios

In the following part, we compare the system optimum for the three different strategies tackled in this work: $\{pv, s, p\}^{\mathcal{V}}|\{b\}^{\mathcal{B}}$, $\{pv, s\}^{\mathcal{V}}|\{p, b\}^{\mathcal{B}}$, and $\{pv, s, p\}^{\mathcal{V}}|\{p, b\}^{\mathcal{B}}$ and for a predetermined network split α. The objective is to evaluate whether by relaxing the assumption that all pooled vehicles must use the bus network, we are able to achieve some improvements by bounding the amount of disruption allowed to buses. The results of the PHT for two different values of α are shown in Fig. 14.

If the fraction of passengers opting for a solo ride is high enough, we observe no improvement by allowing a fraction of the pooling demand to use the bus network, which explains the overlap between the full and dotted lines for large values of β in Fig. 14(a). The gap however starts increasing between the two lines when β is low. Consequently, reducing network delays by allowing pooled vehicles on bus lanes is beneficial if the number of pooled vehicles is capped. Otherwise, the delays caused to buses counterbalance all the advantages associated with pooling. The same analysis applies to Fig. 14(b) where α is set to 0.87 compared to 0.82 in the first case. Nevertheless, we observe here that the gap between the full and dotted lines shrank because of the lower space allocated to the bus network which makes it impossible to be utilized by high-occupancy vehicles.

5. Conclusion

In this paper, we analyze how, by giving pooled ride-hailing vehicles access to dedicated bus lanes, we can improve the performance of the transportation network under some specific settings. Our main methodological contributions lie in the following: (i) we propose a model for demand allocation for multi-modal transportation systems based on macroscopic fundamental diagrams, (ii) we examine the existence of an equilibrium between network supply and multi-modal demand and (iii) we point out the versatility of our model using a numerical example by assessing network delays for the solo-pool split ratio but also for the pooling fraction allowed on bus lanes. The results show that when the bus network is relatively large, assigning pooled drivers to bus lanes

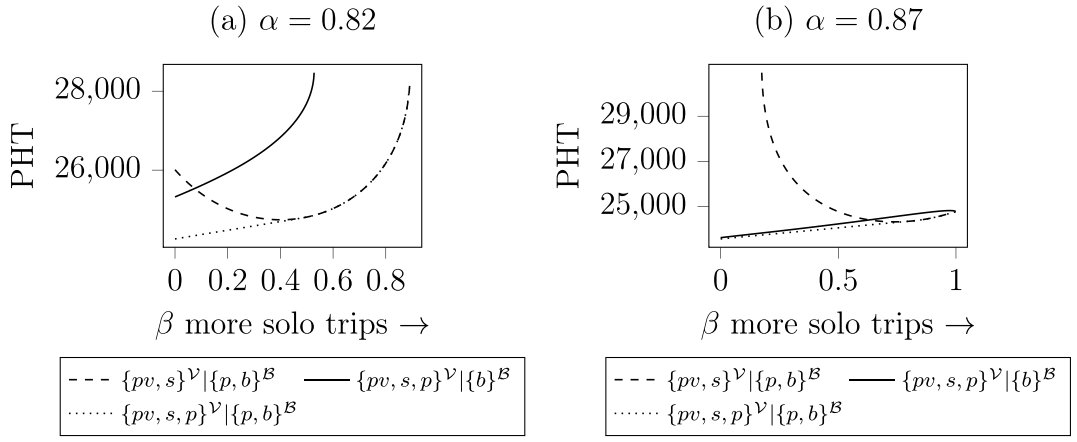


Fig. 14. Comparison of the results for three different strategies for two different values of α.

improves congestion in the vehicle network without causing large disturbances to the bus users in the dedicated bus lanes. When the fraction of pooling demand is high however, the deterioration in the bus network speed becomes more accentuated and an increase in the bus delays is inevitable. In this case, it becomes useful to restrict a portion of the pooled vehicles to travel on bus lanes to keep the bus network speed within acceptable bounds. Additionally, in this study, we assessed one simple pricing strategy that has the potential to drive the network towards its system optimum. Instead of it being an exogenous variable, we consider that the solo-pool demand split is the choice outcome of ride-hailing users. That being the case, we investigate the pooling discount factor that coincides with the system optimum. From the ride-hailing operators' point of view however, this discount does not necessarily maximize their profits. Taking into account the operators' objective requires hence proper modeling of the revenues which are dependent on F^s and ϕ, and the costs which are function of the fleet size N.

Practically, the allocation strategy we propose is easily implementable and effortlessly trackable because it does not require a change in the infrastructure, and is a time-independent modal space allocation strategy. Given the macroscopic feature of the network under consideration, it is possible to determine the recommended number of pool ride-hailing vehicles allowed to use the bus network. BLIP can be also incorporated within this framework in a hierarchical manner, where, after the upper layer determines the number of cars in the bus lane, a BLIP framework can be applied to further enhance the performance. In the future, we plan to investigate a relevant branch of research where regulators set the cap on the maximum number of pooled vehicles allowed on the bus network while operators decide on the pooling discount factor. Moreover, in the current work, we only considered a 2-passenger shared trip. We nevertheless plan to assess the potential of our policy to improve network delays when the service is extended to an on-demand micro-transit service where the capacity of vehicles exceeds two. Finally, investigating the problem in dynamic settings with time-dependent demands, and possibly fleet sizes, is also another research priority. This requires a proper simulation framework that we plan to create in attempt to assess the occupancy-dependent dynamic pricing strategies that we aim to develop.

CRediT authorship contribution statement

Lynn Fayed: Conceptualization, Methodology, Data analysis, Writing – review & editing. **Gustav Nilsson:** Conceptualization, Methodology, Writing – review & editing. **Nikolas Geroliminis:** Conceptualization, Methodology, Writing – review & editing.

Acknowledgments

This work was supported by the Swiss National Science Foundation under NCCR Automation, grant agreement 51NF40_180545.

Appendix. Driver and passenger detour

In the following section, we resort to simulated data to assess the usefulness of pooling in terms of traveled distance savings. The ultimate objective of the approach we undertake is to produce a driver and passenger detour functions Δl_d and Δl_p to utilize in our numerical assessment of the proposed space allocation policy. For this purpose, we make use of the set of taxi trips generated from real data in the Chinese city of Shenzhen by Beojone and Geroliminis (2021) to simulate the set of pooling requests that arrive to the ride-hailing platform.

Considering a scenario where only two requests are allowed to share their rides, enumeration of the possible trip combinations is straightforward. If i and j are the first and second passengers to be picked up respectively with origin nodes O_i and O_j, and destination nodes D_i and D_j, and if l_j is the distance O_jD_j, then the resulting trip possibilities consist of but are not limited to E_1 and E_2 as shown in Fig. 15. In E_2, both passengers experience a detour whereas in E_1, the second passenger to join the trip

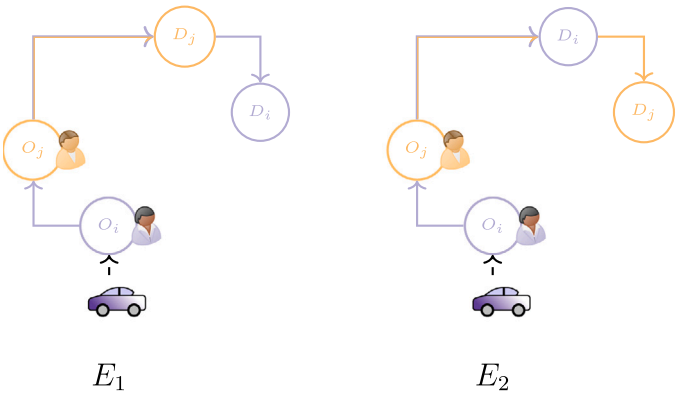


Fig. 15. Schematic of the two different types of pooling trips.

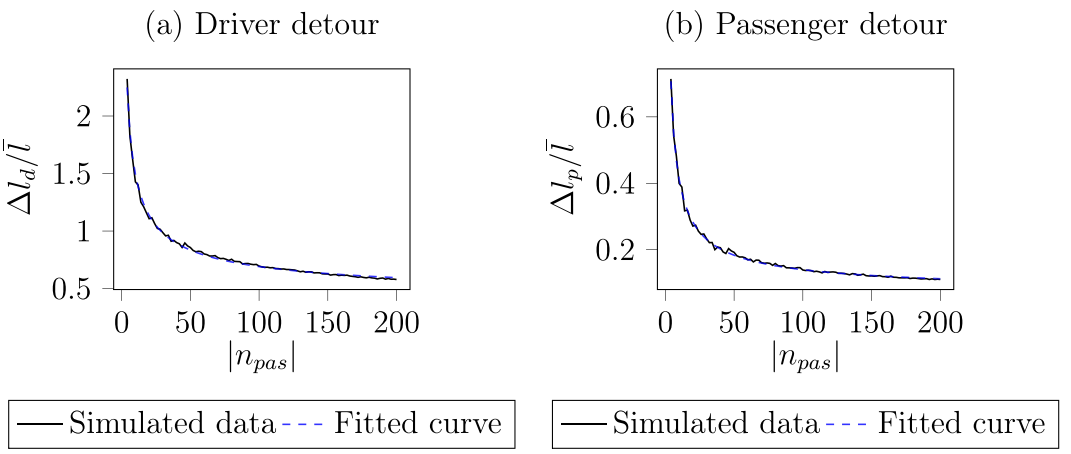


Fig. 16. Simulated data and fitted curve for the driver and passenger detour ratio.

is directly dropped off without any extra distance traveled. We note that the trip possibilities are in fact much larger in dynamic ride-splitting where requests arrive progressively to the platform and hence the sharing opportunities are more elaborate as in Jung et al. (2016). In this study however, we limit our analysis to the static case where every passenger shares the ride with at most one other passenger along the full trip.

If n_{pas} is the set of passengers willing to share their rides with $|n_{pas}|$ being an even number, we look for pooling trips that minimize the total distance traveled for drivers. To do so, we formulate the problem at hand as an LP with z_{ij} being the decision variable set to 1 if passenger i is assigned a pooled trip with passenger j, and to 0 otherwise. Accordingly, we formulate the objective and constraints of this problem as

$$\underset{z\in\{0,1\}^{n_{\text{pas}}\times n_{\text{pas}}}}{\text{minimize}} \quad \sum_i \sum_j c_{ij} z_{ij} \tag{34a}$$

$$\text{subject to} \quad \sum_k (z_{ik} + z_{ki}) = 1 \qquad \forall i \in n_{pas} \tag{34b}$$

$$z \in \{0,1\} \tag{34c}$$

where

$$c_{ij} = \begin{cases} \min(O_iO_j + l_j + D_jD_i, O_iO_j + O_jD_i + D_iD_j,) & \text{if } i \neq j \\ +\infty & \text{if } i = j. \end{cases} \tag{35}$$

The objective function (34a) minimizes the trip distance for all pooled trip possibilities with c_{ij} being the minimum distance over the two trip options E_1 and E_2 and is given by (35). Constraint (34b) guarantees that each passenger in the set of requests n_{pas} is assigned to exactly one pooled trip whereas constraint (34c) ensures that z is a binary variable.

The results for the average normalized driver and passenger detours as function of the number of pooled passengers are displayed in Fig. 16. Intuitively, the detour decreases with the number of passengers as the chances of finding an attractive pooling trip

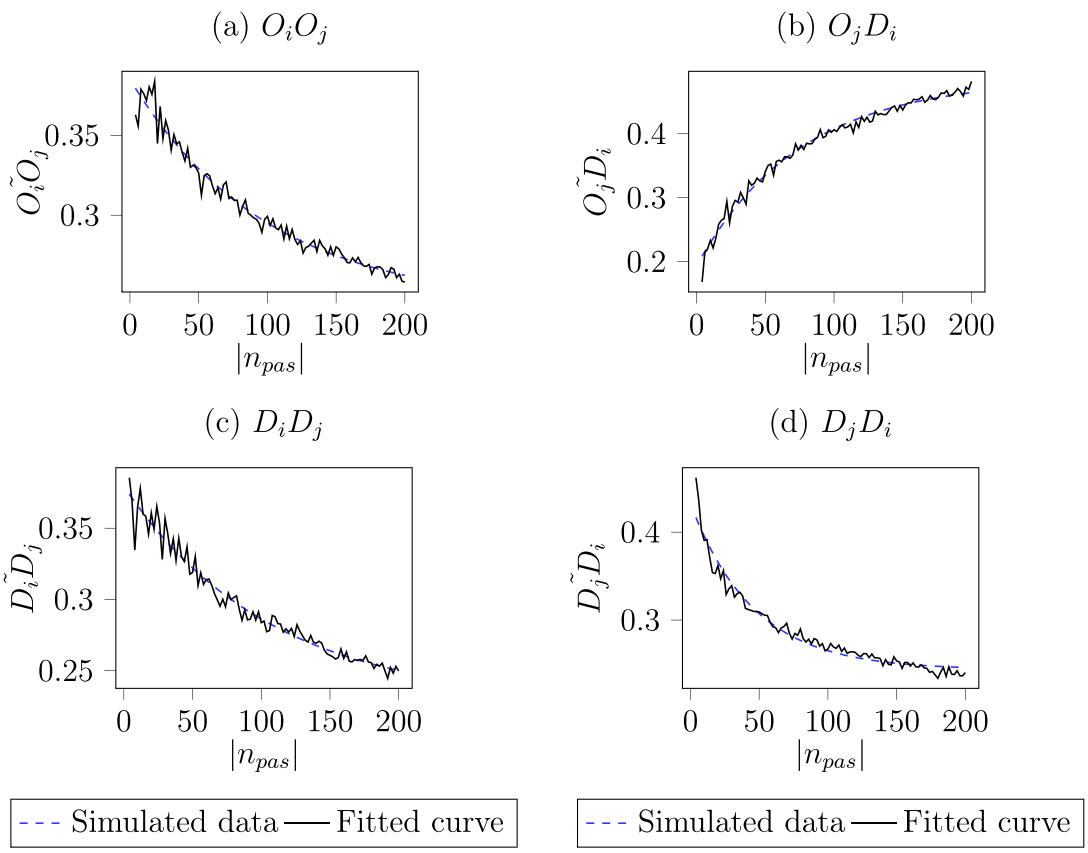


Fig. 17. Normalized distance traveled for different pooled trip segments.

increase. Because in the approach we follow, we do not set a bound on the allowable passenger detour, the total trip distance computed as $\bar{l} + \Delta l_d$ attains very large values when the number of pooling passengers is low.

In addition to the detour, it is interesting to look at the different normalized distance traveled for some of the segments of a pooled trip as function of the number of pooling passengers as shown in Fig. 17. Note that we display the normalized trip portions $O_i\tilde{O}_j$, $O_j\tilde{D}_i$, $D_i\tilde{D}_j$, and $D_j\tilde{D}_i$. The main conclusions that we derive here are that even if the pooled driver trip length decreases because of the smaller detour, the normalized segment of the trip $O_j\tilde{D}_i$ performed with two passengers inside the vehicle grows with the number of pooling passenger. As pooling becomes more efficient, passengers with both neighboring origins and destinations are pooled together, and the resulting trips are labeled as optimal on the network level. In our framework, this is an important observation from an implementation point of view because even if only ride-hailing vehicles with an occupancy of two are allowed in the bus network, the proposed policy has the potential to continue being favorable.

References

Alonso Gonzalez, M., Cats, O., Oort, N., Hoogendoorn-Lanser, S., Hoogendoorn, S., 2020. What are the determinants of the willingness to share rides in pooled on-demand services? Transportation http://dx.doi.org/10.1007/s11116-020-10110-2.

Beojone, C.V., Geroliminis, N., 2021. On the inefficiency of ride-sourcing services towards urban congestion. Transp. Res. C 124, 102890. http://dx.doi.org/10.1016/j.trc.2020.102890.

Bimpikis, K., Candogan, O., Saban, D., 2019. Spatial pricing in ride-sharing networks. Oper. Res. 67 (3), 744–769. http://dx.doi.org/10.1287/opre.2018.1800.

Cachon, G.P., Daniels, K.M., Lobel, R., 2017. The role of surge pricing on a service platform with self-scheduling capacity. Manuf. Service Oper. Manag. 19 (3), 368–384. http://dx.doi.org/10.1287/msom.2017.0618.

Castillo, J.C., Knoepfle, D., Weyl, G., 2017. Surge pricing solves the wild goose chase. In: Proceedings of the 2017 ACM Conference on Economics and Computation. EC '17, Association for Computing Machinery, New York, NY, USA, pp. 241–242. http://dx.doi.org/10.1145/3033274.3085098.

Cheng, Q., Liu, Z., Lin, Y., Zhou, X.S., 2021. An s-shaped three-parameter (S3) traffic stream model with consistent car following relationship. Transp. Res. B 153, 246–271. http://dx.doi.org/10.1016/j.trb.2021.09.004.

Cramer, J., Krueger, A.B., 2016. Disruptive change in the taxi business: The case of Uber. Amer. Econ. Rev. 106 (5), 177–182. http://dx.doi.org/10.1257/aer.p20161002.

Daganzo, C.F., Ouyang, Y., Yang, H., 2020. Analysis of ride-sharing with service time and detour guarantees. Transp. Res. B 140, 130–150. http://dx.doi.org/10.1016/j.trb.2020.07.005.

Eichler, M., Daganzo, C.F., 2006. Bus lanes with intermittent priority: Strategy formulae and an evaluation. Transp. Res. B 40 (9), 731–744. http://dx.doi.org/10.1016/j.trb.2005.10.001.

Erhardt, G., Roy, S., Cooper, D., Sana, B., Chen, M., Castiglione, J., 2019. Do transportation network companies decrease or increase congestion? Sci. Adv. 5, eaau2670. http://dx.doi.org/10.1126/sciadv.aau2670.

Feng, S., Duan, P., Ke, J., Yang, H., 2022. Coordinating ride-sourcing and public transport services with a reinforcement learning approach. Transp. Res. C 138, 103611. http://dx.doi.org/10.1016/j.trc.2022.103611.

Fu, H., Wang, Y., Tang, X., Zheng, N., Geroliminis, N., 2020. Empirical analysis of large-scale multimodal traffic with multi-sensor data. Transp. Res. C 118, 102725. http://dx.doi.org/10.1016/j.trc.2020.102725.

Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transp. Res. B 42 (9), 759–770. http://dx.doi.org/10.1016/j.trb.2008.02.002.

Geroliminis, N., Haddad, J., Ramezani, M., 2013. Optimal perimeter control for Two Urban Regions with macroscopic fundamental diagrams: A model predictive approach. IEEE Trans. Intell. Transp. Syst. 14, 348–359. http://dx.doi.org/10.1109/TITS.2012.2216877.

Geroliminis, N., Zheng, N., Ampountolas, K., 2014. A three-dimensional macroscopic fundamental diagram for mixed bi-modal urban networks. Transp. Res. C 42, 168–181. http://dx.doi.org/10.1016/j.trc.2014.03.004.

Hall, J.D., Palsson, C., Price, J., 2018. Is Uber a substitute or complement for public transit? J. Urban Econ. 108, 36–50. http://dx.doi.org/10.1016/j.jue.2018.09.003.

He, F., Shen, Z.-J.M., 2015. Modeling taxi services with smartphone-based e-hailing applications. Transp. Res. C 58, 93–106. http://dx.doi.org/10.1016/j.trc.2015.06.023.

Henao, A., Marshall, W., 2017. A framework for understanding the impacts of ridesourcing on transportation. In: Meyer, G., Shaheen, S. (Eds.), Disrupting Mobility: Impacts of Sharing Economy and Innovative Transportation on Cities. Springer International Publishing, Cham, pp. 197–209. http://dx.doi.org/10.1007/978-3-319-51602-8_13.

Jung, J., Jayakrishnan, R., Park, J.Y., 2016. Dynamic shared-taxi dispatch algorithm with hybrid-simulated annealing. Comput.-Aided Civ. Infrastruct. Eng. 31 (4), 275–291. http://dx.doi.org/10.1111/mice.12157.

Ke, J., Yang, H., Li, X., Wang, H., Ye, J., 2020a. Pricing and equilibrium in on-demand ride-pooling markets. Transp. Res. B 139, 411–431. http://dx.doi.org/10.1016/j.trb.2020.07.001.

Ke, J., Yang, H., Zheng, Z., 2020b. On ride-pooling and traffic congestion. Transp. Res. B 142, 213–231. http://dx.doi.org/10.1016/j.trb.2020.10.003.

Ke, J., Zheng, Z., Yang, H., Ye, J., 2021a. Data-driven analysis on matching probability, routing distance and detour distance in ride-pooling services. Transp. Res. C 124, 102922. http://dx.doi.org/10.1016/j.trc.2020.102922.

Ke, J., Zhu, Z., Yang, H., He, Q., 2021b. Equilibrium analyses and operational designs of a coupled market with substitutive and complementary ride-sourcing services to public transits. Transp. Res. E 148, 102236. http://dx.doi.org/10.1016/j.tre.2021.102236.

Lamotte, R., de Palma, A., Geroliminis, N., 2017. On the use of reservation-based autonomous vehicles for demand management. Transp. Res. B 99, 205–227. http://dx.doi.org/10.1016/j.trb.2017.01.003.

Lo, J., Morseman, S., 2018. The perfect uberPOOL: A case study on trade-offs. Ethnogr. Prax. Ind. Conf. Proc. 2018 (1), 195–223. http://dx.doi.org/10.1111/1559-8918.2018.01204.

Loder, A., Lukas, e., Menendez, M., Axhausen, K.W., 2017. Empirics of multi-modal traffic networks – using the 3D macroscopic fundamental diagram. Transp. Res. C 82, 88–101. http://dx.doi.org/10.1016/j.trc.2017.06.009.

Ma, S., Zheng, Y., Wolfson, O., 2015. Real-time city-scale taxi ridesharing. IEEE Trans. Knowl. Data Eng. 27, 1782–1795.

Mesbah, M., Sarvi, M., Currie, G., 2011. Optimization of transit priority in the transportation network using a genetic algorithm. IEEE Trans. Intell. Transp. Syst. 12 (3), 908–919. http://dx.doi.org/10.1109/TITS.2011.2144974.

Mo, D., Chen, X.M., Zhang, J., 2022. Modeling and managing mixed on-demand ride services of human-driven vehicles and autonomous vehicles. Transp. Res. B 157, 80–119. http://dx.doi.org/10.1016/j.trb.2022.01.003.

Ni, W., Cassidy, M., 2019. City-wide traffic control: Modeling impacts of cordon queues. Transp. Res. C 113, http://dx.doi.org/10.1016/j.trc.2019.04.024.

Nourinejad, M., Ramezani, M., 2020. Ride-sourcing modeling and pricing in non-equilibrium two-sided markets. Transp. Res. B 132, 340–357. http://dx.doi.org/10.1016/j.trb.2019.05.019, 23rd International Symposium on Transportation and Traffic Theory (ISTTT 23).

Rayle, L., Dai, D., Chan, N., Cervero, R., Shaheen, S., 2016. Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. Transp. Policy 45, 168–178. http://dx.doi.org/10.1016/j.tranpol.2015.10.004.

Schaller, B., 2018. The new automobility: Lyft, uber, and the future of American cities.

Shaheen, S., Cohen, A., 2019. Shared ride services in North America: definitions, impacts, and the future of pooling. Transp. Rev. 39 (4), 427–442. http://dx.doi.org/10.1080/01441647.2018.1497728.

Tirachini, A., 2020. Ride-hailing, travel behaviour and sustainable mobility: an international review. Transportation 47, http://dx.doi.org/10.1007/s11116-019-10070-2.

Tirachini, A., Gómez-Lobo, A., 2019. Does ride-hailing increase or decrease vehicle kilometers traveled (VKT)? A simulation approach for Santiago de Chile. Int. J. Sustain. Transp. 14, 1–18. http://dx.doi.org/10.1080/15568318.2018.1539146.

Tsitsokas, D., Kouvelas, A., Geroliminis, N., 2021. Modeling and optimization of dedicated bus lanes space allocation in large networks with dynamic congestion. Transp. Res. C 127, 103082. http://dx.doi.org/10.1016/j.trc.2021.103082.

Vignon, D.A., Yin, Y., Ke, J., 2021. Regulating ridesourcing services with product differentiation and congestion externality. Transp. Res. C 127, 103088. http://dx.doi.org/10.1016/j.trc.2021.103088.

Xu, Z., Chen, Z., Yin, Y., 2019. Equilibrium analysis of urban traffic networks with ride-sourcing services. Urban Econ. Regional Stud. EJournal.

Xu, Z., Yin, Y., Ye, J., 2020. On the supply curve of ride-hailing systems. Transp. Res. B 132, 29–43. http://dx.doi.org/10.1016/j.trb.2019.02.011, 23rd International Symposium on Transportation and Traffic Theory (ISTTT 23).

Yang, H., Leung, C.W., Wong, S., Bell, M.G., 2010. Equilibria of bilateral taxi–customer searching and meeting on networks. Transp. Res. B 44 (8), 1067–1083. http://dx.doi.org/10.1016/j.trb.2009.12.010.

Yu, J.J., Tang, C.S., Max Shen, Z.-J., Chen, X.M., 2020. A balancing act of regulating on-demand ride services. Manage. Sci. 66 (7), 2975–2992. http://dx.doi.org/10.1287/mnsc.2019.3351.

Zha, L., Yin, Y., Du, Y., 2018a. Surge pricing and labor supply in the ride-sourcing market. Transp. Res. B 117, 708–722. http://dx.doi.org/10.1016/j.trb.2017.09.010, TRB:ISTTT-22.

Zha, L., Yin, Y., Xu, Z., 2018b. Geometric matching and spatial pricing in ride-sourcing markets. Transp. Res. C 92, 58–75. http://dx.doi.org/10.1016/j.trc.2018.04.015.

Zha, L., Yin, Y., Yang, H., 2016. Economic analysis of ride-sourcing markets. Transp. Res. C 71, 249–266. http://dx.doi.org/10.1016/j.trc.2016.07.010.

Zhang, K., Nie, Y.M., 2021. To pool or not to pool: Equilibrium, pricing and regulation. Transp. Res. B 151, 59–90. http://dx.doi.org/10.1016/j.trb.2021.07.001.

Zheng, N., Geroliminis, N., 2013. On the distribution of urban road space for multimodal congested networks. Proc.- Soc. Behav. Sci. 80, 119–138. http://dx.doi.org/10.1016/j.sbspro.2013.05.009, 20th International Symposium on Transportation and Traffic Theory (ISTTT 2013).

Zhu, Z., Qin, X., Ke, J., Zheng, Z., Yang, H., 2020. Analysis of multi-modal commute behavior with feeding and competing ridesplitting services. Transp. Res. A 132, 713–727. http://dx.doi.org/10.1016/j.tra.2019.12.018.