Transportation Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Auction-Based Permit Allocation and Sharing System (A-PASS) for Travel Demand Management

Ruijie Li, Yu (Marco) Nie, Xiaobo Liu

informs
http://pubsonline.informs.org/journal/trsc

TRANSPORTATION SCIENCE
Articles in Advance, pp. 1–16
ISSN 0041-1655 (print), ISSN 1526-5447 (online)

Auction-Based Permit Allocation and Sharing System (A-PASS) for Travel Demand Management

Ruijie Li,[a,b] **Yu (Marco) Nie,**[c] **Xiaobo Liu**[a,b]

[a] School of Transportation and Logistics, Southwest Jiaotong University, Chengdu 610031, China; [b] National Engineering Laboratory of Integrated Transportation Big Data Application Technology, Southwest Jiaotong University, Chengdu 611756, China; [c] Department of Civil and Environmental Engineering, Northwestern University, Evanston, Illinois 60208

Contact: ruijie-li@foxmail.com (RL); y-nie@northwestern.edu, https://orcid.org/0000-0003-2083-470X (Y(M)N); xiaobo.liu@swjtu.cn, https://orcid.org/0000-0001-5722-4943 (XL)

Received: January 31, 2020
Revised: August 14, 2020; April 2, 2021; June 27, 2021
Accepted: July 29, 2021
Published Online in Articles in Advance: January 20, 2022

https://doi.org/10.1287/trsc.2021.1093

Abstract. This paper proposes a novel quantity-based demand management system that aims to promote ridesharing. The system sells a time-dependent permit to access a road facility (conceptualized as a bottleneck) by auction but encourages commuters to share permits with each other. The commuters may be assigned one of three roles: solo driver, ridesharing driver, or rider. At the core of this auction-based permit allocation and sharing system (A-PASS) is a trilateral matching problem (TMP) that matches permits, drivers, and riders. Formulated as an integer program, TMP is first shown to be tightly bounded by its linear relaxation. A pricing policy based on the classical Vickrey–Clarke–Groves (VCG) mechanism is then devised to determine the payment of each commuter. We prove that, under the VCG policy, different commuters pay exactly the same price as long as their role and access time are the same. Importantly, by controlling the number of shared rides, any deficit that may arise from the VCG policy can be eliminated. This may be achieved with a relatively small loss to system efficiency, thanks to the revenue generated from selling permits. Results of a numerical experiment suggest A-PASS strongly promotes ridesharing. As sharing increases, all stakeholders are better off: the ridesharing platform receives greater profits, the commuters enjoy higher utility, and society benefits from more efficient utilization of the road infrastructure.

Funding: This work was supported by the National Natural Science Foundation of China [Grants 72101217 and 71728001] and the U.S. National Science Foundation [Grant CMMI 1922665].
Supplemental Material: The online appendix is available at https://doi.org/10.1287/trsc.2021.1093.

Keywords: bottleneck • trilateral matching • auction • VCG mechanism • ridesharing

1. Introduction

The traditional travel demand–management approach either sets the price or controls the quantity of travel in order to persuade travelers to reduce the number of trips or switch to modes with greater average vehicle occupancy. The price-based approach, widely known as congestion pricing, eliminates excessive congestion by forcing travelers to pay a toll to make up for the discrepancy between their average and marginal travel costs (see, e.g., Pigou 1920, Small and Gómez-Ibáñez 1998). The quantity-based approach, on the other hand, directly chooses an "optimal" amount of traffic allowed to access a facility. The access may be distributed through reservation (Wong 1997), auction (Teodorović et al. 2008, Nie 2012), random draw (Wang, Yang, and Han 2010; Nie 2017), or tradable credits (e.g., Yang and Wang 2011, Nie and Yin 2013).

Promoted as a low-cost strategy to increase average vehicle occupancy, carpool gathered much interest in the late 1970s and early 1980s (Chan and Shaheen 2012, Shaheen and Cohen 2018). Yet, by 1990, the tide had long receded with the share of carpools in U.S. work trips declining from 19% in 1980 to about 13.4% in 1990 (Ferguson 1997). Transportation network companies (TNCs) have revived the enthusiasm for ridesharing as their technology dramatically reduces the cost of pairing riders and pricing rides (Shaheen and Cohen 2019). Santi et al. (2014) show that as much as 80% of the taxi trips in Manhattan could be shared at the expense of a modest increase in travel time. Strictly speaking, the e-hail service provided by most TNCs is not a form of "sharing" as the drivers are in it to make money. Yet TNCs do provide rides that are shared by passengers if only partially. Such a service is often called ride-pooling. Trends toward pooling seem strong in the TNC industry although one has to read some of the lofty claims[1] with caution. In addition, some TNCs have begun to offer true ridesharing services, such as Waze Carpool, Scoop, and RideAmigos (Shaheen and Cohen 2019). Although effective ridesharing can and should play a critical role in managing travel demand, it offers limited incentives to influence travel behaviors. Here, we propose to incentivize ridesharing by combining it with traditional demand-management tools, and we argue that integrating the two creates a more effective and balanced approach.

To showcase the idea, we devise an auction-based permit allocation and sharing system (A-PASS) in the context of ridesharing for the morning commute. In our setting, all commuters own a car and participate in ridesharing through A-PASS as either a solo driver, a ridesharing driver, or a rider. Travel is simplified as passing through a bottleneck with a limited capacity in the spirit of Vickrey (1969) and Arnott, de Palma, and Lindsey (1990). A-PASS aims to eliminate congestion by ensuring the number of commuters arriving at the bottleneck at any time never exceeds the capacity. It accomplishes this goal by instituting an access-by-permit rule. To use the bottleneck, commuters must either acquire a permit from an auction administered by A-PASS or ride with a driver to share a permit. A-PASS not only auctions out permits, but also *simultaneously* matches riders and drivers based on their reported preferences. As such, it solves a *trilateral matching problem* (TMP) that involves permits tied to a time slot, drivers, and riders in order to determine for each commuter (1) the ridesharing role, (2) the time slot, (3) the matching partner, and (4) the payment. Through this mechanism, A-PASS obtains both the resource (i.e., the revenue from selling permits) and the tools (pricing and matching) to achieve its goal: serving as many commuters as possible at a minimum societal cost (total congestion delay less social welfare).

There are two reasons why combining a permit auction with ridesharing creates a win–win solution. First and foremost, because the matching problem leads to a double auction, it is impossible to design a *desirable* pricing policy without running a deficit (Myerson and Satterthwaite 1983). By "desirable," we mean the policy simultaneously leads to efficient matching (*allocative efficiency* (AE)), makes the auction attractive to participants (*individual rationality* (IR)), and eliminates the incentive for lying about one's personal preferences (*incentive compatibility* (IC)). In fact, the prospect of running large deficits is an important reason why many auction schemes fail in practice. Consequently, the auctioneer is often forced to sacrifice efficiency in exchange for *budget balance* (BB). However, combining a permit auction with ridesharing promises a solution to the deficit problem because the revenue obtained from selling bottleneck permits can be used to subsidize ridesharing. The second reason is more technical and has to do with incentive compatibility. If the auction of the bottleneck capacity is separated from the assignment of ridesharing roles (i.e., permits and ridesharing are handled by different auctioneers), it is difficult, if not impossible, to prevent a traveler from lying about preferences in one auction for a greater gain in the other simply because neither auctioneer would learn the traveler's preferences to their full extent. Consequently, the inability to prevent such cheating is bound to lower allocative efficiency and reduce social welfare.

Our contributions to the literature are summarized as follows.

- Through A-PASS, we put forth a new idea for travel demand management that combines ridesharing, an auction, and quantity-based travel demand management.
- The trilateral matching problem created by A-PASS has not been studied in the literature. We formulate this problem as an integer program and show that it is tightly bounded by its linear relaxation thanks to the special constraint structure.
- We prove that the Vickrey–Clarke–Groves (VCG) pricing mechanism (Vickrey 1961, Clarke 1971, Groves 1973) ensures IR, AE, and IC. However, no pricing policy can simultaneously satisfy these three properties in a double auction without running a deficit. The VCG policy is no exception.
- To address the deficit problem, we exploit the trade-off between the deficit and the number of shared rides in trilateral matching. Specifically, we develop a numerical method that determines an optimal number of shared rides to achieve a desired revenue target (revenue neutral or maximization) using a revised VCG policy.
- A strict implementation of A-PASS requires creating a detailed matching table that specifies the role of each commuter, each driver–rider pair, and the access time for everyone. At first glance, such a requirement seems cumbersome, if not impractical. However, the applicability of A-PASS is not restricted by this requirement. Instead, it can be used simply as a tool to elicit travel preferences and to price the time-dependent access to a congested facility according to the ridesharing role. We show this is made possible by a property of the proposed pricing mechanism, which ensures the overall price a commuter has to pay does not vary from individual to individual, but rather is solely determined by the access time and their role in ridesharing.

For the remainder, Section 2 first reviews the related studies. Section 3 introduces A-PASS in the context of a morning commute, and Section 4 defines and formulates the trilateral matching problem. Sections 5 and 6 discuss, respectively, pricing policies and implementation issues. Results of numerical examples are reported and discussed in Section 6, and Section 7 summarizes the main findings and suggests topics for future research.

2. Related Studies

The proposed permit allocation and sharing scheme is closely related to quantity-based travel management and auction-based pricing for ridesharing. We focus on these topics.

2.1. Quantity-Based Travel Management

Directly controlling the access to a road facility can be achieved through various ways. A crude but widely embraced method takes away the access from a subset of drivers based on a random draw, typically according to the last digit of their vehicle's license plate (Cambridge Systematics 2007; Wang, Yang, and Han 2010; Nie 2017). More sophisticated schemes attempt to use the optimal load of the facility to guide the access control. When the demand is greater than the optimal load, the question is how to determine who should be granted what access. Several schemes have been considered in the literature.

Wong (1997) and Akahane and Kuwahara (1996) describe a highway reservation system that receives, evaluates, and accepts/rejects reservations for highway access. Drivers who are turned down are given the opportunity to resubmit alternative reservations. Through a simulation study, De Feijter, Evers, and Lodewijks (2004) show such a system reduces delays and improves travel reliability. Edara and Teodorović (2008) further operationalize the idea by implementing two subsystems, an off-line system that preallocates access to different classes of users and an online system that processes reservations in real time.

Reservation-based allocation does not ensure access is awarded to those who value it the most. This shortcoming can be addressed by selling it to the highest bidders. The transaction can be managed using two methods. The first method, widely known as a tradable credit scheme, first distributes access permits (credits) evenly to travelers, who then trade the permits with each other until a desirable allocation is achieved (e.g., Akamatsu, Sato, and Nguyen 2006; Yang and Wang 2011; Wang et al. 2012; Nie and Yin 2013; Akamatsu and Wada 2017). The focus of this study is the second method, which sells permits to travelers directly through auction.

Teodorović et al. (2008) propose an auction scheme that sets the maximum number of drivers allowed to enter a cordoned area. Drivers submit their sealed bids (the private price they are willing to pay to gain access) to the auctioneer, who grants access by solving a winner-determination problem. Wada and Akamatsu (2013) extend the auction scheme to a general network with multiple bottlenecks and origin–destination pairs. Drivers in such a network must choose a path that consists of multiple bottlenecks. The number of travelers allowed to pass each bottleneck is limited by its capacity, and drivers must acquire, through a combinatorial auction, a unique permit for each bottleneck on the path. The winner-determination problem is decomposed into two subproblems that are solved iteratively. The first subproblem allocates a certain number of permit bundles to each path, and in the second, the fixed permit bundles are sold to drivers through an ascending auction that ensures truthful reporting. Su and Park (2015) implement an auction scheme using a commercial traffic simulator. In their case study, drivers employ a "blind search" strategy to choose between a fallback option (a slow arterial road) and bidding for the permit to use a faster expressway. The winner-determination problem is solved using a heuristic that ranks all the submitted bids and approves feasible requests in descending order of the bids. Through a stated preference survey, Basar and Cetin (2017) found no "outright rejection" of the auction scheme among drivers who have experience with congestion pricing.

2.2. Pricing Ridesharing by Auction

The literature on ridesharing has grown substantially in recent years. The reader is referred to Agatz et al. (2012); Furuhata et al. (2013); and Mourad, Puchinger, and Chu (2019) for comprehensive reviews. Our focus here is auction-based pricing for ridesharing.

An auction is often used to solve matching problems arising from two-sided markets, such as assigning riders to drivers. When solving the winner-determination problem, a challenge is to ensure agents report their bids truthfully. The classical solution to this problem is to set the price based on the VCG mechanism (Vickrey 1961, Clarke 1971, Groves 1973). Hence, this type of pricing policy has been widely used in ridesharing. In the agent-based ridesharing (ABC) system proposed by Kamar and Horvitz (2009), agents report their private driving cost, and ABC computes a payment using a VCG pricing policy. Following Parkes, Kalagnanam, and Eso (2001), a budget-balance constraint is added into the winner-determination problem, which solves the deficit problem caused by the VCG policy at the expense of losing the insurance of truthful reporting. Kleiner, Nebel, and Ziparo (2011) allow agents to declare their role (rider or driver) and use an auction to rank and assign riders to each driver. To ensure truthful reporting, a second-price policy (i.e., the winner pays the price of the next bidder in the ranking) is employed. In Zhao et al. (2014), an agent's\valuation of a shared ride is computed based on the reported preferences for (1) departure time window, (2) number of available seats, and (3) private trip cost. To address the deficit problem, two revised pricing schemes are proposed: the first is a fixed payment scheme, and the second is a two-sided reserve pricing scheme. Although both schemes can achieve budget balance, their impact on the matching rate is not controlled. Zhao, Ramchurn, and Jennings (2015) further examine the pricing issue when agents may not be able to complete their trip because of uncertainty. They show that ensuring truthful reporting is much more difficult in this case. Zhang, Wen, and Zeng (2016) design a discounted trade reduction mechanism to ensure a high matching rate in ridesharing. Zhang, Wu, and Bei (2018) allow the drivers to impose a reserve price according to the origin–destination information and show it is an individually rational, incentive compatible, and computationally efficient mechanism. Balseiro et al. (2019) design a ridesharing pricing scheme applied over a finite horizon and impose a set of constraints to ensure periodic individual rationality, dynamic incentive compatibility, no positive transfers, and promise keeping. Their design objective, however, is to maximize the profit of the platform. Li, Nie, and Liu (2020) consider a ridesharing system in which participants price shared rides based on both operating cost and schedule displacement. To avoid deficits, they propose a single-side reward pricing policy, which only compensates participants who are forced to endure schedule displacement.

To summarize, the trade-off between deficits and other desired properties (AE, IC, and IR) is at the heart of the auction design for ridesharing. In practice, balancing the budget is typically achieved at the expense of allocative efficiency. By combining a permit auction with ridesharing, this study offers a unique and potentially win–win solution to this challenge.

3. Permit Allocation and Sharing System

In Section 3.1, we describe a permit allocation and sharing system for managing a morning commute. Because the travel permits are distributed using an auction, Section 3.2 explains how commuters value their bids for permits and how their utility is determined by the system's pricing policy.

3.1. Preliminaries

Consider a set of heterogeneous travelers $\mathcal{I} = \{1, 2, \cdots, i, \cdots, I\}$, who commute from home to a workplace via a bottleneck with a constant capacity c. Without loss of generality, we assume all travel occurs within a time window $\mathcal{T} = [0, T]$ and every commuter has a preferred arrival time $t \in \mathcal{T}$ (Arnott, de Palma, and Lindsey 1990). In this paper, we assume the preferred arrival time is heterogeneous. Given the finite capacity of the bottleneck, commuters cannot all arrive at their preferred arrival time t, and those whose arrival time is displaced bear a schedule cost proportional to the displacement. Because the desire for punctual arrival creates congestion in the form of queueing at the bottleneck, commuters are forced to make a trade-off between the schedule cost and queueing delay by adjusting their departure time. Eventually, this trade-off leads to a Nash equilibrium. However, such an equilibrium is not desirable because the queuing delay is a deadweight loss to the system. Vickrey (1969) suggests the queueing delay be completely eliminated by a time-varying toll. He shows that, faced with such a toll, commuters would depart at a rate exactly equal to the bottleneck capacity. Therefore, an alternative to Vickrey's toll is to enforce the number of commuters passing through the bottleneck not to exceed its capacity at any time. This can be achieved by dividing $\mathcal{T}$ into small intervals and selling permits by auction to match the capacity of each interval (Wang et al. 2018). In the following, we propose a new travel-management system that integrates such a permit auction scheme with ridesharing. Because the system effectively encourages drivers to share their permits with riders, it is called an A-PASS.

As illustrated in Figure 1, A-PASS first receives requests from commuters that detail their preferences. It then processes these requests through allocation and sharing functions. The former sells permits tied to each passing time slot to both solo and ridesharing drivers, effectively "allocating" them into each slot. The sharing function is responsible for matching ridesharing drivers to riders who are willing to purchase a seat from them. It is worth emphasizing that these functions are seamlessly integrated through A-PASS and executed simultaneously as a trilateral matching problem. We next explain how each component of A-PASS works.

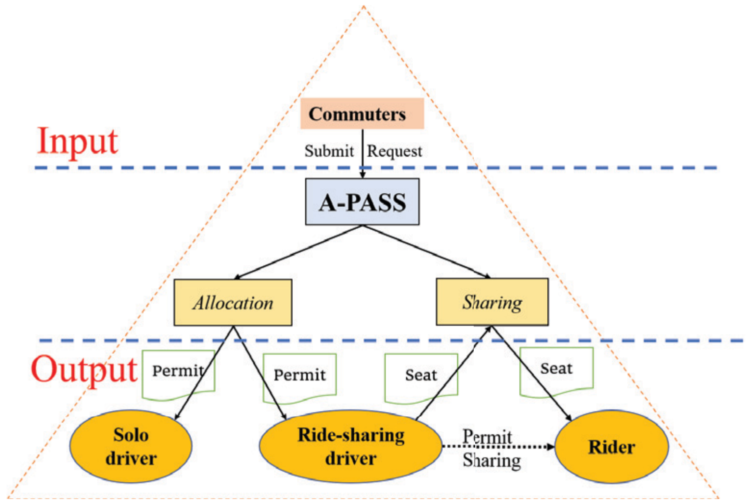
A-PASS divides the entire analysis period $\mathcal{T}$ into a set of discrete time slots $\mathcal{M} = \{0, 1, 2, \cdots, m, \cdots, M\}$ with equal lengths $\Delta t = \frac{T}{M+1}$. For each given time slot m, the number of permits sold is capped by $C = \lfloor c\Delta t \rfloor$, where $\lfloor a \rfloor$ is the largest integer less than or equal to a. We assume each time slot is short enough so that permit holders would arrive randomly within each slot, causing only negligible congestion.

To acquire the right to pass through the bottleneck at a prespecified time, each commuter must declare their request as follows:

Definition 1 (Commuter Request). The request of commuter $i \in \mathcal{I}$ consists of five pieces of private information regarding travel preferences: (i) the maximum willingness to pay (MWTP) for a bottleneck permit as a driver (λ_i), (ii) the desired price for sharing a seat in the vehicle with a rider (p_i), (iii) the MWTP for a seat in a shared ride (μ_i), (iv) the private cost per unit schedule displacement (b_i),[2] and (v) the preferred arrival time (t_i). Commuter i's request is written as a tuple $(\lambda_i, p_i, \mu_i, b_i, t_i)$.

To simplify the analysis, the following assumptions are introduced.

Figure 1. (Color online) Auction-Based Permit Allocation and Sharing System



Assumption 1. *Every commuter has a car and is willing to accept one of three possible roles—solo driver, ridesharing driver, or rider—assigned by A-PASS.*

Assumption 2. *Each driver can accept at most one rider and is paid by the rider at a price determined by A-PASS. Other than the payment, the costs associated with ridesharing (e.g., waiting, pick up, and drop off) are not explicitly modeled and are implicitly included in the commuter's MWTP.*

Assumption 3. *By restricting access, A-PASS ensures no congestion ever arises at the bottleneck. Hence, all commuters experience exactly the same travel time, which is normalized to zero.*

Assumptions 1 and 2 are fairly common in recent ridesharing studies (e.g., Xu et al. 2015; Liu and Li 2017; Wang, Ban, and Huang 2019), but Assumption 3 warrants a bit more explanation. Riders tend to value their in-vehicle time more highly because they can use it more productively (Ma and Zhang 2017, Zhong et al. 2020). Normalizing the free flow travel time to zero appears to wipe out this difference. We note that, however, the commuters in our model should be able to discern this difference because they report μ_i (the maximum price paid to get a seat) and p_i (the price of sharing a seat) separately.

When a decision is made based on user-reported preferences, an important concern has to do with the truthfulness of the reported information. Here, we assume that commuters are self-interested individuals who are eager to *independently* exploit any loopholes in the system.

Assumption 4. *Commuters have incentive to misreport their travel preferences if doing so benefits them, but they would not collude with others.*

3.2. Valuation and Utility

We are now ready to explain how A-PASS computes a commuter's valuation for a given role and a time slot based on the commuter's request. Let $\mathcal{K} = \{-1, 0, 1\}$ be the set of possible roles that a commuter can take on, where –1, 0, and 1 represent, respectively, rider, solo driver, and ridesharing driver.

Definition 2 (Valuation Policy). Suppose commuter i is assigned a role $k \in \mathcal{K}$ in a time slot m. The implicit bidding price, $v^k_{i,m}$, is defined as

$$v^k_{i,m} = \begin{cases} \lambda_i - b_i \varrho_{i,m} - p_i & k = 1 \\ \lambda_i - b_i \varrho_{i,m} & k = 0 \\ \mu_i - b_i \varrho_{i,m} & k = -1 \end{cases}, \ \forall i \in \mathcal{I}, m \in \mathcal{M}, \qquad (1)$$

where $\varrho_{i,m}$ represents the schedule displacement for commuter i passing through the bottleneck within time slot m.[3]

For solo drivers or riders, the price they are willing to pay is their MWTP less the schedule cost. Thus, the valuation of any slot m decreases as $\varrho_{i,m}$ and/or b_i increases. Everything else equal, commuters with a larger b_i are more likely to be granted a smaller displacement because it leads to a higher valuation. A ridesharing driver's valuation of the same slot is that of a solo driver less the price the driver expects to charge the rider. Clearly, when p_i is sufficiently large, the ridesharing driver would expect to be paid by the system to drive through the bottleneck. This valuation policy implies commuters value the same schedule displacement differently based on b_i. The policy also forbids a ridesharing driver from varying the seat price according to the time slot.

We proceed to define the utility of a commuter based on this valuation policy. To this end, we first define the charge levied by A-PASS on commuter i as

q_i. Note that q_i can be negative, which means a commuter could get paid. Further, we distinguish the valuations based on truthful reporting from others. Hereafter, a symbol with a bar header (i.e., $\bar{\cdot}$) is used to represent a variable associated with truthful reporting. Specifically, we define

$$\bar{v}^k_{i,m} = \begin{cases} \bar{\lambda}_i - \bar{b}_i\bar{\varrho}_{i,m} - \bar{p}_i & k=1 \\ \bar{\lambda}_i - \bar{b}_i\bar{\varrho}_{i,m} & k=0 \\ \bar{\mu}_i - \bar{b}_i\bar{\varrho}_{i,m} & k=-1 \end{cases}, \ \forall i \in \mathcal{I}, m \in \mathcal{M}, \tag{2}$$

as the true valuation of commuter i for role k and time slot m.

Definition 3 (Utility of a Commuter). If commuter i's request to use the bottleneck is turned down, commuter i's utility is zero. Otherwise, supposing commuter i is assigned a role k in time slot m, commuter i's utility is defined as commuter i's true valuation less the price paid to A-PASS for a seat or a permit, that is,

$$u^k_{i,m} = \bar{v}^k_{i,m} - q_i. \tag{3}$$

A-PASS may reject some requests when the bottleneck capacity is tight relative to the number of commuters. Commuters who are not awarded a bottleneck pass are assumed to have a fallback option with a utility of zero. This is a reasonable assumption as any bidder would naturally have a fallback option against which the bid (in our case, the four private parameters included in the request) is made. A bid is acceptable only if it yields a better utility than that of the fallback option, which is typically normalized to zero. The utility of a commuter whose request is accepted depends on a number of decisions made by A-PASS: the time slot and role and with whom the commuter is matched. We now turn to these decisions.

4. Trilateral Matching Problem

Upon receiving all requests, A-PASS must determine for all commuters: (1) their role $k \in \mathcal{K}$, (2) to which time slot m they are assigned, and (3) their rideshare partner if $k = 1$ or -1. The system makes these decisions by solving a trilateral matching problem that aims to maximize the social welfare (or the total valuations).

To represent the decisions, let z^k_i be the binary role assignment variable $z^k_i = 1$ if commuter i is assigned role k and zero otherwise, $x_{i,m}$ be the binary time slot assignment variable $x_{i,m} = 1$ if driver (solo or not) i is allocated into time slot m and zero otherwise, and $y_{i,j,m}$ represent the binary trilateral matching variable $y_{i,j,m} = 1$ if ridesharing driver i is matched with rider j and the pair is allocated into time slot m. We use $\boldsymbol{z}$, $\boldsymbol{x}$, and $\boldsymbol{y}$ to represent the corresponding vectors.

The TMP for A-PASS can be formulated as the following integer program:

$$\textbf{TMP} \quad \max \sum_{i\in\mathcal{I}}\sum_{m\in\mathcal{M}} v^0_{i,m}x_{i,m} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}}\sum_{m\in\mathcal{M}} \left(v^{-1}_{j,m} - p_i\right) y_{i,j,m} \tag{4a}$$

subject to:

$$\sum_{m\in\mathcal{M}} x_{i,m} \le 1, \ \forall i \in \mathcal{I}, \tag{4b}$$

$$\sum_{i\in\mathcal{I}}\sum_{m\in\mathcal{M}} y_{i,j,m} \le 1, \ \forall j \in \mathcal{I}, \tag{4c}$$

$$\sum_{j\in\mathcal{I}}\sum_{m\in\mathcal{M}} y_{i,j,m} \le 1, \ \forall i \in \mathcal{I}, \tag{4d}$$

$$\sum_{j\in\mathcal{I}} y_{i,j,m} \le x_{i,m}, \ \forall i \in \mathcal{I}, \ \forall m \in \mathcal{M}, \tag{4e}$$

$$\sum_{i\in\mathcal{I}} x_{i,m} \le C, \ \forall m \in \mathcal{M}, \tag{4f}$$

$$z^1_i = \sum_{j\in\mathcal{I}}\sum_{m\in\mathcal{M}} y_{i,j,m}, \ \forall i \in \mathcal{I}, \tag{4g}$$

$$z^{-1}_j = \sum_{i\in\mathcal{I}}\sum_{m\in\mathcal{M}} y_{i,j,m}, \ \forall j \in \mathcal{I}, \tag{4h}$$

$$z^0_i = \sum_{m\in\mathcal{M}} x_{i,m} - \sum_{j\in\mathcal{I}}\sum_{m\in\mathcal{M}} y_{i,j,m}, \ \forall i \in \mathcal{I}, \tag{4i}$$

$$\sum_{k\in\mathcal{K}} z^k_i \le 1, \ \forall i \in \mathcal{I}, \tag{4j}$$

$$z^k_i \in \{0,1\}, \ \forall i \in \mathcal{I}, \ \forall k \in \mathcal{K}, \tag{4k}$$

$$x_{i,m} \in \{0,1\}, \ \forall m \in \mathcal{M}, \ \forall i \in \mathcal{I}, \tag{4l}$$

$$y_{i,j,m} \in \{0,1\}, \ \forall m \in \mathcal{M}, \ \forall i \in \mathcal{I}, \ \forall j \in \mathcal{I}. \tag{4m}$$

Objective (4a) maximizes the total private valuations, which can also be interpreted either as the total social welfare or the system allocative efficiency. Constraint (4b) ensures that every commuter is allocated into no more than one time slot. Constraints (4c) and (4d) enforce one-to-one matching between riders and drivers. Constraint (4e) states that, if a driver is matched with a rider, they must be allocated into the same time slot. For a ridesharing driver, the inequality in this constraint becomes equality. Constraint (4f) requires the total number of permits sold not to exceed the bottleneck capacity. Constraints (4g)–(4j) specify the relation between role assignment variables $\boldsymbol{z}$, time slot assignment variables $\boldsymbol{x}$, and matching variables $\boldsymbol{y}$. Constraints (4k)–(4m) dictate that all decision variables are binary.

A general integer program such as Problem (4) is NP-hard. However, Problem (4) has a special structure that makes it relatively easy to solve because its linear relaxation generally provides high-quality lower bounds. We formally state this result as follows.

Proposition 1. *Construct a linear relaxation of* Problem (4) *as follows:*

RL − TMP

$$\max \sum_{i\in\mathcal{I}}\sum_{m\in\mathcal{M}} v^0_{i,m}x_{i,m} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}}\sum_{m\in\mathcal{M}}\left(v^{-1}_{j,m}\text{-}p_i\right)y_{i,j,m} \tag{5a}$$

subject to:

$$(4b) - (4j), \tag{5b}$$

$$0 \le z^k_i \le 1,\ \forall i \in \mathcal{I},\ \forall k \in \mathcal{K}, \tag{5c}$$

$$0 \le x_{i,m} \le 1,\ \forall m \in \mathcal{M},\ \forall i \in \mathcal{I}, \tag{5d}$$

$$0 \le y_{i,j,m} \le 1,\ \forall m \in \mathcal{M},\ \forall i \in \mathcal{I},\ \forall j \in \mathcal{I}. \tag{5e}$$

The optimal objective function value of Problem (4) *equals that of Problem* (5).

Proof. See Online Appendix B. □

Solving TMP gives optimal allocation and sharing decision vector $[x^*, y^*, z^*]$. Let $\tilde{\mathcal{I}}$ denote the set of all winning commuters in $[x^*, y^*, z^*]$. Accordingly, we define the *total system throughput* (i.e., the number of commuters allowed to pass the bottleneck within $\mathcal{T}$) as

$$Z = \sum_{i\in\mathcal{I}}\sum_{k\in\mathcal{K}} z^{k,*}_i, \tag{6}$$

and the maximum social welfare (the optimal objective function value of Problem (5)) as

$$V = \sum_{i\in\mathcal{I}}\sum_{m\in\mathcal{M}} v^0_{i,m}x^*_{i,m} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}}\sum_{m\in\mathcal{M}}\left(v^{-1}_{j,m} - p_i\right)y^*_{i,j,m}. \tag{7}$$

Further, V_{-i} is introduced to represent the maximum social welfare of the system when the request of commuter i is removed, $V_{-i,m-}$ denotes the maximum social welfare when the request of commuter i is removed and the capacity of time slot m decreases from C to $C-1$, and $V_{-ij,m-}$ denotes the maximum social welfare when the requests of commuter i and j are both removed and the capacity of time slot m decreases from C to $C-1$.

Note that we only need to solve TMP once to determine optimal allocation and sharing decisions. To price the participants—which must be done repeatedly as explained in the next section—we only need to evaluate the value V_{-i}, which can be obtained by solving the much easier LP relaxation.

5. Pricing Policies

Once the optimal permit allocation and sharing decisions are reached, A-PASS needs to decide how to price permits and seats. This price is set for each commuter as q_i, which specifies how much each commuter needs to pay A-PASS ($q_i \ge 0$) or be paid ($q_i < 0$). We denote a pricing policy as $\boldsymbol{q}$ and the profit of the system as

$$W = \sum_{i\in\tilde{\mathcal{I}}} q_i. \tag{8}$$

Before discussing various policies that the system can choose, we first explain what properties are desired for such a policy.

Definition 4 (Desired Properties of a Pricing Policy). A pricing policy q is said to be BB if $W = 0$ or weak budget balancing if $W \ge 0$; it is IR if $q_i \le v^k_{i,m}$ for each winning commuter $i \in \tilde{\mathcal{I}}$ and IC if no commuter can improve utility by misreporting the bid valuation.

The most straightforward policy is called the *first price policy*, denoted as $\boldsymbol{q}^-$. Using this policy, A-PASS simply sets the price to match each commuter's reported valuation as defined in Equation (1). The fact that $\boldsymbol{q}^-$ is IR is self-evident by inspecting the definition. It must be BB because V under the policy equals the profit, and it should always be nonnegative. To see why V can never be negative, note that, in the worst case, A-PASS can simply choose to do nothing, which would balance the budget (i.e., $V = 0$). The fact that any permit allocation and sharing occurs at all implies that $V \ge 0$.

The problem with the first price policy is that the system has no protection at all against speculative behaviors of commuters. If a significant portion of commuters misrepresent their preferences, the allocation of the permits would be severely distorted from the optimal solution. A standard approach to addressing this concern is to invoke the VCG policy, which may be simply described as a *second price policy*, originating from having the winner whose bid price is the highest pay the price offered by the second highest bidder in a single-round, sealed auction (Vickrey 1961, Clarke 1971, Groves 1973). In what follows, we show that the VCG policy works as intended in our setting. We denote the second pricing policy, or the VCG policy, as $\boldsymbol{q}^+$. Algorithm 1 describes an implementation of the policy in A-PASS. Our focus is to show that the implementation ensures truthful reporting in trilateral matching.

Algorithm 1 (Second Price Policy)
1: **Input:** The request $(\lambda_i, \mu_i, b_i, p_i, t_i)$ from each commuter $i \in \mathcal{I}$.
2: **Output:** Commuters winning travel permits within $\mathcal{T}$, that is, $[x^*, y^*, z^*]$ and a pricing policy $\boldsymbol{q}^+$.
3: **Trilateral matching problem:**
4: Solve Problem (4) to obtain V and $[x^*, y^*, z^*]$.
5: **Second price problem:**
6: **for** every winner $i \in \tilde{\mathcal{I}}$ **do**
7: Calculate V_{-i} by resolving Problem (5) without commuter i.
8: Set the bonus for commuter i as $\rho^+_i = V - V_{-i}$.

9: Set the second price on commuter i as $q_i^+ = v_{i,m}^k - \rho_i^+$.
10: **end for**
11: Return $[x^*, y^*, z^*]$ and q^+.

We first show that the second price policy is IR, which is straightforward.

Proposition 2. *The second price policy q^+ obtained from* Algorithm 1 *satisfies IR in* Definition 4.

Proof. It is easy to see that $V \geq V_{-i}$ because V_{-i} is the social welfare of a subset of the users included in the system that yields V. Hence, ρ_i^+ must be nonnegative as per Algorithm 1, which implies that $\rho_i^+ = v_{i,m}^k - q_i^+ \geq 0$. □

It should be noted that the second price policy does not automatically ensure truthful reporting (Conitzer and Sandholm 2006). Its applicability in our setting is not obvious because commuters' utility depends on their role. We separate the proof in several steps. Let us first propose and prove the following lemmas.

Lemma 1. *Suppose that commuter i reports a valuation as $(\lambda_i, \mu_i, b_i, p_i, t_i)$; then,*

i. *If the commuter is not allocated into any time slot, that is, $z_i^* = (z_i^0, z_i^1, z_i^{-1}) = (0,0,0)$, the commuter's utility is zero.*

ii. *If the commuter is allocated into time slot m as a solo driver, that is, $z_i^* = (1,0,0)$, the commuter's utility is*

$$u_{i,m}^0 = \bar{v}_{i,m}^0 + V_{-i,m-} - V_{-i}, \tag{9}$$

iii. *If the commuter is allocated into time slot m as a ridesharing driver and is matched with rider e, that is, $z_i^* = (0,1,0)$ and $z_e^* = (0,0,1)$, the commuter's utility is*

$$u_{i,m}^1 = \bar{v}_{i,m}^1 + v_{e,m}^{-1} + V_{-ie,m-} - V_{-i}, \tag{10}$$

iv. *If the commuter is allocated into time slot m as a rider and is matched with ridesharing driver f, that is, $z_i^* = (0,0,1)$ and $z_f^* = (0,1,0)$, the commuter's utility is*

$$u_{i,m}^{-1} = \bar{v}_{f,m}^1 + v_{i,m}^{-1} + V_{-fi,m-} - V_{-i}. \tag{11}$$

Proof. See Online Appendix C. □

Lemma 2. *Commuters cannot improve their utility by misreporting their request, whether they are designated as a solo driver, a ridesharing driver, or a rider when they act truthfully.*

Proof. See Online Appendix D. □

Lemma 3. *Commuters cannot improve their utility by misreporting their request if their truthful request would be rejected.*

Proof. See Online Appendix E. □

We are now ready to present the first main result.

Proposition 3. *The second price policy q^+ obtained from* Algorithm 1 *satisfies IC in Definition* 4.

Proof. Lemmas 2 and 3 list all four possibilities for a commuter's request: rejected or accepted as a rider/ridesharing driver/solo driver. Lemma 3 states that commuters cannot improve their utility if by truthful reporting their request would be rejected. Lemma 2 asserts the same is true if their request would be accepted. Thus, under no circumstance they can do better by misreporting. □

Our next main result asserts that, under the second price policy, once a commuter's role and time slot are fixed, the payment is also fixed. This feature is important because it ensures fairness; that is, nobody should be discriminated based on personal preferences. We formally state the result as follows.

Proposition 4. *Under the second price policy (Algorithm 1), different commuters pay exactly the same price as long as their role and time slot are the same.*

Proof. See Online Appendix F. □

A few remarks about Proposition 4 are in order here. First, it not only guarantees fairness, but also facilitates computation. In order to compute the bonus for a winner $i \in \tilde{\mathcal{I}}$, Algorithm 1 solves Problem (4) excluding i to evaluate V_{-i}. With Proposition 4, we only need to do this computation once for each role and time slot, which is a significant savings when the number of commuters is large. Second, Proposition 4 suggests that, when ridesharing is prohibited (i.e., all commuters solo drive), A-PASS under the second price policy is degraded to a time-dependent tolling scheme, much like what is proposed by Vickrey (1969). Finally, although alternative auction-based pricing schemes exist (Wada and Akamatsu 2013, Wang et al. 2018), none is demonstrated to satisfy this desired property.

6. Implementation Issues

The second price policy presented in the previous section encourages truthful reporting with a bonus. This could lead to large deficits, however. It is well known that, in a two-sided market such as considered herein, maximizing social welfare (i.e., optimal matching), ensuring truthful reporting, and balancing the budget cannot be achieved simultaneously (Myerson and Satterthwaite 1983). In Section 6.1, we propose a practical remedy to address the deficit problem. Section 6.2 shows A-PASS may be implemented as a pricing tool.

6.1. Deficit Control by a Budget Balance Constraint

Our idea is to trade social welfare for profit. To this end, note that the potential deficit comes from the need to compensate the two parties involved in ridesharing, that is, drivers and riders. Should we eliminate ridesharing altogether, A-PASS is reduced to a permit auction system, which can never yield a deficit. Let us define the number of ridesharing pairs as E, called the matching control. The feasible range of E is $[0, I/2]$ because there are no more than $I/2$ matched pairs with I commuters. Thus, a simple search in the feasible range yields an E^* that achieves the desired trade-off between profit and social welfare. For each given matching control E, the controlled trilateral matching problem can be reformulated as follows:

Controled TMP

$$\max \sum_{i\in\mathcal{I}}\sum_{m\in\mathcal{M}} v^0_{i,m}x_{i,m} + \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{I}}\sum_{m\in\mathcal{M}} \left(v^{-1}_{j,m} - p_i\right)y_{i,j,m} \tag{12a}$$

subject to:

$$(4b) - (4m), \tag{12b}$$

$$\sum_{i\in\mathcal{I}} z^{-1}_i \le E. \tag{12c}$$

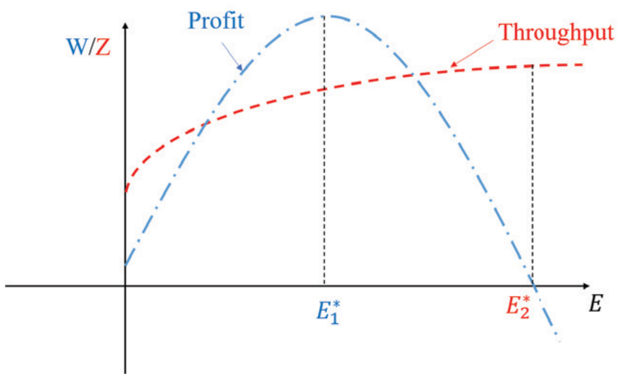
Because of the extra constraint (12c), Problem (12) always generates social welfare equal to or lower than that of Problem (5). Yet the extra constraint changes little the analytical property of the problem.

Proposition 5. *If E is an integer, the coefficient matrix of* Constraints (12b) *and* (12c) *satisfies total unimodularity.*

Proof. See Online Appendix G. □

This result asserts that the controlled TMP can also be solved by its linear relaxation. As E increases from zero to $I/2$, it is clear that both the social welfare (V) and the total system throughput (Z) increase monotonically. The implication for the profit (W) is less clear. We postulate that the relationship between W and E may be roughly depicted as a concave function illustrated in Figure 2. For a small E, the revenue generated from selling seats to riders is likely more than what is needed to keep them honest. This is because, when available seats are scarce, the competition for them would be strong enough to hold the trading price at a high level. As more seats become available (larger E), the profit keeps rising initially. When the decreasing price is finally unable to offset the loss from the need to compensate an increasing number of commuters, the profit peaks at E^*_1 in Figure 2. The exact location of E^* likely depends on inputs. The profit may never peak within the feasible range or peak early and plummet to the

Figure 2. (Color online) Illustration of the Relation Between the Matching Control E, the Profit, and the System Throughput



Note. E^*_1 maximizes the profit and E^*_2 maximizes the throughput while balancing the budget.

negative territory before reaching $I/2$ (E^*_2 marks the point at which the profit reaches zero).

In reality, the commuter requests received by A-PASS, including the number of requests and the information contained in each request, may vary from day to day. Yet the matching cannot be frequently adjusted according to the requests. Doing so implies commuters can influence the pricing policy itself by changing their requests, leading to a violation of truthful reporting.

Algorithm 2 (Matching Control Optimization Under Stochastic Demand)

1: **Input:** Distribution of all demand parameters: $(F_I, F_\lambda, F_p, F_\mu, F_b, F_t)$.
2: **Output:** Optimal matching control: E^*.
3: **Set the upper bound for E:** Calculate the maximum possible values for E as $E_{max} = F_I^{-1}(0.99)$.
4: **Sample demand:** Set the sample size S.
5: **Main iteration**:
6: **for** $e = 0 : \frac{E_{max}}{2}$ **do**
7: **for** every $s = 1 : S$ **do**
8: Draw a random demand N_s from F_I.
9: Generate a sample of N_s from each of the four distributions $F_\lambda, F_p, F_\mu, F_b, F_t$.
10: Set $E = e$ and solve the TMP (12) to get $\tilde{\mathcal{I}}$.
11: Determine $\mathbf{q}^+$ using Algorithm 1.
12: Set the profit $W_{e,s}$ using Equation (8) and the throughput $Z_{e,s}$ using Equation (6).
13: **end for**
14: Set $W_e = \frac{1}{S}\sum_{s=1}^{S} W_{e,s}$ and $Z_e = \frac{1}{S}\sum_{s=1}^{S} Z_{e,s}$.
15: **end for**
16: $E^*_W = \arg\max\{W_e,\ \forall e = 0, \cdots, \frac{E_{max}}{2}\}$ and $E^*_Z = \arg\max\{Z_e,\ \forall e = 0, \cdots, \frac{E_{max}}{2}\}$.

To account for the stochasticity in the demand process, we assume that A-PASS has access to distribution information about the key parameters. Such information may be collected gradually from the bidding process itself. Specifically, let $F_I, F_\lambda, F_p, F_\mu$, F_b, and F_t be, respectively, the distribution function for the number of commuters, the MWTP for a permit, the desired price for sharing a seat, the MWTP for a ride, the cost of schedule displacement, and the preferred arrival time. With such information, A-PASS can find the matching control that is optimal in a stochastic sense. The simplest implementation is to locate the matching control that delivers the best "expected" performance for a random sample drawn from the distributions. Algorithm 2 details this procedure.

6.2. A-PASS as a Pricing Tool

Under A-PASS, commuters not only have to acquire permit prior to travel, they must also accept the role assigned by the platform, including with whom to share the trip. Such an overly restrictive arrangement could face many practical challenges. However, A-PASS can be simply used as a tool to price the access to the bottleneck according to the ridesharing role. In such an application of A-PASS, the platform still needs to know the distributional information about the preferences of the travelers, which again may be collected from a bidding process. Thus, the platform still organizes the auction but would not provide explicit matching results. Rather, it only announces the price based on the role and access time. Thanks to Proposition 4, this price is determined solely by the role and access time, independent of the solution to the trilateral matching problem.

Algorithm 3 (Generate a Time- and Role-Specific Pricing Policy)

1: **Input:** Distribution of travelers' preferences: $(F_I, F_\lambda, F_p, F_\mu, F_b, F_t)$, and E^*.
2: **Output:** Expected pricing policy $\mathbf{Q} = \{Q_{m,k}\}$, where $Q_{m,k}$ represents the price for time slot $m \in \mathcal{T}$ and role $k \in \mathcal{K}$.
3: **Sample demand:** Set the sample size S.
4: **Main iteration**:
5: Set $Q_{m,k} = 0, \ \forall m, k$.
6: **for** every $s = 1 : S$ **do**
7: Draw a random demand N_s from F_I.
8: Generate a sample of N_s from each of the four distributions $F_\lambda, F_p, F_\mu, F_b, F_t$.
9: Solve the TMP (12) to get $\tilde{\mathcal{I}}$ and determine $\boldsymbol{q}_s^+$ using Algorithm 1.
10: Set $q^s_{m,k}$ as the price for time slot m and role k based on $\boldsymbol{q}_s^+$.
11: Set $Q_{m,k} = Q_{m,k} + q^s_{m,k}$.
12: **end for**
13: Set $Q_{m,k} = Q_{m,k}/S, \ \forall m, k$.

Supposing that $(F_I, F_\lambda, F_p, F_\mu, F_b, F_t)$ (the distribution of the parameters that define travelers' preferences) is given, a time- and role-specific pricing policy may be set using Algorithm 3. For each random sample drawn from the distributions, a TMP is solved, and a second price policy is generated. However, the price information is stored and averaged over the sample only for each time slot and role.

To apply such a pricing scheme, one does not have to worry about implementing the detailed matching results at all. Instead, it suffices to announce the price one has to pay at a given access time in a given role. For example, the policy may dictate that between 8:00 a.m. and 8:05 a.m., a solo driver would pay \$10 for the permit, a ridesharing driver would receive \$4, and a ridesharing rider would pay \$15. With this information, there is no need to explicitly acquire a permit prior and commuters are left with the decision to find a ridesharing partner (or not) assisted by and executed through the platform.

7. Numerical Experiments

In this section, we first illustrate the trilateral matching and pricing schemes of A-PASS using a small example. Then, a large-scale simulation experiment is conducted. The simulation results highlight the tradeoff between the system throughput and profit of A-PASS. All numerical results are obtained on a laptop with Inter(R) E5-1620 v4 CPU at 3.50 GHz and 32 G RAM. The TMP is solved using intlinprog solver in Gurobi.

7.1. Illustrative Example

Consider a bottleneck with a discrete capacity $c = 1$. All travel occurs within the time window $\mathcal{T} = [0, 2]$ and the set of discrete departure time intervals $\mathcal{M} = \{0, 1\}$. The schedule deviation function is $\varrho_{i,m} = |t_i - m|$. Thus, a commuter passing through the bottleneck at $m = 1$ should have a displacement of one time interval. Suppose A-PASS only receives requests from four commuters as detailed in Table 1.

We first consider the case in which all commuters report truthfully. Commuter 1's bidding price as a solo driver is $\bar{v}^0_{1,0} = \bar{\lambda}_1 - \bar{b}_1\bar{\varrho}_{1,0} = 5 - 2 \times |0 - 0| = 5$ and $\bar{v}^0_{1,1} = \bar{\lambda}_1 - \bar{b}_1\bar{\varrho}_{1,1} = 5 - 2 \times |0 - 1| = 3$. Because the desired ridesharing price is $\bar{p}_1 = 4$, the truthful bidding price as a ridesharing driver is $\bar{v}^1_{1,0} = \bar{v}^0_{1,1} - \bar{p}_1 = 5 - 4 = 1$ and $\bar{v}^1_{1,1} = \bar{v}^0_{1,1} - \bar{p}_1 = 3 - 4 = -1$. As a rider the commuter is willing to pay $\bar{v}^{-1}_{1,0} = \bar{\mu}_1 - \bar{b}_1\bar{\varrho}_{1,0} = 14 - 2 \times |0 - 0| = 14$ and $\bar{v}^{-1}_{1,1} = \bar{\mu}_1 - \bar{b}_1\bar{\varrho}_{1,1} = 14 - 2 \times |0 - 1| = 12$. The valuation of other commuters can be computed similarly and are omitted here for brevity.

Solving Problem (4) yields social welfare $V = 29$, and the permit allocation and sharing decisions are (i) commuter 2 as a ridesharing driver is matched with commuter 1 as a rider, and they are both allocated in the first time slot, and (ii) commuter 3 as a

Table 1. Commuter Requests in the Illustrative Example

Notation	Truthful reporting				Commuter 1 Misreporting			
	Commuter 1	Commuter 2	Commuter 3	Commuter 4	Case 1	Case 2	Case 3	Case 4
$\lambda_i/\bar{\lambda}_i$	5	5	6	4	5	5	14	5
$\mu_i/\bar{\mu}_i$	14	8	12	15	10	14	14	14
$b_i/\bar{b}_i$	2	3	1	2	2	1	2	2
$p_i/\bar{p}_i$	4	6	2	5	4	4	10	4
$t_i/\bar{t}_i$	0	0	0	0	0	0	0	1

ridesharing driver is matched commuter 4 as a rider, and they are both allocated in the second time slot. When commuter 1 is removed, the social welfare is $V_{-1} = 21$. According to Algorithm 1, thus, the bonus is $\bar{\rho}_1^+ = V - V_{-1} = 8$, and so the commuter pays $\bar{q}_1^+ = \bar{v}_{1,0}^{-1} - \bar{\rho}_1^+ = 14 - 8 = 6$ for the ride. The utility equals the bonus, that is, eight. The results associated with different reported valuations are summarized in Table 2.

Suppose now commuter 1 tries to lower the MWTP for a seat from $\bar{\mu}_1 = 14$ to $\mu_1 = 10$ (row 4 in Table 2). With this request, commuter 1 is a ridesharing driver instead of a rider. This reduces the social welfare by four, which, in turn, reduces the bonus accordingly as computed by the second price policy. Eventually, the commuter is worse off with a positive utility of four compared with eight when reporting truthfully.

If commuter 1 attempts to misrepresent the sensitivity to schedule displacement as one instead of two (row 5 in Table 2), commuter 1 is still a rider but is moved to the second time slot. In this case, both the social welfare and the bonus increase by one. The payment is reduced to four, but the utility remains at eight because their true valuation is down from 14 to 12.

If commuter 1 overstates the valuation of both MWTP for a permit and the price of a shared ride (i.e., commuter 1 tries to use aggressive seat pricing to gain the best time slot; row 6 in Table 2), commuter 1 becomes a solo driver in the second time slot. Commuter 1 ends up with a utility of one, the lowest of all strategies.

Table 2. Commuter 1's Utility Associated with Different Bidding Strategies

Reported valuation $(\lambda_1, \mu_1, b_1, p_1, t_1)$	Role $(z_1^{0,*}, z_1^{1,*}, z_1^{-1,*})$	Time slot m	Second pricing policy						
			V	V_{-1}	ρ_1^+	$v_{1,m}^k$	q_1^+	$\bar{v}_{1,m}^k$	$u_{1,m}^k$
(5,14,2,4,0)	(0,0,1)	0	29	21	8	14	6	14	8
(5,10,2,4,0)	(0,1,0)	0	25	21	4	1	−3	1	4
(5,14,1,4,0)	(0,0,1)	1	30	21	9	13	4	12	8
(14,14,2,10,0)	(1,0,0)	1	31	21	10	12	2	3	1
(5,14,2,4,1)	(0,0,1)	1	31	21	10	14	4	12	8

If commuter 1 attempts to misrepresent the desired arrival time as one (row 7 in Table 2), commuter 1 will still be a rider and is assigned to the second time slot. Thus, the consequence would be the same as if commuter 1 misreports the sensitivity to schedule displacement as one.

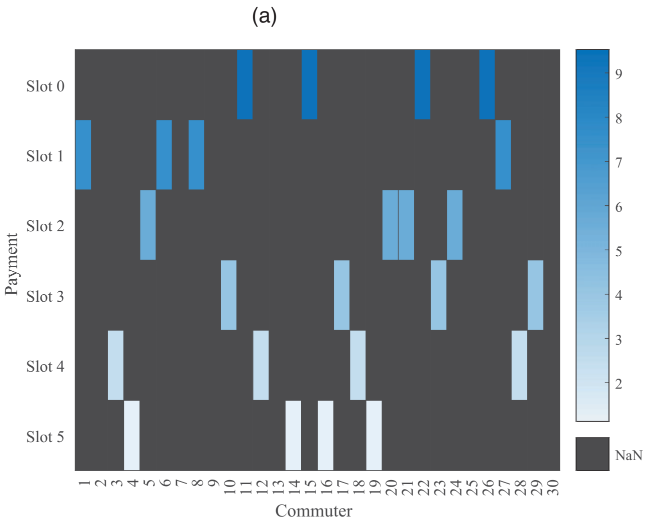
This analysis is not exhaustive but clearly illustrates why misreporting is not going to help anyone under the second price policy. However, this policy is expensive to implement. The reader can verify that, when all commuters truthfully report, the second price bonuses for commuters 1–4 are 8, 7, 11, and 9, respectively. This leads to a profit of A-PASS at $V - \sum_{i \in \bar{\mathcal{I}}} \rho_i^+ = 29 - 8 - 7 - 11 - 9 = -6$.

As explained in Section 5.2, we can eliminate deficit by implementing a matching control. If we set $E = 1$ and solve Problem (12), the optimal permit allocation and sharing decisions are (i) commuter 3 as a ridesharing driver matched with commuter 2, and they are both assigned to the first time slot; (ii) commuter 1 is assigned to the second time slot as a solo driver; and (iii) commuter 2's request is rejected. In this case, the social welfare decreases from 28 to 22, but the bonus required for truthful reporting is reduced to one, four, and two for commuters 1, 3, and 4, respectively. Consequently, the profit of A-PASS increases from −6 to 15.

7.2. Simulation Experiment

In this experiment, we evaluate the performance of A-PASS from the perspective of both the commuters and the system. For commuters, the performance metrics include the payment under the second price policy and the utility. The system performance metrics are the profit W and the throughput Z. Unless otherwise specified, we consider a bottleneck with a capacity c = 4 and an analysis period from zero to T = 6. The desired arrival time is set to $t^* = 0$ for all commuters in order to mimic the situation in rush hour when the time to pass the bottleneck is largely driven by the start time at work. All four parameters in the request are assumed to follow a normal distribution. Specifically, $F_\lambda = \mathbf{N}(10,1)$, $F_\mu = \mathbf{N}(25,2)$, $F_b = \mathbf{N}(1.5,0.5)$ and $F_p \sim \mathbf{N}(6,1)$. By default, the sample size S = 100, and

Figure 3. (Color online) Payment and Utility of Commuters Under the Second Price Policy



Note. $E = 0$, that is, permit sharing is banned, and all commuters are solo drivers.

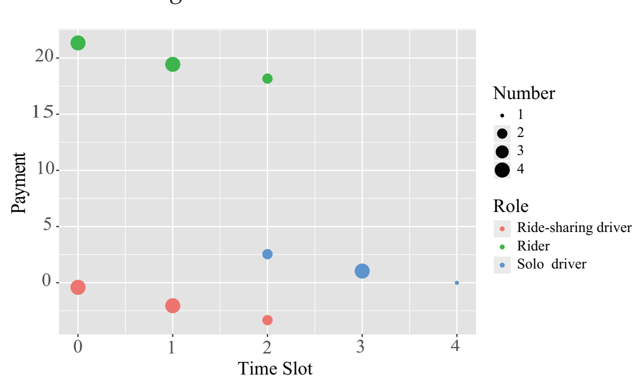
the average performance metrics are reported. We fix the number of commuters I for simplicity. In the benchmark case, $I = 30$, and the length of each discrete interval is set to one. Hence, $\mathcal{M} = [0, 1, 2, 3, 4, 5]$, and $C = 4$.

We first consider the case when $E = 0$, which effectively bans permit sharing and reduces the system to a permit auction system. Figure 3 reports the payment and utility of commuters under second price policy from a single run. Figure 3(a) reveals a couple of interesting patterns. First, as expected, the payment for a permit decreases as the time slot deviates further away from the arrival time desired by everyone (i.e., $t = 0$). The furthest time slot is worthless because the payment for this time slot is close to zero. Second and more important, all commuters allocated into the same time slot are charged exactly the same price for the permit, which verifies that the permit auction is effectively equivalent to an anonymous step tolling scheme as asserted by Proposition 4.

Figure 3(b) shows that, in general, a commuter's utility increases under second pricing when the commuter is allocated to a time slot further away from t^*. Not all commuters in the same slot have the same utility because of heterogeneity. In this case, those traveling closer to their desired time are worse off likely because of the intensive competition for t^*.

We next set the matching control $E = 10$. Figure 4 shows how payments aggregated for different roles from a single run vary over different time slots. In the plot, the location of a circle represents time slot (x) and payment (y). Its color and size indicate the role and the number of a commuter at the location, respectively. First, we note that commuters serving the same role in the same slot make exactly the same payment. For example, all four riders in time slot 0 pay the same price 21.865. Again, this confirms the assertion in Proposition 4. Second, for each role, the payment decreases as the time slot moves away from t^*, consistent with the finding from Figure 3(a). Third, in general, a rider pays the highest price, and a ridesharing driver pays the lowest. This is expected because the ridesharing driver has to cover the vehicle operating cost as well as the inconvenience cost associated with ride sharing. Finally, solo drivers are pushed away from t^*. Also noted in Liu and Li (2017), this phenomenon results from the fact that the willingness to pay of a solo driver for a highly desired slot is likely lower than that of a pair of two commuters. Hence, in general, solo drivers tend to concede more competitive slots to ridesharing partners.

Figure 4. (Color online) Payment Aggregated for Different Roles from a Single Run



Note. Matching control $E = 10$.

Figure 5. (Color online) Average Utility of Different Roles as a Function of Time Slot

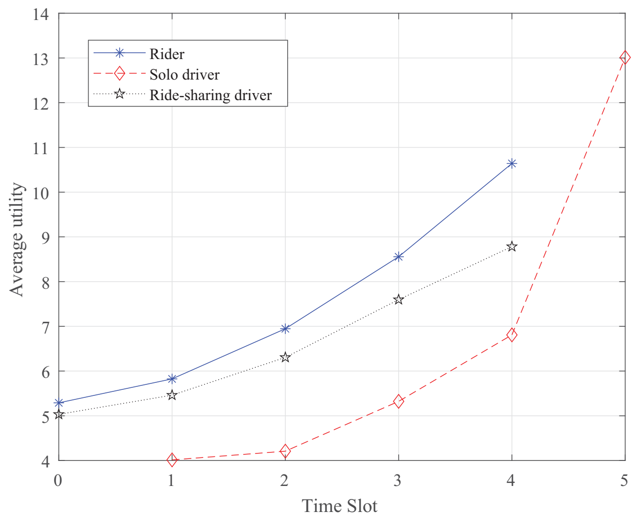


Figure 5 compares the average utility of each role in the different time slots. The results confirm that, on average, the utility of each role increases with schedule displacement, consistent with the finding from Figure 3(b). Furthermore, when they are placed in the same slot, solo drivers have significantly lower utility compared with ridesharing drivers and riders. This finding is strong evidence that A-PASS promotes ride sharing. The utility of ridesharing drivers and riders is comparable, but the riders consistently outperform the drivers with a small margin.

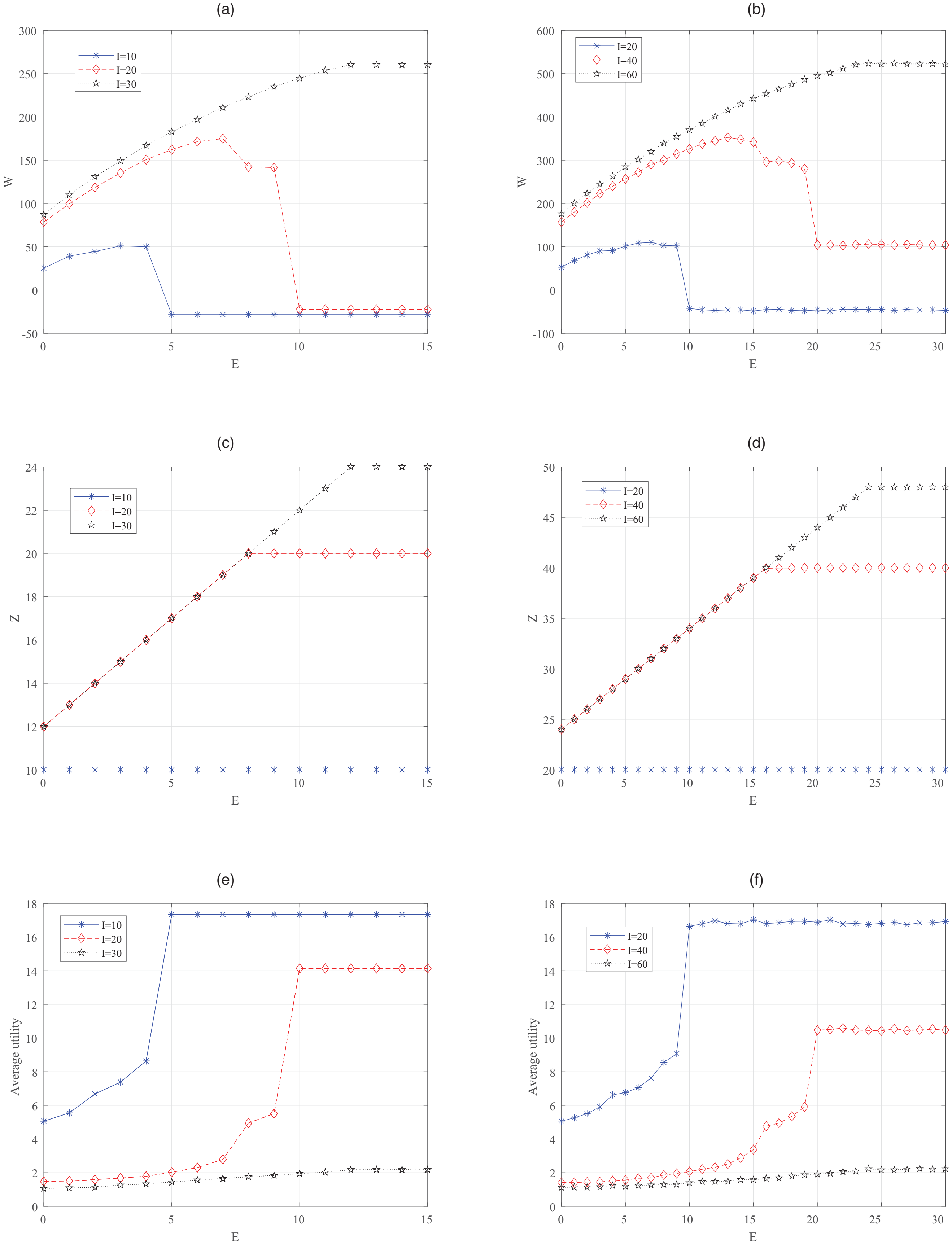
We next examine how the matching control E affects average system and commuter performance indexes. We set $c = 2$ and 4. When c equals two, the maximum system throughput is $2 \times 6 \times 2 = 24$ (everyone participates in ridesharing). For $c = 4$, the maximum throughput is 48. For each capacity level, three different demand levels are considered: for $c = 2$, we set $I = 10, 20, 30$, and for $c = 4$, $I = 20, 40, 60$. In both cases, the lowest demand level corresponds to an uncongested state because I is far less than the maximum throughput. On the contrary, the highest demand level leads to intense congestion because I is larger than the maximum throughput.

Figure 6 reports the results in all cases. We start from the case for $c = 2$, that is, the left three plots in the figure. Figure 6(a) shows that, for the low- and medium-demand levels ($I = 10$ and 20), the profit first increases and then decreases with E as predicted in Figure 2. When $I = 10$, the profit peaks at $E = 4$, and when $I = 20$, it peaks at seven. In both cases, reaching a perfect match (i.e., every commuter is matched with a ridesharing partner) leads to a deficit. For the high-demand level $I = 30$, however, the profit keeps increasing with E until it peaks at $E = 12$. As expected, the system throughput increases with E except for the lowest demand level (Figure 6(c)). In that case, the demand is so low relative to the capacity that everyone can be accommodated without any ridesharing. For $I = 20$ and 30, the throughput is maximized when E is increased to the level that maximizes the profit. This is welcoming news because both outcomes are desired. From the perspective of A-PASS, such a matching control is optimal. Figure 6(e) shows that commuters benefit from ridesharing in general: their average utility increases with E. Another interesting—though not surprising—finding from this plot is that congestion hurts commuters' utility. When I increases from 10 to 30, the average utility drops almost 90%. The results for $c = 4$ (the right column in Figure 6) are very similar.

Overall, Figure 6 offers two important takeaways. First, operating with a deficit seems a rare event in A-PASS regardless of the setting. It occurs only three times in more than 90 tests reported in Figure 6. In all three incidents, the system faced a low demand relative to its capacity and implemented a loose matching control, leading to a perfect match. These features can be easily identified and used to guide the operation of A-PASS. Second, A-PASS strongly promotes ridesharing through the trilateral matching scheme. As the matching control is relaxed, all stakeholders are better off in general: the operator receives greater profits, the commuters enjoy higher utility, and the society benefits from more efficient utilization of infrastructure.

Finally, Table 3 reports the average CPU times required to solve the TMP exactly (with a gap of zero) when the input parameters are closer to those from real-world applications. Here, we consider 20 time slots, which could cover roughly a two-hour analysis period if each lot is five minutes long. The capacity of the bottleneck ranges from 10 to 25 vehicles per slot, and the total number of commuters ranges from 600 to 1,000. The results show that the computation time rises at a relatively mild pace with the number of commuters and the bottleneck capacity. In the largest case tested, it takes a modest personal computer about 100 seconds to solve one instance. In order to set the second price policy, variants of the instance must be solved $20 \times 3 = 60$ times, which brings the total computation time to almost two hours. Yet this is likely a pessimistic estimation given many of these variants are very similar and, hence, can potentially be solved from a warm start.

Figure 6. (Color online) Average Performance Indexes vs. Matching Control Under Different Congestion Levels (Measured by the Ratio Between the Bottleneck Capacity C and the Number of Commuters I)



Note. Left column: $C = 2$; right column: $C = 4$.

Table 3. Computation Performance of Gurobi in Solving TMP (Averaged over 100 Runs)

Number of commuters	600	700	800	900	1,000
Bottleneck capacity	10	15	15	20	25
Time interval	20	20	20	20	20
Intlinprog solver runtime, s	20.1	28.0	35.8	44.8	56.1
Total runtime, s	34.0	47.8	66.7	80.1	101.4
Gap	0	0	0	0	0

8. Conclusions

In this study, we devised and analyzed a quantity-based travel demand management system aiming to promote ridesharing. The system sells permits to access a road facility (conceptualized as a bottleneck) through an auction but encourages travelers to share the permits with each other using ridesharing. The permit is classified according to access time, and the travelers may be assigned by the system one of the three roles: solo driver, rideshare driver, or rider. At the core of this A-PASS is a TMP that optimally matches permits, drivers, and riders. We first formulated the TMP as an integer program and then proved it can be reduced to an equivalent linear program thanks to the total unimodularity of the constraint structure. A pricing policy based on the classical VCG mechanism is proposed to determine the payment for each traveler. Although this policy guarantees all participants of the auction truthfully report their private information, it may not balance the budget. As a remedy, we propose a revised pricing policy that allows the operator of A-PASS to eliminate any deficit by controlling the number of shared rides. We also demonstrate A-PASS can be simply used as a tool to price the facility based on the access time and ridesharing role. Main findings from numerical experiments are summarized as follows:

- Consistent with the theoretical result, the experiments show the payment of a commuter depends only on the commuter's role and the access time.
- A traveler's utility increases as the access time deviates further away from the access time desired by everyone regardless of the role. Given the same access time, solo drivers have significantly lower utility compared with ridesharing drivers and riders.
- A-PASS does not usually operate with a deficit even with the VCG pricing policy. Yet a deficit may arise when the system faces a low demand relative to its capacity and the matching is not properly controlled.
- The trilateral matching scheme strongly promotes ridesharing. As sharing increases, all stakeholders are better off: the operator receives greater profits, the commuters enjoy higher utility, and society benefits from more efficient utilization of infrastructure.

The work presented in this paper can be extended in several directions. The obvious next step is to consider a more general representation of the transportation system than a single bottleneck. In a network setting, the proposed framework has to be extended to allow the valuation and payment of a traveler to depend on route choice, departure time, and ridesharing role. Wada and Akamatsu (2013) implement and analyze a tradeable network permit system in a network setting in which the permits are distributed through an auction scheme. Given that their system is also based on the bottleneck model, it may be used as a prototype to develop a networked A-PASS. Also, in the current setting, ridesharing can only occur between a driver and a rider. A future study can allow a driver to take multiple riders. Finally, other schemes, such as tradable credits, may be used to replace the auction in the A-PASS to distribute permits.

Acknowledgments

R. Li thanks Professor He Shuangchi at National University of Singapore for his discussion and support. The authors are grateful to the constructive comments provided by the associate editor and three anonymous reviewers. The remaining errors are those of the authors alone.

Endnotes

[1] For example, Uber claims that, by 2016, 20% of its trips were shared rides on UberPool: https://techcrunch.com/2016/05/10/uber-says-that-20-of-its-rides-globally-are-now-on-uber-pool/.

[2] Although empirical evidence suggests most commuters prefer an early over late arrival (see, e.g., Small 1982), here we assume, for simplicity, the value of a displacement does not depend on whether it leads to an early or late arrival. We note that all analytical results presented herein can be readily extended to accommodate the case of asymmetric schedule displacement.

[3] Because the entire trip is subject to no congestion, the time passing the bottleneck dictates the schedule displacement (i.e., the discrepancy between the desired and actual arrival time).

References

Agatz N, Erera A, Savelsbergh M, Wang X (2012) Optimization for dynamic ride-sharing: A review. *Eur. J. Oper. Res.* 223(2):295–303.

Akahane H, Kuwahara M (1996) A basic study on trip reservation systems for recreational trips on motorways. *Proc. Third World Congress Intelligent Transportation Systems*, 43–62.

Akamatsu T, Wada K (2017) Tradable network permits: A new scheme for the most efficient use of network capacity. *Transportation Res. Part C Emerging Tech.* 79:178–195.

Akamatsu T, Sato S, Nguyen L (2006) Tradable time-of-day bottleneck permits for morning commuters. *JSCE J. Infrastructure Planning Management* 62(4):605–620.

Arnott R, de Palma A, Lindsey R (1990) Economics of a bottleneck. *J. Urban Econom.* 27(1):111–130.

Balseiro SR, Mirrokni V, Leme RP, Zuo S (2019) Dynamic double auctions: Toward first best. *Proc. 30th Annual ACM-SIAM Sympos. Discrete Algorithms* (SIAM), 157–172.

Basar G, Cetin M (2017) Auction-based tolling systems in a connected and automated vehicles environment: Public opinion

and implications for toll revenue and capacity utilization. *Transportation Res. Part C Emerging Tech.* 81:268–285.
Cambridge Analytics (2007) License plate rationing evaluation. Accessed June 22, 2020, https://www.dot.ny.gov/programs/repository/Tech Memo on License Plate Rationing.pdf.
Chan ND, Shaheen SA (2012) Ridesharing in North America: Past, present, and future. *Transport Rev.* 32(1):93–112.
Clarke EH (1971) Multipart pricing of public goods. *Public Choice* 11(1):17–33.
Conitzer V, Sandholm T (2006) Failures of the VCG mechanism in combinatorial auctions and exchanges. *Proc. Fifth Internat. Joint Conf. Autonomous Agents Multiagent Systems*, 521–528.
De Feijter R, Evers JJM, Lodewijks G (2004) Improving travel-time reliability by the use of trip booking. *IEEE Trans. Intelligent Transportation Systems*, 5(4):288–292.
Edara P, Teodorović D (2008) Model of an advance-booking system for highway trips. *Transportation Res. Part C Emerging Tech.* 16(1):36–53.
Ferguson E (1997) The rise and fall of the American carpool: 1970–1990. *Transportation* 24:349–376.
Furuhata M, Dessouky M, Ordóñez F, Brunet ME, Wang X, Koenig S (2013) Ridesharing: The state-of-the-art and future directions. *Transportation Res. Part B Methodological* 57(57):28–46.
Groves T (1973) Incentives in teams. *Econometrica* 41(4):617–631.
Kamar E, Horvitz E (2009) Collaboration and shared plans in the open world: Studies of ridesharing. *Proc. Internat. Joint Conf. Artificial Intelligence*, 187.
Kleiner A, Nebel B, Ziparo VA (2011) A mechanism for dynamic ride sharing based on parallel auctions. *Internat. Joint Conf. Artificial Intelligence*, 266–272.
Li R, Nie YM, Liu X (2020) Pricing carpool rides based on schedule displacement. *Transportation Sci.* 54(4):1134–1152.
Liu Y, Li Y (2017) Pricing scheme design of ridesharing program in morning commute problem. *Transportation Res. Part C Emerging Tech.* 79:156–177.
Ma R, Zhang H (2017) The morning commute problem with ridesharing and dynamic parking charges. *Transportation Res. Part B Methodological* 106(C):345–374.
Mourad A, Puchinger J, Chu C (2019) A survey of models and algorithms for optimizing shared mobility. *Transportation Res. Part B Methodological* 123:323–346.
Myerson RB, Satterthwaite MA (1983) Efficient mechanisms for bilateral trading. *J. Econom. Theory* 29(2):265–281.
Nie Y (2012) Transaction costs and tradable mobility credits. *Transportation Res. Part B Methodological* 46(1):189–203.
Nie YM (2017) Why is license plate rationing not a good transport policy? *Transportmetrica A* 13(1):1–23.
Nie YM, Yin Y (2013) Managing rush hour travel choices with tradable credit scheme. *Transportation Res. Part B Methodological* 50:1–19.
Parkes DC, Kalagnanam J, Eso M (2001) Achieving budget-balance with Vickrey-based payment schemes in exchanges. *Internat. Joint Conf. Artificial Intelligence*, 1161–1168.
Pigou AC (1920) *The Economics of Welfare*, 1st ed. (Macmillan and Company, London).
Santi P, Resta G, Szell M, Sobolevsky S, Strogatz S, Ratti C (2014) Quantifying the benefits of vehicle pooling with shareability networks. *Proc. Natl. Acad. Sci. USA* 111(37):13290–13294.
Shaheen S, Cohen A (2018) The benefits of carpooling. Accessed June 18, 2020, https://escholarship.org/uc/item/7jx6z631.
Shaheen S, Cohen A (2019) Shared ride services in North America: Definitions, impacts, and the future of pooling. *Transport Rev.* 39(4):427–442.
Small K (1982) The scheduling of consumer activities: Work trips. *Amer. Econom. Rev.* 72(3):467–479.
Small KA, Gómez-Ibáñez JA (1998) Road pricing for congestion management: The transition from theory to policy. Button KJ, Verhoef ET, eds. *Road Pricing, Traffic Congestion and the Environment: Issues of Efficiency and Social Feasibility* (Edward Elgar, Cheltenham, UK), 213–246.
Su P, Park BB (2015) Auction-based highway reservation system an agent-based simulation study. *Transportation Res. Part C Emerging Tech.* 60:211–226.
Teodorović D, Triantis K, Edara P, Zhao Y, Mladenović S (2008) Auction-based congestion pricing. *Transportation Planning Tech.* 31(4):399–416.
Vickrey W (1961) Counterspeculation, auctions, and competitive sealed tenders. *J. Finance* 16(1):8–37.
Vickrey W (1969) Congestion theory and transport investment. *Amer. Econom. Rev.* 59(2):251–260.
Wada K, Akamatsu T (2013) A hybrid implementation mechanism of tradable network permits system which obviates path enumeration: An auction mechanism with day-to-day capacity control. *Procedia Soc. Behav. Sci.* 80:304–326.
Wang J-P, Ban XJ, Huang H-J (2019) Dynamic ridesharing with variable-ratio charging-compensation scheme for morning commute. *Transportation Res. Part B Methodological* 122: 390–415.
Wang X, Yang H, Han D (2010) Traffic rationing and short-term and long-term equilibrium. *Transportation Res. Rec.* (2196(1):131–141.
Wang X, Yang H, Zhu D, Li C (2012) Tradable travel credits for congestion management with heterogeneous users. *Transportation Res. Part E Logist. Transportation Rev.* 48(2):426–437.
Wang P, Wada K, Akamatsu T, Nagae T (2018) Trading mechanisms for bottleneck permits with multiple purchase opportunities. *Transportation Res. Part C Emerging Tech.* 95: 414–430.
Wong J-T (1997) Basic concepts for a system for advance booking for highway use. *Transport Policy* 4(2):109–114.
Xu H, Pang JS, Ordóñez F, Dessouky M (2015) Complementarity models for traffic equilibrium with ridesharing. *Transportation Res. Part B Methodological* 81:161–182.
Yang H, Wang X (2011) Managing network mobility with tradable credits. *Transportation Res. Part B Methodological* 45(3):580–594.
Zhang J, Wen D, Zeng S (2016) A discounted trade reduction mechanism for dynamic ridesharing pricing. *IEEE Trans. Intelligent Transportation Systems* 17(6):1586–1595.
Zhang C, Wu F, Bei X (2018) An efficient auction with variable reserve prices for ridesourcing. *Pacific Rim Internat. Conf. Artificial Intelligence* (Springer), 361–374.
Zhao D, Ramchurn SD, Jennings NR (2015) Incentive design for ridesharing with uncertainty. Preprint, submitted May 7, https://arxiv.org/abs/1505.01617.
Zhao D, Zhang D, Gerding EH, Sakurai Y, Yokoo M (2014) Incentives in ridesharing with deficit control. *Internat. Conf. Autonomous Agents Multi-Agent Systems*, 1021–1028.
Zhong L, Zhang K, Nie YM, Xu J (2020) Dynamic carpool in morning commute: Role of high-occupancy-vehicle (HOV) and high-occupancy-toll (HOT) lanes. *Transportation Res. Part B Methodological* 135:98–119.