Transportation Research Part B 93 (2016) 470–494

Contents lists available at ScienceDirect

Transportation Research Part B

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

Modeling the morning commute for urban networks with cruising-for-parking: An MFD approach

Wei Liu, Nikolas Geroliminis*

Urban Transport Systems Laboratory (LUTS), École Polytechnique Fédérale de Lausanne (EPFL), CH-1015 Lausanne, Switzerland

ARTICLE INFO

Article history:
Received 10 March 2015
Revised 10 August 2016
Accepted 13 August 2016

Keywords:
Morning commute
Cruising-for-parking
MFD
Pricing

ABSTRACT

This study focuses on the morning commute problem with explicit consideration of cruising-for-parking, and its adverse impacts on traffic congestion. The cruising-for-parking is modeled through a dynamic aggregated traffic model for networks: the Macroscopic Fundamental Diagram (MFD). Firstly, we formulate the commuting equilibrium in a congested downtown network where travelers have to cruise for curbside parking spaces. The cruising-for-parking would yield longer trip distance and smaller network outflow, and thus can induce severe congestion and lengthen the morning peak. We then develop a dynamic model of pricing for the network to reduce total social cost, which includes cruising time cost, moving time cost (moving or in-transit time, which is the duration during which vehicles move close to the destination but do not cruise for parking yet), and schedule delay cost. We show that under specific assumptions, at the system optimum, the downtown network should be operating at the maximum production of its MFD. However, the cruising effect is not fully eliminated. We also show that the time-dependent toll to support the system optimum has a different shape than the classical fine toll in Vickrey's bottleneck model. In the end, analytical results are illustrated and verified with numerical experiments.

1. Introduction

Parking is not only a headache for travelers heading for the city center, but also a challenging issue for the transport system planners, operators and regulators. In some cities, the time spent on searching for a vacant parking space can be up to 40% of the total travel time (Axhausen et al. 1994), due to the limitation of parking supply. Shoup (2006) summarized the findings of several studies done between 1927 and 2001, which show that between 8 and 74 percent of the traffic was cruising for parking, and the average time to find a curb space can be up to 14 minutes. However, those studies summarized in Shoup (2006) generally will report locations where cruising for parking cannot be neglected. Cruising might be less significant, especially for those areas with sufficient parking and low demand. More empirical evidence for cruising is needed for a more comprehensive understanding. Cruising-for-parking can also influence drivers not involved in cruising and create severe congestion even under medium demand conditions ("medium travel demand" will lead to no or very light congestion if travelers can find parking very easily when they are close to final destinations, and do not cruise for parking). This is because the outflow of the transport network (arrivals to the parking spaces) can reach very low values when finding a vacant parking space is extremely difficult (Geroliminis, 2015). Due to its inefficiency, cruising-for-parking is one of the

* Corresponding author.
E-mail addresses: w.liu@epfl.ch (W. Liu), nikolas.geroliminis@epfl.ch (N. Geroliminis).

http://dx.doi.org/10.1016/j.trb.2016.08.004

most studied topics in the economics of parking. Understanding the effect of cruising-for-parking for congested networks can help improve efficiency in the flow of vehicles and facilitate the development of more equitable management strategies as trips with cruising might contribute to congestion more than trips without that, e.g. trips with destinations outside the limited parking zones.

Glazer and Niskanen (1992) has modeled the congestion caused by through-traffic and by traffic destined for the area where consumers park. To evaluate different parking policies, Bifulco (1993) introduced the parking search times in a static stochastic traffic assignment model. Anderson and de Palma (2004) studied the parking problem under a private parking operator in a monopolistically competitive market, with an emphasis on the commuter's time spent on searching for a vacant parking space. There is a branch of literature looking into the interaction between cruising-for-parking and traffic congestion (e.g., Arnott and Rowse, 1999, 2009; Arnott and Inci, 2006, 2010). For example, Arnott and Rowse (1999) developed a structural model of parking for a ring-road on which travelers' choice of parking lot is uniformly distributed; the expected parking time, driving time and cruising distance for finding available parking spaces are derived. These studies focusing on cruising-for-parking have provided insightful ideas of the complex interaction among cruising, traffic congestion and network performance. However, they often overlook the rush hour traffic dynamics and time-varying traffic conditions. For a recent review of the economic studies of cruising-for-parking, one may refer to Inci (2015).

Another branch of literature has focused on integrating the parking problem into the well-known morning commute model (Vickrey 1969). In this context, Arnott et al. (1991) showed that a parking fee alone can effectively increase social welfare, and that a combination of dynamic road toll and dynamic parking fee can yield the system optimum. Zhang et al. (2008) further extended Arnott et al. (1991) by deriving the daily commuting pattern that combines both the morning and evening commutes. More recently, attentions have been paid to how parking capacity allocations, parking fees, parking permits and parking reservations can be designed to improve traffic efficiency in a dynamic network with one roadway bottleneck (Zhang et al. 2011; Qian et al. 2011, 2012; Fosgerau and de Palma, 2013; Yang et al. 2013; Liu et al. 2014a,b). However, in most of these studies, the cruising for parking is not modeled. Very recently, Qian and Rajagopal (2014, 2015) modeled how travelers make parking location choices and departure time choices to minimize their generalized travel cost. More importantly, they incorporated cruising-for-parking by using a cruising time function dependent on parking occupancy. However, their study treated cruising time as a cost at the end of trip, but ignored the impacts of cruising-for-parking on the roadway traffic congestion, as well as the interaction between cruising and moving traffic. By "moving traffic", we mean that the vehicles are moving towards their destinations, but have not started to search for a parking space yet.

To the best of our knowledge, this study is the first to incorporate not only the cruising-for-parking, but also its adverse impacts on traffic congestion in the context of dynamic commuting equilibrium. Alternatively, this study explores how interactions between cruising and traffic congestion will re-shape the morning commute. Following a recent macroscopic simulation study of parking (Geroliminis, 2015), the impact of cruising-for-parking is modeled through an aggregated network-level traffic model: the Macroscopic Fundamental Diagram (MFD), see Daganzo and Geroliminis (2008) for empirical evidence. Different from Geroliminis (2015), this study considers travelers' scheduling cost and time of departure choices rather than assuming given demand profile over time, which bring much more complexities. By adopting the MFD approach, one of the advantages is that the downward-sloping part of the curve between traffic flow and density, known as hypercongestion in economic terms (e.g., Small and Chu, 2003), can be modeled. The MFD approach has already been used to study the recurrent morning commute problem without consideration of cruising-for-parking (e.g., Geroliminis and Levinson, 2009; Arnott, 2013; Fosgerau, 2015).

As mentioned in Arnott (2013), the dynamic user equilibrium problem with hypercongestion is analytically intractable (which is to solve a delay differential equation with an endogenous delay). To deal with this intractability, Arnott (2013) assumes that the outflow from the downtown area depends on the contemporaneous traffic density, which is termed as the "bathtub" model. In an earlier time, to solve the no-toll commuting equilibrium, Small and Chu (2003) assumed that a commuter's travel time depends on traffic density at his or her arrival time. This assumption is also adopted in some other studies, e.g., Mahmassani and Herman (1984), Yang and Huang (1997). Later, Geroliminis and Levinson (2009) extended Small and Chu (2003) by considering user heterogeneity in desired arrival time, and incorporating various pricing strategies to eliminate congestion. An interesting finding of this study is that the duration of pricing period is smaller than the congested period in the no-toll case, and the total savings (in travel delay and scheduling penalties) are higher than the total toll paid (note that in a classical bottleneck, these two quantities are equal). A simplified tractable version of the MFD model considering capacity drop facing queueing is adopted in some recent studies on the morning commute problem (e.g., Fosgerau and Small, 2013; Liu et al., 2015b). Unfortunately, extension of their models would lead to a tedious proliferation of cases. Very recently, Fosgerau (2015) proposed a similar "bathtub" model as Arnott (2013), given the heterogeneity in trip length of the population. Some of the analysis in Fosgerau (2015) relies on that, for all travelers, the speed does not drop too quickly at times of departure or rise too quickly at times of arrival. The paper identified that under some conditions, a regular sorting property arises, where shorter trips take place within the durations of longer trips. In the current study, to tackle the intractability, we approximate travelers' travel time with instantaneous speed (depends on traffic density) and trip length (depends on parking availability). With this "instantaneous travel time" approximation, the problem is significantly simplified and still analytically tractable. Later our numerical analysis can compare the difference between the analytical travel time (calculated with instantaneous conditions) and the estimated travel time (calculated with the estimated departure/arrival traffic pattern). While a number of assumptions have been made to keep some level of analytical tractability in the model, the analysis of this work contains interesting findings both from a mathematical and policy point of view and could help policy makers in dealing with congestion and parking issues.

Under the MFD framework, the traffic arrival rate at destinations or the outflow of the network depends on the traffic accumulation in the network and the trip length of the traffic. The existing MFD models often assume that the trip length is constant over time and independent of destination, and ignore phenomena which may change trip length, e.g., when vehicles are cruising for parking (increase in trip length due to route choice is analyzed in Yildirimoglu and Geroliminis, 2014). In reality, parking capacity in the downtown is often limited. Moreover, due to travelers' arrival, the available parking capacity will decrease over time in the morning peak. It is then more difficult for a later traveler to find a vacant parking space. In an average sense, cruising distance to find a vacant parking space will increase over time, which leads to a decrease in network outflow (arrival rate at parking spaces). Macroscopic models that ignore this phenomenon will underestimate trip length of travelers and overestimate the outflow of the network. Furthermore, the increased travel distance due to cruising will lead to more severe congestion in the network. If we look at the network traffic dynamics, given the future traffic inflow, the decreased outflow due to cruising-for-parking would in return intensify the network accumulation in the future, and then decrease the traveling speed of traffic and create more severe congestion.

This study models the above interactions in the context of dynamic user equilibrium in the morning commute. Firstly, we formulate the commuting equilibrium in a congested downtown network where travelers have to search for vacant curbside parking spaces around their destinations. Since the cruising-for-parking would yield smaller network outflow, more traffic congestion, and more travel delays, we develop a dynamic model of pricing for the network to improve efficiency, and reduce total social cost including cruising time cost, moving time cost (where moving time is the duration during which vehicles move to the destination but do not cruise for parking yet), and schedule delay cost. Our analysis shows that, due to the consideration of cruising-for-parking, the dynamic toll to support the system optimum has a different shape (over time) than the classical triangular toll in Vickrey's bottleneck model.

The rest of the paper is organized as follows. Section 2 describes the problem and presents the model formulation and major assumptions adopted. In Section 3, the morning commute equilibrium with cruising-for-parking is discussed. Section 4 introduces the optimal time-varying toll to reduce total social cost and improve traffic efficiency. Numerical studies are presented in Section 5 to illustrate and verify the essential ideas in the paper. Finally, Section 6 concludes the paper and provides some discussions.

2. Model formulation

We start with a thumbnail sketch of the problem, and follow it with more detailed formulations. Specifically, we list the main model assumptions of this work A1-A7 in subsection 2.2, while some initial or boundary traffic conditions are summarized as A8-A10 in subsection 3.1.

2.1. Thumbnail description

We consider a downtown area, which exhibits an MFD with low scatter (a lower scatter requires that the congestion is more homogeneously distributed over the network). Basically, the MFD of a network describes the relationships among network vehicle density, network average speed of traveling traffic, and network space-mean flow (or network travel production). A formal definition of the variables is provided right afterwards. The MFD of a network or region can be estimated with various real data (from loop detectors, GPS etc. see Geroliminis and Daganzo, 2008 and Leclercq et al., 2014).[1] In this paper, the network MFD function is assumed to be given.

In the morning commute, a number of travelers have to travel through the downtown network to reach their downtown destinations. In this work, all of them have identical desired arrival time. All travel is by car and all parking is on street. Before arrival, travelers need to find vacant parking spaces to park their cars. Since all parking is on street, travelers will cruise for parking on the street, which depends on parking availability. Indeed, there will be two types of traffic: moving traffic (or in-transit traffic, which means that vehicles are moving towards their destinations, but have not started to find a parking space yet), and cruising traffic (vehicles are searching around their final destinations to find vacant parking spaces). The traffic dynamics after taking into account cruising-for-parking are modeled through the Macroscopic Fundamental Diagram (MFD), see Geroliminis and Daganzo (2008). By utilizing the MFD framework, we can model the downward-sloping part of the curve between network traffic flow and density.[2]

A commuter will choose his or her departure time to minimize the travel cost, which includes travel delay cost and schedule delay cost. The travel delay includes both moving time and cruising time. We look at the long-term dynamic user equilibrium such that we assume travelers to be aware of traffic conditions and parking vacancies after their long-term experience. Equilibrium is achieved when no one can reduce his or her travel cost by unilaterally changing departure time.

[1] The MFD of a network can be affected by, e.g., systematic changes of signal control plans, expanded roads (see for example Leclercq and Geroliminis, 2013 or Geroliminis and Boyaci, 2012).

[2] While this study focuses on a single-region downtown network, the dynamic relations for traffic in a multi-region system with a known time-of-departure for all travelers and no cruising are described in detail by Ramezani et al. (2015).

We end the subsection with listing the notations we employ in the paper. Those listed as exogenous variables or parameters are considered as input to the model and can be estimated with field experiments or surveys.

MFD variables:	
n	Accumulation (total number of vehicles in the network)
n_c	Critical level of accumulation (exogenous)
n_{jam}	Jam accumulation (exogenous)
v	Velocity; $v = v(n)$ (exogenous relation between speed and accumulation)
v_f	Free-flow speed (exogenous)
P	Network production; $P \equiv v \cdot n$
L	Trip length
o	Network outflow; where $o = P/L$

Exogenous parameters:	
N	Number of commuters
N_p	Number of parking spaces
l_m	Moving or in-transit distance (identical for all travelers)
d	Distance or spacing between parking spaces
p_0	Parking occupancy rate at the start of the peak
t^*	Desired arrival time (identical for all travelers)

Endogenous variables:	
t	Clock time
$I(t)$	Cumulative departure from home at time t
$I'(t)$	Departure rate at time t; $I'(t) \equiv dI(t)/dt$
$A(t)$	Cumulative arrival at destination at time t
$A'(t)$	Arrival rate at time t; $A'(t) \equiv dA(t)/dt$
$p(t)$	Parking vacancy rate experienced by travelers departing at time t
$\tilde{p}(t)$	Parking vacancy rate at time t
$n(t)$	Accumulation at time t ($\equiv I(t) - A(t)$)
$\hat{v}(t)$	Velocity at time t; and $\hat{v}(t) = v(n(t))$
$\hat{o}(t)$	Network outflow at time t
$l_s(t)$	Cruising-for-parking distance for travelers departing at time t
$\hat{L}(t)$	Trip length for travelers departing at time t; $\hat{L}(t) \equiv l_m + l_s(t)$
$\tau_m(t)$	In-transit travel time or moving time for travelers departing at time t
$\tau_s(t)$	Cruising-for-parking time for travelers departing at time t
$\hat{\tau}(t)$	Total travel time for travelers departing at time t ($\equiv \tau_m + \tau_s$)
$T(t)$	Time-varying toll at (departure) time t

Note that the time-varying accumulation $n(t)$ follows the notation n from the MFD variables. However, the notations $\hat{v}$, $\hat{o}$, $\hat{L}$ and $\hat{\tau}$ (all are time-varying endogenous variables) are adopted to distinguish them from relevant functions. Other parameters and variables if not mentioned in the above will be specified in the text.

2.2. Dynamic traffic model with cruising-for-parking

Regarding the MFD of the downtown network, we employ the following assumptions A1-A2. The speed-accumulation ($v - n$) relationship defined in assumptions A1 is also illustrated in Fig. 1.

A1. *Velocity is a continuous function of accumulation in the network:*

$$v = v(n)$$

In particular, we assume that

$$v(n) = v_f \text{ for } n \le n_c;$$

$$v(n) > 0 \text{ and } v'(n) < 0 \text{ for } n > n_c \text{ and } n < n_{jam};$$

$$v(n) = 0 \text{ for } n \ge n_{jam};$$

where n_c is the critical accumulation, and n_{jam} is the jam accumulation.

The above assumption A1 means that traffic flows at free-flow velocity up to a critical level n_c, above which the velocity declines monotonically as accumulation increases (before jam accumulation), reflecting (hyper)congestion.[3] Note that the

[3] The assumption (A1) also indicates that we do not differentiate the direct impacts of cruising and moving vehicles on traffic congestion. This is often not the case in reality. Vehicles cruising on the street might be slower, and might affect surrounding traffic when they try to park at vacant spaces they find. Future research might model the specific impacts of cruising traffic. To achieve this, we should identify cars that are cruising, which can be difficult as we need to track the state of every vehicle. However, from a system perspective, we might approximate the proportions of moving and cruising vehicles

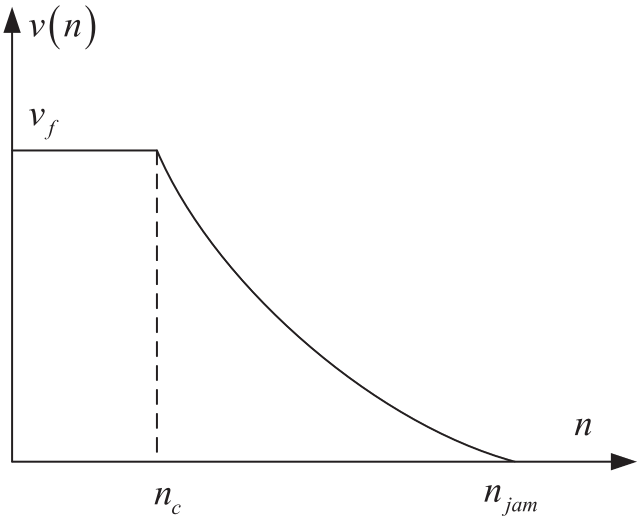


Fig. 1. The speed-accumulation relationship for the downtown network.

speed might not be differentiable while continuous at $n = n_c$ and at $n = n_{jam}$. This assumption (A1) is adopted in many traffic simulation studies, and empirical observations from Yokohama (Geroliminis and Daganzo, 2008 and Geroliminis and Levinson, 2009) suggest this to be a reasonable approximation. For later use, we here define $v^{-1}(\cdot)$ as the inverse function of $v = v(n)$ for $n \in [n_c, n_{jam})$. Since we are focusing on the dynamic user equilibrium with cruising for parking, the accumulation n will never reach the jam accumulation n_{jam} (otherwise, travelers will have an unbounded travel cost). If at time t, accumulation is $n(t)$, velocity is then $\hat{v}(t) = v(n(t))$.

A2. *Production $P(n) = n \cdot v(n)$, i) increases with n for $n < n_c$; and ii) decreases with n for $n > n_c$ and $n < n_{jam}$.*

Part (i) of assumption A2 holds as long as assumption A1 holds. Part (ii) of Assumption A2 implies that, for $n > n_c$ and $n < n_{jam}$, we have $dP(n)/dn < 0$. This requires that $\frac{dv(n)}{dn} < -\frac{v(n)}{n}$. The production P reaches its maximum when $n = n_c$.

The network outflow $o = P/L$ would have the same shape (over accumulation n) as the production P if trip length L is constant. Note that $o = P/L$ is a reasonable approximation (see, e.g., Geroliminis and Daganzo, 2008), which is used in the flow conservation equation for representing the network dynamics. In this paper, trip distance L depends on parking availability, which is changing over time. Therefore, network outflow can be different even if the production is identical and the $o - n$ plot (we do not show it here) might experience hysteresis phenomena due to variable L. Moreover, even if the production is at its maximum, the network outflow can be smaller when trip length is longer due to cruising.

The purpose of this study is to examine the integrated problem of morning commute and downtown parking (on-street parking) in the context of dynamic user equilibrium. This means that accumulation, traffic velocity, and parking vacancy rate would all be time-dependent. Thus, during the journey of a traveler (both the moving part and cruising part), both the velocity and parking vacancy rate are changing, which make the model analytically intractable (if without any approximations). The problem is that the exact model gives rise to delay differential equations with an endogenous delay (as described by Arnott (2013) even for the case of no cruising). To circumvent sources of analytical intractability, we make the following assumptions (A3-A7).

A3. *The moving time (in-transit travel time) for a traveler who departs from home at time t depends on the traffic accumulation at the beginning of his trip (at time t):*

$$\tau_m(t) = \frac{l_m}{v(n(t))}. \tag{1}$$

Assumption A3 simply means that, given the moving distance, the moving time of a trip depends on instantaneous speed (at departure time), instead of velocity during the moving duration of the trip. As will be discussed later, assumptions A3-A7 together lead to an "instantaneous model". In this model, we work with departure-time-dependent accumulation and speed (rather than arrival-time-dependent) to estimate travel time as we consider that capturing the evolution of congestion during onset is critical.

(among the total accumulation) by assuming that they are proportional to moving distance and cruising distance (as briefly discussed in Geroliminis (2015), this is reasonable only if the system traffic conditions change slowly over time). Detailed GPS data (to identify if a vehicle follows a straight or circuitous trajectory) combined with parking occupancy measures that identify how often each parking spot changes its state from busy to available, would be two important aspects of a future analysis.

A4. *The parking vacancy rate for a traveler who departs from home at time t depends on cumulative departures at time t:*

$$p(t) = 1 - \frac{I(t)}{N_p} - p_0. \tag{2}$$

Eq. (2) implies that a traveler departing earlier will experience a higher parking vacancy rate as $I(t)$ will be smaller. This means that first-in-first-out (FIFO) is implicitly assumed. It follows that the cumulative departure at the departure time of a traveler will be equal to the cumulative arrival at the arrival time of this traveler. Also note that $p_0 \cdot N_p$ parking spaces are occupied at the start of the rush hour. If p_0 is larger, the initial parking vacancy rate is lower, and the cruising effect will be more significant.

A5. *The cruising for parking distance for a traveler who departs from home at time t equals the distance between parking spaces divided by parking vacancy rate defined in Eq. (2):*

$$l_s(t) = \frac{d}{p(t)}. \tag{3}$$

The above assumption A5 is inherently coupled with A4. We use parking vacancy rate in Eq. (2) (depending on cumulative departure at departure time t) to estimate the cruising distance of a traveler departing at time t (rather than parking vacancy rate depending on cumulative arrivals at time t), i.e., $l_s(t)$ depends on $p(t)$. We consider this as a more realistic assumption. The searching for parking occurs at the end of the trip, which means that the parking vacancy rate experienced by the traveler departing at time t would be closer to the one when the traveler arrives.[4] If FIFO is assumed, when the traveler departing at time t arrives, the cumulative arrival would be equal to the cumulative departure at time t, which means the exact parking vacancy rate when a traveler departing at time t arrives at the parking would be as in Eq. (2).

Note also that Eq. (3) denotes an average searching distance for a given d and p, i.e., $l_s = d/p$. We indeed follow Anderson and de Palma (2004), where readers can find a detailed derivation. It regards searching for parking spaces as a stochastic process with replacement, which means a traveler who is cruising for parking "forgets" whether he or she has previously checked on a space. It is worth mentioning that alternative formulations of searching distance can be readily accommodated in our model as long as the searching distance l_s decreases with parking vacancy rate p, and is convex over p.

A6. *Cruising-for-parking time for a traveler who departs from home at time t equals the cruising distance for a traveler who departs at time t divided by the velocity at time t:*

$$\tau_s(t) = \frac{l_s(t)}{v(n(t))}. \tag{4}$$

Assumption A6 means that, similar to assumption A3, we use departure-time-based speed to estimate cruising time of travelers. One may argue that cruising is closer to the end of the trip thus it might be more appropriate to use speed at arrival time. However, analytical intractability would arise if we adopt different speeds for moving time and cruising time.[5]

The trip distance of a traveler departing at time t can be given as

$$\hat{L}(t) = l_m + l_s(t) = l_m + \frac{d}{p(t)} \equiv L(p(t)). \tag{5}$$

It can be seen in Eq. (5) that $L(p)$ is decreasing in p. From assumptions A3 and A6, we further have

$$\hat{\tau}(t) = \frac{\hat{L}(t)}{\hat{v}(t)} = \frac{L(p(t))}{v(n(t))} \equiv \tau(n(t), p(t)). \tag{6}$$

From Eq. (1), Eq. (4), Eq. (5) and Eq. (6), it is obvious that $\hat{\tau}(t) = \tau_m(t) + \tau_s(t)$. And $\tau(n, p)$ is decreasing in p, and is increasing in n when $n \geq n_c$ and $n < n_{jam}$ (since $v(n)$ is decreasing with n).

At time t, total arrival at destination (as well as parking) is $A(t)$, the parking vacancy rate at time t is then

$$\tilde{p}(t) = 1 - \frac{A(t)}{N_p} - p_0. \tag{7}$$

At time t, the traveler departing at time t have not arrived yet, i.e., $I(t) > A(t)$. Therefore, this parking vacancy rate in Eq. (7) is smaller than the one encountered by the traveler departing at time t, which is given in Eq. (2). The traveler departing at time t is still far from starting to cruise for parking, and this is why we do not use the parking vacancy rate in Eq. (7) to estimate the cruising distance (as well as the total trip distance) for the traveler departing at time t.

A7. *The network outflow at time t equals the production at time t divided by trip length for travelers arriving at time t,* i.e.,

$$\hat{o}(t) = \frac{n(t) \cdot v(n(t))}{L(\tilde{p}(t))} \equiv o(n(t), \tilde{p}(t)). \tag{8}$$

[4] We can expect that the exact parking vacancy rate at the departure time of a traveler (depending on cumulative arrivals at the departure time, as given in Eq. (7)) can be far from (much smaller than) that at the arrival time of this traveler.

[5] If we adopt for both moving time and cruising time the arrival-time-dependent speed, one may also argue that it might be better to use departure-time-based speed to estimate moving time. It can be said that, since the departure time and arrival time are fully dependent on each other, formulating the problem based on either departure time or arrival time will make this "instantaneous travel time" assumption equally strong. We work with departure-time-based accumulation and speed as we consider that capturing the evolution of congestion during onset is more important and errors will accumulate less.

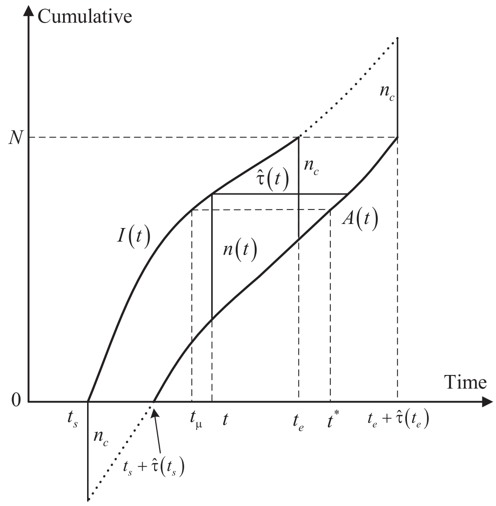


Fig. 2. Cumulative departure and arrival at the user equilibrium.

Based on assumptions A4 and A5, the parking vacancy rate at time t given in Eq. (7) is the one encountered and experienced by travelers departing at time $\tilde{t} < t$ where $I(\tilde{t}) = A(t)$, and we have $p(\tilde{t}) = \tilde{p}(t)$. At time t, the travelers departing at time $\tilde{t}$ arrive. Therefore, the outflow or arrival rate at time t given in Eq. (8) is for the arrival of travelers departing at time $\tilde{t}$. The trip distance assumed to calculate outflow is equal to $L(p(\tilde{t})) = L(\tilde{p}(t))$.

Assumptions A3-A7 together might be termed as an "instantaneous model".[6] The most critical component is that we assume "instantaneous travel time" (departure-time-based).[7] We would like to highlight that trip length formulation is also departure-time-based. We use cumulative departure at the departure time of the traveler to estimate parking vacancy rate and trip length experienced by the traveler (When FIFO is assumed, cumulative departure at the departure time of a traveler will be equal to cumulative arrival at the arrival time of the same traveler).

With assumptions A3-A7, the problem is simplified and the model becomes analytically tractable. Later our numerical analysis compares the difference between the analytical travel time (calculated with instantaneous conditions, which is given by Eq. (6)) and the "estimated" travel time (calculated with the estimated departure/arrival traffic pattern). More specifically, we compute $I(t)$ and $A(t)$ based on our "instantaneous model" (how to estimate the solution is presented in Section 3). And the "estimated" travel time is the horizontal gap between cumulative departure curve and cumulative arrival curve. For a traveler departing at time t_1, "estimated" travel time would be $\hat{\tau}_{est}(t_1) = t_2 - t_1$, where $A(t_2) = I(t_1)$ (one may refer to Fig. 2 for better understanding). "Estimated" schedule delay costs then can be determined accordingly. Note that the "estimated" values only appear when we want to compare them with the solution of the analytical model defined by assumptions A1-A7.[8]

Given the above formulations and assumptions, the full trip cost of a commuter departing from home at time t is given by

$$c(t,t^*) = c_w \cdot \tau(n(t), p(t)) + c_s \cdot (t^* - t - \tau(n(t), p(t))), \tag{9}$$

where $\tau(n(t), p(t))$ is the travel time defined by Eq. (6), c_w is the value of unit travel time, and c_s is the schedule penalty of unit time.[9] And the schedule penalty is $c_s = e$ for a unit time of early arrival, i.e., $t^* \geq t + \tau(n(t), p(t))$, and is $c_s = -l$ for a unit time of late arrival, i.e., $t^* < t + \tau(n(t), p(t))$. It is assumed that $e < c_w < l$, which is consistent with empirical evidence.

[6] While in this paper the "instantaneous travel time" formulation is considered, a more accurate estimation of experienced travel time (e.g. as described in Yildirimoglu and Geroliminis, 2013) will not allow for analytical derivations.

[7] To approximate travel time with arrival-time-dependent accumulation and speed, we can define the t in Eq. (9) as arrival time, and let $c(t,t^*) = c_w \cdot \tau(n(t), p(t)) + c_s \cdot (t^* - t)$. Then we can conduct similar derivations as those in the current paper. This is indeed applied to estimate the solution presented in numerical analysis (Fig. 7(b)) where we compare our results with the solution based on arrival-time-dependent formulation.

[8] In Section 5, it is numerically shown that the "estimated" travel time over the time horizon still follows a similar pattern as the analytical travel time under our approximation. Also, the discrepancies in system efficiency measures such as total travel time, total travel cost and total schedule delay cost will be less than 10% (user equilibrium solution), while we omit the detailed discussion. Indeed, a vast literature in travel time estimations indicates that errors associated with instantaneous travel times for smoothly varying traffic conditions might be in the range of 5-10% (e.g. Yildirimoglu and Geroliminis, 2014).

[9] In the travel cost formulation in Eq. (9), the walking time between the parking spaces and final destination (e.g., workplaces) is not considered, i.e., we only consider the driving time. If walking time is increasing over time (later arrival indicates parking further away), which is similar to cruising time, incorporating walking time will likely give us similar results as those from this study. However, with more detailed spatial consideration of parking, the walking time will be related to specific parking spatial distribution, and might affect how travelers are cruising and are choosing where to park. In this case, to incorporate walking time can be much more challenging, which is under our consideration for further research.

3. Morning commute equilibrium with cruising-for-parking

3.1. User equilibrium description

If the parking vacancy rate p is constant during the morning peak, our study is similar to Small and Chu (2003) and Geroliminis and Levinson (2009). Dynamic user equilibrium is achieved when no one can reduce travel cost by unilaterally changing his or her departure time. However, the time-dependent p would affect the cruising-for-parking and congestion dynamics in the network, and then re-shape the dynamic user equilibrium.

Compared to the standard bottleneck studies, this research incorporates a more realistic traffic flow model to capture traffic dynamics, which makes it necessary to describe the initial traffic conditions (this is common in traffic simulation studies with more detailed traffic flow models). Suppose the first and last departures (from home) occur at time t_s and at time t_e respectively, as shown in Fig. 2 (note that both time points are endogenous, and need to be estimated). We have the following assumption for initial traffic conditions.

A8. *At peak start time t_s, the accumulation equals the critical value,* i.e., $n(t_s) = n_c$ *(onset of congestion).*

The assumption above (A8) means that the onset of congestion is at the peak start time t_s. This assumption ensures that, based on the "instantaneous travel time" formulation, the first traveler will experience free-flow speed and zero congestion delay. This is similar to the standard bottleneck model where the first traveler also experiences free-flow speed and zero congestion delay.

Between time t_s and time t_e (one may refer to Fig. 2), the accumulation will go beyond the critical value n_c, and the speed is below the free-flow speed (congestion). At the time point of last departure t_e, the accumulation will go back to the critical value, i.e., $n(t_e) = n_c$, thus the last traveler will also experience free-flow speed (based on the "instantaneous travel time"). Note that if the accumulation $n(t_e) > n_c$ (corresponding case in standard bottleneck model is that the last traveler will experience non-zero queuing delay), we can construct an equilibrium with a later t_s such that $n(t_e) = n_c$, and equilibrium individual travel cost is less. An equilibrium with $n(t_e) < n_c$ also cannot exist. This is explained as follows. If $n(t_e) < n_c$, there must be a small time duration just before t_e where the accumulation will be less than n_c. For travelers departing during this small time duration, they all experience a free-flow speed. Among these travelers, a later departure indicates a larger travel cost (equilibrium condition does not hold), since a later departure means a larger travel time (trip length is larger as parking vacancy will decline), and a larger schedule delay (later arrival).

Between t_s and $t_s + \hat{\tau}(t_s)$, there will also be arrivals (of non-peak traffic, i.e., the traffic already in the network at time t_s), which are displayed by dotted line in Fig. 2. To compute the outflow (arrival rates) during this time interval, as well as the accumulation, we need the trip distance information of these non-peak traffic. We have the following assumption.

A9. *For the travelers already in the network at time t_s, the parking vacancy rate experienced is $p = p(t_s) = 1 - p_0$, and trip distance is then constantly equal to $l_m + d/(1 - p_0)$.*

Assumption A9 gives information about the trip length of the non-peak traffic already in the network at time t_s. For simplicity, we assume that the trip length of them is constantly equal to the trip length of the first traveler in the peak demand N (who departs from home at time t_s). Firstly, this assumption assures that early non-peak traffic will encounter a reasonably small cruising distance (and parking vacancy rate is relatively high for them). Secondly, alternative assumptions on trip length of the non-peak traffic can be easily accommodated in our model.

Furthermore, after t_e, there will also be departures (which do not belong to the peak demand N). These departures of non-peak traffic can be less intensive thus we have $n(t) \le n_c$ and the speed is at its maximum. The outflow after t_e, as well as the accumulation is related to the departures of non-peak traffic. For simplicity, we assume the following.

A10. *Between t_e and $t_e + \hat{\tau}(t_e)$, $n(t) = n_c$.*

From the discussion after assumption A8, we know that $n(t_e) = n_c$. Assumption A10 then means that between t_e and $t_e + \hat{\tau}(t_e)$, the accumulation will remain constant, which indicates that the network inflow equals the outflow during the mentioned time interval. However, we would like to point out that, assumption A10 is not critical to our model, as the model can easily accommodate alternative assumptions, e.g., $n(t) < n_c$ during this time interval (non-peak traffic after t_e is very light such that inflow is less than outflow).

The conditions assumed in the above related to the start and end of the peak might not be acceptable unless the peak hour is significantly longer than the travel time of a single trip with zero congestion. However, the model, while adopting initial or boundary condition assumptions A8, A9, and A10, still incorporates the standard bottleneck model (with constant highway capacity) as a special case.

3.2. User equilibrium conditions

We now derive the dynamic user equilibrium conditions. As mentioned, equilibrium is achieved when no one can reduce travel cost by unilaterally changing his or her departure time. By taking the first-order derivative of the individual travel cost given by Eq. (9) with respect to t, we have

$$\frac{\partial c(t,t^*)}{\partial t} = (c_w - c_s) \cdot \left[\frac{\partial \tau(n(t),p(t))}{\partial n} \cdot \frac{dn(t)}{dt} + \frac{\partial \tau(n(t),p(t))}{\partial p} \cdot \frac{dp(t)}{dt} \right] - c_s. \tag{10}$$

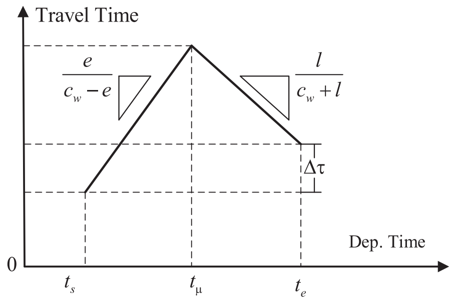


Fig. 3. The equilibrium travel time profile over departure time.

Equilibrium requires that $\partial c(t, t^*)/\partial t = 0$, indicating a traveler cannot reduce his or her travel cost by changing departure time. Then we have

$$\frac{\partial \tau(n(t), p(t))}{\partial n} \cdot \frac{dn(t)}{dt} + \frac{\partial \tau(n(t), p(t))}{\partial p} \cdot \frac{dp(t)}{dt} = \frac{c_s}{c_w - c_s}. \tag{11}$$

Note that in Eq. (11), the schedule penalty c_s is different for early and late arrivals.

For the first traveler departing at time t_s, travel time is given by $\hat{\tau}(t_s) = \tau(n_c, 1 - p_0)$. The on time traveler departs at time t_μ, thus $t_\mu + \tau(n(t_\mu), p(t_\mu)) = t^*$, and $\hat{\tau}(t_\mu) = \tau(n(t_\mu), p(t_\mu))$. The last traveler will depart at time t_e and $\hat{\tau}(t_e) = \tau(n_c, 1 - N/N_p - p_0) > \hat{\tau}(t_s)$. The estimation of t_s, t_μ and t_e, as well as the dynamic traffic equilibrium, will be discussed later in Section 3.3. With Eq. (11) and the boundary conditions described in Section 3.1, we can construct the equilibrium travel time profile, which is given as follows

$$\left(\hat{\tau}(t)\right)^* = \begin{cases} \frac{e}{c_w - e}(t - t_s) + \hat{\tau}(t_s) & \text{for } t_s \le t < t_\mu \\ \hat{\tau}\left(t_\mu\right) - \frac{l}{c_w + l}\left(t - t_\mu\right) & \text{for } t_\mu \le t \le t_e \end{cases}. \tag{12}$$

Eq. (12) is also depicted in Fig. 3.

As mentioned, $\hat{\tau}(t_e) > \hat{\tau}(t_s)$ holds (also shown in Fig. 3), i.e., $\Delta\tau = \frac{1}{1-p_0} \cdot \frac{N}{(1-p_0)\cdot N_p - N} \cdot \frac{d}{v(n_c)} > 0$. This means that even both the first and last travelers can enjoy free-flow speed and zero congestion delay (similar to standard bottleneck model), they will encounter different travel times, due to cruising-for-parking. When there is no cruising, i.e., $N_p \to \infty$, we have $\Delta\tau = 0$. The above also implies that the last commuter experiencing a longer travel time (due to cruising) than the first commuter will experience less schedule delay cost as compensation. In an alternative way, to enjoy less cruising-for-parking, commuters have to travel earlier and might encounter larger schedule delay cost.

As all travelers have identical $t*$, for given t_s and $\hat{\tau}(t_s)$, the departure time of on-time traveler t_μ can even be explicitly determined by these two equations: $\hat{\tau}(t_\mu) = e \cdot (t_\mu - t_s)/(c_w - e) + \hat{\tau}(t_s)$ and $t_\mu + \hat{\tau}(t_\mu) = t^*$, which is

$$t_\mu = \frac{c_w - e}{c_w}\left(t^* - \hat{\tau}(t_s)\right) + \frac{e}{c_w} t_s, \tag{13}$$

and $\hat{\tau}(t_\mu)$ can be determined accordingly. However, if $t*$ is not identical for the population, then the closed-form solution of t_μ is not available, and t_μ has to be determined by numerically solving the equation $t_\mu + \tau(n(t_\mu), p(t_\mu)) = t^*$.

The existence of the dynamic user equilibrium (as well as the existence of Eq. (12)) relies on exclusion of extreme cases with $\hat{\tau}(t_e) \to \infty$. This is discussed as follows. Eq. (12) implies that the inequality $\hat{\tau}(t_e) \le \hat{\tau}(t_\mu)$ holds, i.e., $\tau(n_c, p(t_e)) \le \frac{e}{c_w - e}(t_\mu - t_s) + \hat{\tau}(t_s)$. Suppose now that the network has to accommodate a relatively large demand where $\frac{N}{N_p} \to 1 - p_0$, the parking vacancy rate $p(t_e)$ would be small, and travel time $\tau(n_c, p(t_e))$ will be very large. The above inequality might still hold, as under a larger N the peak may start earlier (smaller t_s), and $t_\mu - t_s$, as well as $\frac{e}{c_w - e}(t_\mu - t_s) + \hat{\tau}(t_s)$ will become larger. However, the cruising distance function $L(p)$ in Eq. (5) is strictly decreasing, convex over p. And it is unbounded when $p \to 0$. This means that the increase of $\tau(n_c, p(t_e))$ resulting from a larger demand N can be much larger than that of $\frac{e}{c_w - e}(t_\mu - t_s) + \tau_s$ (please note that the slope $\frac{e}{c_w - e}$ is bounded), especially when $p(t_e) \to 0$. $\tau(n_c, p(t_e)) \le \frac{e}{c_w - e}(t_\mu - t_s) + \hat{\tau}(t_s)$ will not hold then as N becomes close enough to N_p. In summary, the existence of equilibrium requires that the increase of travel time due to increasing cruising distance (even with free-flow speed) cannot exceed the maximum travel time increase allowed by the equilibrium condition.[10]

For given N_p, there exists a critical t_s and corresponding N such that $\hat{\tau}(t_\mu) = \hat{\tau}(t_e)$, and $t_\mu = t_e$, and the last traveler arrives at the destination just on time.[11] Denote the t_s and N in this critical case by $\bar{t}_s$ and $\bar{N}$. For $N \le \bar{N}$, we can construct the

[10] The equilibrium travel time profile (over time) depends on schedule preference of travelers. In our model, the equilibrium travel time profile would always be piecewise linear (over departure time, as shown in Fig. 3). This determines the marginal travel time increase or decrease with respect to departure time.

[11] To identify the critical case, we can start by selecting a t_s, which is close to $t*$ (but $t_s + \hat{\tau}(t_s) < t^*$), and estimating the traffic pattern according to the estimation procedure described in Section 3.3 (given in Appendix A). Note that in this case N is not fixed and indeed dependent on t_s. As a larger t_s

user equilibrium solution. We would like to mention that excluding the extreme cases with $\frac{N}{N_p} \to 1 - p_0$ and $\hat{\tau}(t_e) \to \infty$ for equilibrium analysis is reasonable. If the cruising time is extremely large (travel cost becomes very large), from a long-term perspective, people will shift to public transportation with lower travel cost. Thus, the equilibrium $\hat{\tau}(t_e)$ will be bounded and never reach infinity.

We conjecture that the dynamic user equilibrium is unique given the existence. This is explained as follows. Suppose we know the peak start time t_s. Firstly, given t_s, the time t_e can be uniquely determined. This is because, for given N and N_p, $\hat{\tau}(t_s)$ and $\hat{\tau}(t_e)$ are fixed, and the first and last travelers should have identical travel cost (t_e can be fully determined by $\hat{\tau}(t_s)$, $\hat{\tau}(t_e)$ and t_s through the equation of identical travel cost). Secondly, given t_s and t_e, due to the monotonicity of Eq. (2), Eq. (5), Eq. (6), Eq. (7) and Eq. (8), the traffic pattern (both departure and arrival) can be uniquely determined. However, the monotonicity of Eq. (6) and Eq. (8) relies on that during the peak, we always have $n \geq n_c$and $n < n_{jam}$. More specifically, the monotonicity of Eq. (6) relies on that $v(n)$ decreases when $n \in [n_c, n_{jam})$ (assumption A1); and the monotonicity of Eq. (8) relies on that $P(n) = n \cdot v(n)$ decreases when $n \in [n_c, n_{jam})$ (assumption A2). Note that empirical observations indicate these two assumptions (on monotonicity) to be reasonable. Also note that, every t_s corresponds to a specific travel demand N (an earlier t_s indicates a larger N, we observe this from numerical experiments), which is similar to Vickrey's bottleneck model. This means that with given N we then have a unique solution of t_s.

3.3. Computing user equilibrium

We now discuss how to compute the user equilibrium solution. From Eq. (6), we see that $\tau(n(t), p(t)) = L(p(t))/v(n(t)) = \hat{\tau}(t)$. As we know $(\hat{\tau}(t))^*$ from Eq. (12), the equilibrium $n(t)$ can be determined as follows:

$$(n(t))^* = v^{-1}\left(L(p(t)) \cdot \left[\left(\hat{\tau}(t)\right)^*\right]^{-1}\right), \tag{14}$$

where $v^{-1}(\cdot)$ is the inverse function of $v = v(n)$ over $[n_c, n_{jam})$. Later for ease of presentation, we may simply use "equilibrium condition" to refer to Eq. (14). But indeed, Eq. (14) is obtained by integrating the time-distance-speed relation in Eq. (6) into equilibrium travel time profile in Eq. (12) (derived from equilibrium condition in Eq. (11) and boundary conditions).[12]

The estimation of network outflow $\hat{o}(t)$ relies on Eq. (8), which also closely relates to $n(t)$. The conservation of traffic requires that

$$\frac{dn(t)}{dt} = \frac{dI(t)}{dt} - \hat{o}(t). \tag{15}$$

It is worth mentioning that $A(t)$ is the cumulative arrival of travelers belonging to the demand N. For $t \in [t_s + \hat{\tau}(t_s), t_e + \hat{\tau}(t_e)]$, we have $dA(t)/dt = \hat{o}(t)$, while for $t < t_s + \hat{\tau}(t_s)$, $A(t) = 0$ since no commuter in the total demand N has arrived at the parking spaces yet.

Computing the Dynamic User Equilibrium solution is more challenging compared to the traditional analysis of Vickrey's bottleneck model. This is mainly due to the time-dependent traffic conditions and cruising-for-parking. The estimation of solution relies on simultaneously solving a system of equations (over the time horizon): equilibrium condition, traffic flow dynamics, traffic conservation, parking dynamics. We discuss the estimation of the equilibrium solution in the following.

The estimation can be, technically, divided into two levels. i) The first level is to estimate the peak start time t_s, given equilibrium is achieved for travelers departing at every $t \in [t_s, t_e]$. Note that given t_s, time point t_e is determined in the second level based on $n(t_e) = n_c$ (as discussed, this is required for equilibrium). For an intermediate t_s during the estimation, it is not necessary that $I(t_e) = N$ (this should hold at the equilibrium solution). And the gap between $I(t_e)$ and N gives us information on how to adjust t_s. Simply speaking, we adjust t_s in the way such that $I(t_e)$ approaches the target demand N. ii) The second level is to estimate, given t_s from the first level, the cumulative inflow and outflow, and the rates of inflow and outflow, parking vacancy rates, accumulation and speed for every $t \in [t_s, t_e]$, based on equilibrium condition in Eq. (14), traffic flow dynamics in Eq. (8), traffic conservation in Eq. (15), parking dynamics in Eq. (2) and Eq. (7).

More details of the estimation procedure are provided in Appendix A. Roughly speaking, the first and second levels correspond to mainly **Step 1** and **Step 2** in the estimation procedure described in Appendix A. Moreover, we have highlighted places where the mentioned equations of equilibrium condition, traffic flow dynamics, traffic conservation and parking dynamics are exactly used in Appendix A. We suggest readers to combine with the flowchart in Fig. 4 when reading the estimation procedure.[13]

associates with less traffic (similar to Vickrey's model), in the beginning, $\hat{\tau}(t_e) \leq \hat{\tau}(t_\mu)$ would hold. We then decrease t_s (N will increase, as well as $\hat{\tau}(t_e)$) until $\hat{\tau}(t_e) = \hat{\tau}(t_\mu)$.

[12] In Geroliminis and Levinson (2009), since $p(t)$ is constant (without consideration of cruising-for-parking), the equilibrium $n(t)$ can be determined explicitly with Eq. (14) given $(\hat{\tau}(t))^*$ in Eq. (12). However, in this study, $p(t)$ is dependent on the cumulative departure $I(t)$, thus is related to $n(t)$ profile over the time interval $[t_s, t]$. Therefore, $n(t)$ and $p(t)$ have to be jointly estimated.

[13] The time horizon is discretized into small intervals with identical length of δt for numerical computation. We adopt $\varepsilon = 10^{-3}$ and $\delta t = 0.1$(min) in this paper for numerical analysis.

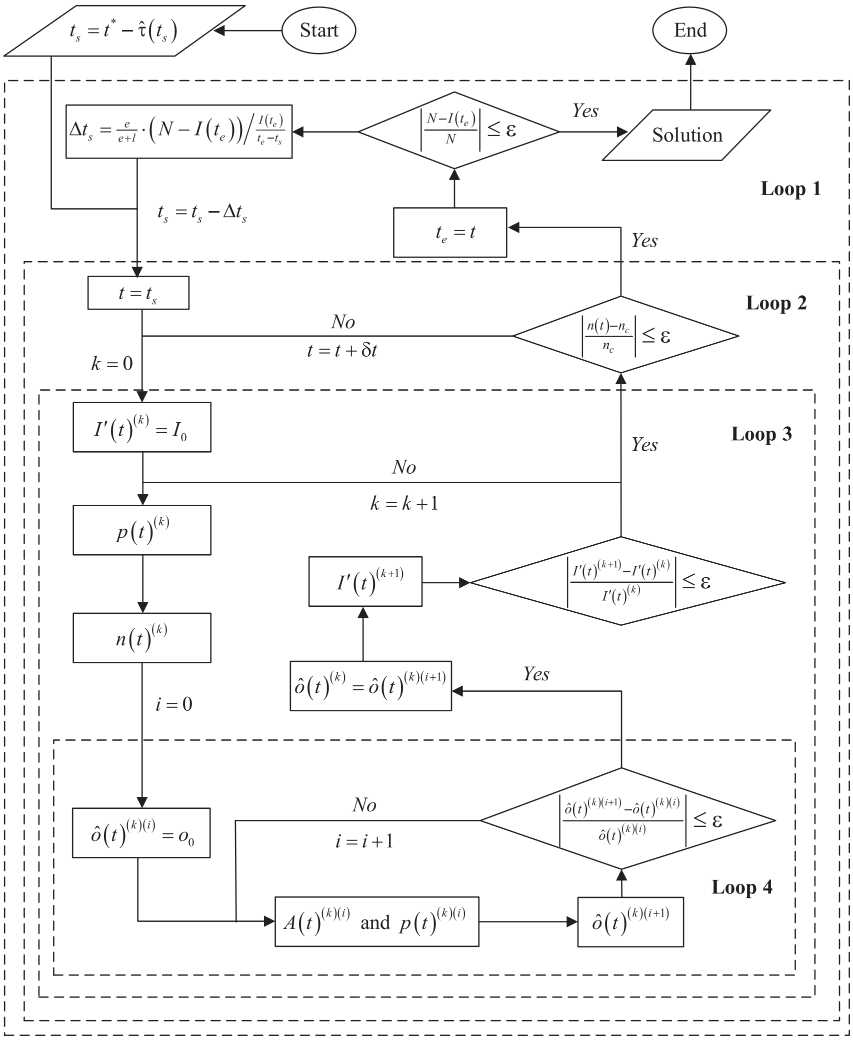


Fig. 4. The estimation procedure for User Equilibrium.

4. Optimal time-varying pricing for the downtown network

The user equilibrium experiences (hyper)congestion. And there are travel delays due to roadway congestion (high accumulation because of concentrated schedule preference and cruising-for-parking), and increased schedule delays due to competition for smaller cruising distance for parking. We now introduce a time-varying toll to minimize total social cost including travel time cost and schedule delay cost. We consider that, a toll is a transfer of money from travelers to government, but not social cost.

4.1. System optimum

We briefly show in Appendix B that, for a single-region network, the system optimum (total social cost consists of travel time cost and schedule delay cost is minimized) must occur at when the network is operating at the maximum production of the MFD during the peak,[14] i.e., $n(t) = n_c$ and $\hat{v}(t) = v(n_c)$, and $P(t) = n_c \cdot v(n_c)$.[15] However, this result relies on the network MFD specification in Section 2 (assumptions A1-A2), i.e., when the production is maximized (when $n = n_c$), the speed is also at its maximum. Also, the proof presented in Appendix B takes advantage of our formulation with “instantaneous

[14] Note that the production is the veh-km travelled (of all the traffic in the network) per unit time, while the outflow is the rate of trip endings. The two quantities are associated with the trip length. When the trip length is time-dependent (depends on cruising for parking), even if the production of the network reaches the maximum at n_c (as assumed in Section 2), network outflow might not be at its maximum, as a longer trip length will lead to smaller outflow.

[15] As shown later, to minimize schedule delay cost, an appropriate time interval of arrival should be chosen, as well as the departure.

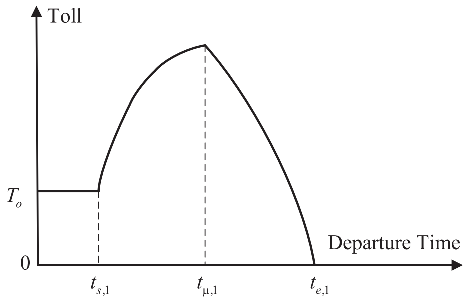


Fig. 5. The time-varying toll supporting $n(t) = n_c$.

travel time" (which means that assumptions A3-A10 are adopted). This result is consistent with those described in Daganzo (2007), Gonzales and Daganzo (2012). If these assumptions are not valid, the system optimum might occur at $n < n_c$, which is explained in the following.

Suppose that the network exhibits a different MFD than that assumed in the current paper (then assumptions A1-A2 will not hold anymore), under which the speed is decreasing over n even for n under critical accumulation n_c. The speed is then not at its maximum when the production is maximized. If the network is operating at some $n < n_c$ where $v(n) > v(n_c)$, even if the production and outflow might be smaller than those under $n = n_c$ (schedule delay will increase), the savings in travel delay (a higher speed) might be significant enough (overweighs the increase in schedule delay cost) to reduce total travel cost. Furthermore, for multi-region cities, even if different regions have MFDs similar to that assumed in this study, more complex control strategies to coordinate different regions have to be introduced (see, e.g., Haddad et al., 2013; Ramezani et al., 2015) to achieve the multi-region system optimum.

We now develop the optimal time-varying toll (note that how to implement such a toll has to be further investigated). Let $T(t)$ be the toll for the commuters departing from home at time t, individual full trip cost including the toll can be written as follows:

$$c(t, t^*) = c_w \cdot \tau(n(t), p(t)) + c_s \cdot (t^* - t - \tau(n(t), p(t))) + T(t). \tag{16}$$

Similar to the User Equilibrium case, we take the first-order derivative of Eq. (16) with respect to t, and let it be zero, then we have

$$\frac{\partial \tau(n(t), p(t))}{\partial n} \cdot \frac{dn(t)}{dt} + \frac{\partial \tau(n(t), p(t))}{\partial p} \cdot \frac{dp(t)}{dt} + \frac{1}{c_w - c_s}\frac{dT(t)}{dt} = \frac{c_s}{c_w - c_s}. \tag{17}$$

Suppose under the time-varying toll, the peak starts at $t_{s,\,1}$, of which the estimation will be discussed later. For $t \leq t_{s,\,1}$ we set $T(t) = T_0$. After $t_{s,\,1}$, to achieve the system optimum, we maintain $n(t) = n_c$, $dn(t)/dt = 0$. By adding this condition into Eq. (17), we have

$$\frac{dT(t)}{dt} = c_s - (c_w - c_s) \cdot \frac{\partial \tau(n_c, p(t))}{\partial p} \cdot \frac{dp(t)}{dt}. \tag{18}$$

With $\tau(n(t), p(t)) = L(p(t))/v(n(t))$, Eq. (18) can be immediately written as

$$\frac{dT(t)}{dt} = c_s - (c_w - c_s) \cdot \frac{1}{v(n_c)} \cdot \frac{dL(p(t))}{dp(t)} \cdot \frac{dp(t)}{dt}. \tag{19}$$

With Eq. (19), we can derive the time-varying toll to support $n(t) = n_c$ during the peak, which is given as follows

$$T(t) = \begin{cases} T_0 & \text{for } t < t_{s,1} \\ T_0 + e \cdot (t - t_{s,1}) - T_p(t) & \text{for } t_{s,1} \leq t \leq t_{\mu,1} \\ T\left(t_{\mu,1}\right) - l \cdot \left(t - t_{\mu,1}\right) - T_p(t) & \text{for } t_{\mu,1} < t \leq t_{e,1} \\ T(t_{e,1}) & \text{for } t > t_{e,1} \end{cases}, \tag{20}$$

where $T_p(t)$ is

$$T_p(t) = \begin{cases} (c_w - e) \cdot \frac{L(p(t)) - L(p(t_{s,1}))}{v(n_c)} & \text{for } t_{s,1} \leq t \leq t_{\mu,1} \\ (c_w + l) \cdot \frac{L(p(t)) - L\left(p\left(t_{\mu,1}\right)\right)}{v(n_c)} & \text{for } t_{\mu,1} < t < t_{e,1} \end{cases}, \tag{21}$$

and $t_{\mu,\,1}$ is the departure time of the on-time traveler and $t_{e,\,1}$ is the latest departure time. For $t > t_{e,\,1}$ we set $T(t) = T(t_{e,1})$.

Fig. 5 shows the pattern of the optimal time-varying toll when the minimum toll is zero, i.e., we let $T(t_{e,1}) = 0$.[16] The toll is non-linear over time since the impact of cruising is generally non-linear over time, i.e., $T_p(t)$ is nonlinear over time.[17] Indeed, $-T_p(t)$ in the toll $T(t)$ is to compensate the loss of a later departing traveler with a larger cruising distance. If we look into Eq. (21), the compensation for one unit of additional travel time is $c_w - e$ for early arrival traveler; and is $c_w + l$ for late arrival travelers. This is because, one unit of additional travel time indicates, besides one more unit of travel delay penalty (c_w), either one unit less early arrival penalty ($-e$) or one unit more late arrival penalty ($+l$).

Furthermore, the first commuter would experience a higher toll than the last commuter, i.e., $T_o > 0$ will hold if we let $T(t_{e,1}) = 0$, as shown in Fig. 5. This T_o is to prevent travelers from departing too earlier and enjoy less cruising, thus to reduce the additional schedule delay cost.

Besides the toll to support $n(t) = n_c$, we also need to appropriately choose the $t_{s,\,1}$ for the system optimum, which is related to schedule delay cost. For different but given $t_{s,\,1}$, if we implement the toll design in Eq. (20), we will obtain identical departure/arrival traffic pattern where $n(t) = n_c$ (however, they start at different $t_{s,\,1}$). The total travel time cost of all travelers will be identical. However, different $t_{s,\,1}$ associates with different schedule delay costs. To minimize total social cost, we then have to choose an appropriate $t_{s,\,1}$ to minimize schedule delay cost, which is to solve the following problem:

$$\min : SC(t_{s,1}) = \int_0^{\Delta t} \tfrac{dI(x)}{dx} \cdot c_s \cdot (t^* - (t_{s,1} + x) - \tau_{so}(x))dx, \tag{22}$$

where Δt is the length of the travelers' departure duration, $I(x)$ is the cumulative departure (inflow) at time $t_{s,1} + x$, and $\tau_{so}(x)$ is the travel time for travelers departing at time $t_{s,1} + x$. Note that as the departure/arrival patterns are identical under different $t_{s,\,1}$, $I(x)$, $\tau_{so}(x)$ and Δt in Eq. (22) are independent of $t_{s,\,1}$. For $t_{s,1} \neq t'_{s,1}$, if we implement the toll design in Eq. (20), we would have

$$t_{e,1} = t_{s,1} + \Delta t, t'_{e,1} = t'_{s,1} + \Delta t. \tag{23}$$

For $t \in [0, \Delta t]$, we have $n(t_{s,1} + t) = n({t'}_{s,1} + t) = n_c$, it follows

$$I(t_{s,1} + t) = I\left({t'}_{s,1} + t\right), p(t_{s,1} + t) = p\left({t'}_{s,1} + t\right), \tilde{p}(t_{s,1} + t) = \tilde{p}\left({t'}_{s,1} + t\right). \tag{24}$$

Since $\hat{v}(t_{s,1} + t) = \hat{v}({t'}_{s,1} + t) = v(n_c)$, we further have

$$\hat{L}(t_{s,1} + t) = \hat{L}\left({t'}_{s,1} + t\right), \hat{o}(t_{s,1} + t) = \hat{o}\left({t'}_{s,1} + t\right), \hat{\tau}(t_{s,1} + t) = \hat{\tau}\left({t'}_{s,1} + t\right). \tag{25}$$

However, the toll pattern over time would be different, i.e., $T(t_{s,1} + t) \neq T({t'}_{s,1} + t)$.

Taking the first order derivative of the objective function in Eq. (22) with respect to $t_{s,\,1}$, we have

$$\frac{dSC(t_{s,1})}{dt_{s,1}} = \int_0^{\Delta t} d\frac{\frac{dI(x)}{dx} \cdot c_s \cdot (t^* - (t_{s,1} + x) - \tau_{so}(x))}{dt_{s,1}}dx = \int_0^{\Delta t} -\frac{dI(x)}{dx} \cdot c_s dx. \tag{26}$$

Let $\Delta t_{\mu,1} + \tau_{so}(\Delta t_{\mu,1}) = t^* - t_{s,1}$, then $\Delta t_{\mu,\,1}$ corresponds to the on time traveler, and it can be verified that $d\Delta t_{\mu,\,1}/dt_{s,\,1} < 0$. With $\Delta t_{\mu,\,1}$, Eq. (26) can be rewritten as

$$\frac{dSC(t_{s,1})}{dt_{s,1}} = -e \cdot \int_0^{\Delta t_{\mu,1}} \frac{dI(x)}{dx}dx + l \cdot \int_{\Delta t_{\mu,1}}^{\Delta t} \frac{dI(x)}{dx}dx. \tag{27}$$

We then look at the second order derivative of the objective function in Eq. (22), which is given as

$$\frac{d^2SC(t_{s,1})}{d(t_{s,1})^2} = -e \cdot \left(\frac{dI(x)}{dx}\bigg|_{x=\Delta t_{\mu,1}} \right) \cdot \frac{d\Delta t_{\mu,1}}{dt_{s,1}} - l \cdot \left(\frac{dI(x)}{dx}\bigg|_{x=\Delta t_{\mu,1}} \right) \cdot \frac{d\Delta t_{\mu,1}}{dt_{s,1}} > 0. \tag{28}$$

Total schedule delay cost is minimized if we let Eq. (27) be zero, i.e.,

$$\frac{dSC(t_{s,1})}{dt_{s,1}} = 0 \Leftrightarrow \frac{\int_0^{\Delta t_{\mu,1}} \frac{dI(x)}{dx}dx}{\int_{\Delta t_{\mu,1}}^{\Delta t} \frac{dI(x)}{dx}dx} = \frac{N_e}{N_l} = \frac{l}{e}, \tag{29}$$

which says the early arrival traffic N_e should be l/e times as much as the late arrival traffic N_l. This is consistent with both the case without cruising and the case in Vickrey's bottleneck model.

[16] We would like to point out that the maximum toll might not always be experienced by the on time travelers (however, Fig. 5 takes this case as an illustrative example). This is explained as follows. Eq. (19) is the first-order derivative of the toll with respect to time. For some $t < t_\mu$, cruising might be already very costly, thus $|\frac{dL(p(t))}{dp(t)} \cdot \frac{dp(t)}{dt}|$ can be large, and $\frac{dT(t)}{dt} < 0$ holds, which indicates that the toll will start to decrease before t_μ. However, at the system optimum, the percentage of vacant parking spaces can be relatively large when the on-time traveler departs from home (early traffic is still much less than the parking capacity), cruising is then not very significant. Thus, we have $\frac{dT(t)}{dt} \geq 0$ for $t < t*$. Besides, for $t > t*$, it can be easily verified that $\frac{dT(t)}{dt} < 0$. In this case, the maximum toll will arise at time t_μ.

[17] If approximating $\tilde{p}(t)$ with $p(t)$, and noting the cumulative departure (network inflow) is parallel to the cumulative arrival (network outflow) at the system optimum, it can be shown that the time-varying toll should be concave over time as $L(p(t))$ is convex over time.

For given N and N_p, denote the $t_{s,1}$ under which the last traveler just arrives on time by $t^1_{s,1}$, then $t^1_{s,1} = t^* - \hat{\tau}(t_e) - \Delta t$, and denote the $t_{s,1}$ under which the first traveler just arrives on time by $t^2_{s,1}$, then we have $t^2_{s,1} = t^* - \hat{\tau}(t_s)$. It can be shown that the derivative in Eq. (27) will be negative when $t_{s,1} = t^1_{s,1}$, and will be positive when $t_{s,1} = t^2_{s,1}$. Given Eq. (28), the value of $t_{s,1}$ that solves Eq. (29) will be within $[t^1_{s,1}, t^2_{s,1}]$, which means that the optimal $t_{s,1}$ is both lower and upper bounded. The lower and upper bounds will be utilized to compute the optimal $t_{s,1}$, which solves Eq. (29), as discussed next.

4.2. Computing system optimum

Principally, the procedure for computing the system optimum solution is similar to that for user equilibrium solution. This is because, traffic flow dynamics in Eq. (8), traffic conservation in Eq. (15), and parking dynamics in Eq. (2) and Eq. (7) are still valid. The major difference is that, for the system optimum, Eq. (14) representing equilibrium condition (for user equilibrium in Section 3) will no longer hold. Instead, we know that the time-dependent accumulation should remain constantly at the critical value, i.e., $n(t) = n_c$. And we have an additional equation, i.e., Eq. (20), to compute the additional time-dependent variable, i.e., the toll $T(t)$, which is to support $n(t) = n_c$, as an equilibrium.

The estimation of system optimum can also be divided into two levels. i) In the first level, we use a bi-section based approach to determine the departure time of the first traveler $t_{s,1}$, given that the traffic pattern with $n(t) = n_c$ for $t \in [t_{s,1}, t_{e,1}]$ is achieved in the second level. Specifically, the determination of $t_{s,1}$ takes advantage of the result in Eq. (29), which states that the early arrival traffic N_e should be l/e times as much as the late arrival traffic N_l. ii) In the second level, given $t_{s,1}$ from the first level, we estimate the time-varying variables, i.e., $n(t)$, $\hat{v}(t)$, $I(t)$, $p(t)$, $A(t)$, $\tilde{p}(t)$, $\hat{o}(t)$ and $T(t)$ for all $t \in [t_{s,1}, t_{e,1}]$, where $t_{e,1}$ is determined based on $I(t_{e,1}) = N$.

To be more specific, firstly, we choose the initial lower bound $t^l_{s,1} = t^1_{s,1}$ and upper bound $t^u_{s,1} = t^2_{s,1}$ for $t_{s,1}$, and set $t_{s,1} = \frac{1}{2}(t^l_{s,1} + t^u_{s,1})$ as an initial solution. Given $t_{s,1}$, we can estimate all the time-varying variables with a similar approach as the estimation for User Equilibrium starting from $t = t_{s,1}$ and ending $t = t_{e,1}$ where $I(t_{e,1}) = N$. However, instead of computing $n(t)$ directly through Eq. (14) (**Step 2-1-2** in Appendix A for computing User Equilibrium), we firstly compute the toll $T(t)$ through Eq. (20), and then compute the $n(t)$ as follows:

$$n(t) = v^{-1}\left(L(p(t)) \cdot \left(\int_{t_{s,1}}^{t} \frac{c_s}{c_w - c_s} - \frac{1}{c_w - c_s}\frac{dT(w)}{dw}dw\right)^{-1}\right). \tag{30}$$

Eq. (30) comes from Eq. (17), i.e., the equilibrium condition with tolling introduced. Note that as $T(t)$ in Eq. (20) is determined from letting $n(t) = n_c$, Eq. (30) will give us $n(t) = n_c$. The other parts of estimation follow those for user equilibrium. However, when checking whether the current $t_{s,1}$ is the solution or not, we compare the numbers of early and late traffic N_e and N_l. If $N_e/N_l < l/e$, it means the current $t_{s,1}$ is too large, then we can set $t^u_{s,1} = t_{s,1}$, and update $t_{s,1} = \frac{1}{2}(t^l_{s,1} + t^u_{s,1})$; If $N_e/N_l > l/e$, it means the current $t_{s,1}$ is too small, then we can set $t^l_{s,1} = t_{s,1}$, and update $t_{s,1} = \frac{1}{2}(t^l_{s,1} + t^u_{s,1})$. The System Optimum solution is achieved as $|(t^u_{s,1} - t^l_{s,1})/t^u_{s,1}| < \varepsilon$. We adopt $\varepsilon = 10^{-3}$ in this paper for numerical analysis. The convergence of the estimation procedure is numerically illustrated in Appendix C.

4.3. Approximate solution for p(t) *under the system optimum*

The following analysis provides an approximate closed-form solution for $p(t)$ at the System Optimum. It is mentioned that $p(t) = 1 - \frac{I(t)}{N_p} - p_0$. With Eq. (15) and $n(t) = n_c$at System Optimum, we have

$$\frac{dp(t)}{dt} = -\frac{1}{N_p} \cdot \frac{dI(t)}{dt} = -\frac{1}{N_p} \cdot \left(\frac{dn(t)}{dt} + \hat{o}(t)\right) = -\frac{1}{N_p} \cdot \hat{o}(t). \tag{31}$$

With $\hat{o}(t) = o(n(t), \tilde{p}(t))$, and with $p(t)$ to approximate $\tilde{p}(t)$, i.e., assume $p(t) = p(\tilde{t}) = \tilde{p}(t)$, we have

$$\frac{dp(t)}{dt} = -\frac{1}{N_p}\frac{n_c \cdot v(n_c)}{L(p(t))}. \tag{32}$$

Since $L(p(t)) = l_m + d/p(t)$, after some manipulations of Eq. (32), we obtain

$$l_m \cdot p(t) \cdot \frac{dp(t)}{dt} + d \cdot \frac{dp(t)}{dt} + \theta \cdot p(t) = 0. \tag{33}$$

where $\theta = n_c \cdot v_f / N_p$. We then have

$$p(t) = \frac{d}{l_m} W\left(\frac{l_m}{d}\exp\left(\frac{k_1 - \theta t}{d}\right)\right), \tag{34}$$

where $W(\cdot)$ is the inverse function of $f(W) = W \cdot e^W$, and k_1 is determined by $p(t_{s,1}) = 1 - p_0$. Eq. (34) implies that, with information of d, l_m, $p(t_{s,1}) = 1 - p_0$, n_c, $v(n_c)$ and N_p, we can estimate the shape of the parking vacancy profile over time in the system optimum.

Table 1
Summary of values of parameters and variables.

Parameters or Functions	Specification
Travel demand	$N = 6000$
Initial parking occupancy rate	$p_0 = 0\%$[a]
Total parking capacity	$N_p = 6500$ if not specified, indicating parking occupancy rate at peak end is 92.31%.
Critical accumulation	$n_c = 1000$ (veh)
Travelling Speed	$v(n) = v_0 \cdot e^{-v_1 \cdot n}$ (km/h) for $n \geq n_c$[b] $v(n) = v(n_c)$ (km/h) for $n < n_c$
Speed function parameters	$v_0 = 68$, $v_1 = 10^{-3}$, indicating $v_f = 25$(km/h)
Trip distance function parameters	$l_m = 5$ (km) and $d = 0.2$(km)
Value of travel time	$c_w = 9.91$ (EUR\$)
Early arrival penalty	$e = 4.66$ (EUR\$)
Late arrival penalty	$l = 14.48$ (EUR\$)
Desired arrival time	$t^* = 200$(min)

Note: c_w, e and l are from Tseng et al. (2005).
[a] Our numerical analysis assumes fully empty parking spaces at the peak start. However, the cruising effect can be significant even for people traveling very early if we adopt a larger p_0, and traffic efficiency improvement through pricing can be larger as the potential for reducing cruising time and traffic congestion is larger.
[b] This speed specification indicates that we have $n_{jam} \to +\infty$.

5. Numerical studies

In this section, we conduct some numerical experiments to illustrate and verify the models and analysis in the previous sections. Table 1 summarizes the values of parameters and variables valid for the analysis. We consider the downtown network with a total peak travel demand of $N = 6000$(veh), and a critical accumulation of $n_c = 1000$(veh), and a maximum speed of 25(km/h). The travel demand can be generated at both the boundary and interior of the downtown network. The peak travel demand is six times of critical accumulation, i.e., $N = 6n_c$. This indicates that, if we consider parking capacity is infinity, i.e., $p = 1$, at the System Optimum, i.e., $n(t) = n_c$ and $\hat{v}(t) = v_f$, the departure or arrival duration of travelers in the peak will be six times of free-flow travel time.[18]

Travelers' moving distance before starting to find a parking space is $l_m = 5$km, and the distance traveled in each trial to find a space is $d = 0.2$km. A single distance d is negligible when compared to l_m. However, when parking availability is low, even a very small d can lead to a very long cruising distance l_s. A larger d indicates more cruising for parking and longer trip length, which leads to more travel time and schedule delays (this can be verified with numerical experiments).[19]

5.1. User equilibrium

5.1.1. Time-varying traffic and parking vacancy under user equilibrium

The parking capacity is $N_p = 6500$ in the benchmark case. Fig. 6(a) presents the cumulative departure and arrival, i.e., $I(t)$ and $A(t)$, at equilibrium, while Fig. 6(b) presents the inflow (departure rate) and outflow (arrival rate). Fig. 6(c) and Fig. 6(d) depict the time-varying accumulation and associated traveling speed, and the parking vacancy rate and associated trip distance respectively.

As shown in Fig. 6(b), the network outflow (traffic exiting the network per unit time) decreases after the start of the peak and until $t_\mu = 149.5$(min), which is partly due to the increasing accumulation and decreasing speed as shown in Fig. 6(c)), and partly due to the decreasing parking availability and increasing trip length as shown in Fig. 6(d). After $t_\mu = 149.5$(min), the accumulation starts to decrease, which leads to increase in outflow shown in Fig. 6(b). The impact of decreasing accumulation (less congestion) overweighs the impact of the decreasing parking availability (longer trip length). However, the network outflow at the end of the peak cannot go back to the level at the peak start, owing to the decreased parking availability (100% $>$ 7.76%) and increased trip length (7.58 km $>$ 5.20 km) as shown in Fig. 6(d). Besides, as travelers can

[18] At User Equilibrium with finite parking capacity, the peak duration will be even larger (more than six times of free-flow travel time with infinite parking capacity). By doing so, we try to reduce the impacts of the boundary traffic conditions at the peak start and end, as the peak hour is much longer than the travel time for the travelers departing at the peak start or end.

[19] Note that d is not simply the spacing between two adjacent parking spaces (e.g., less than 10 meters), but the average spacing of all the parking spaces near the destination, which depends on specific parking distribution over the streets and the network topology. For example, travelers heading for a specific building will consider to park at curbside spaces just around the building, or one street away (e.g., 200 meters), or two streets away (e.g., 400 meters). To check parking availability at different locations, cruising travelers have to drive to different streets (even worse, these streets might not be in the same direction). Also, even for parking spaces quite close to drivers but on the opposite side of the street, drivers might have to cover a long distance, e.g., make a U-turn. More importantly, as travelers have the incentive to park close to the building they are heading for, they may cruise around the building and wait for vacancies. This can induce more cruising. While this study relies on average and deterministic cruising distance based on parking availability for insights, simulation approach might be considered to integrate detailed cruising behavior and stochasticity.

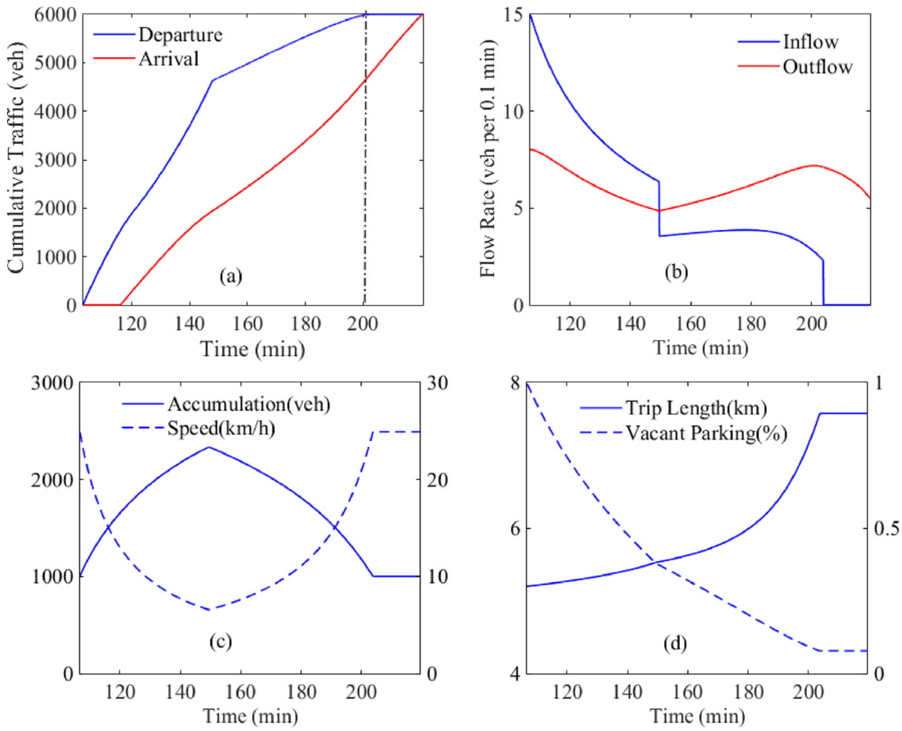
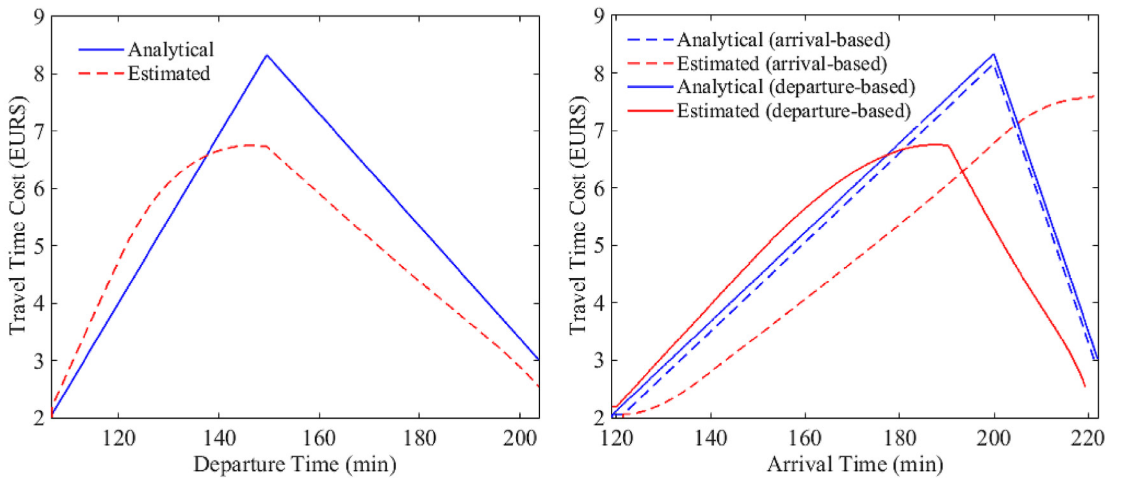


Fig. 6. User Equilibrium: (a) flow pattern; (b) inflow and outflow; (c) speed vs. accumulation; (d) trip length vs. vacant parking.



(a) Travel time cost varies over departure time (b) Travel time cost varies over arrival time

Fig. 7. Travel Time Cost (analytical vs. estimated).

enjoy smaller cruising distance (as well as the whole trip length) by departing earlier, there are more early arrival traffic ($N_e/N_l = 3.7$) when compared to the case with infinite parking capacity and zero cruising ($N_e/N_l = 2.4$). Note that in Vickrey's model, $N_e/N_l = l/e = 3.1$.

As discussed in Section 2, we approximate travelers' travel time with instantaneous speed and parking availability. Fig. 7(a) displays the difference (in the unit of monetary cost) between the analytical travel time (based on instantaneous speed and parking availability, which is given in Eq. (6)) and "estimated" travel time (obtained by calculating the horizontal gap between the cumulative departure and arrival depicted in Fig. 6(a), which is described in subsection 2.2, after A7). Note that, the travel time cost profile over the time horizon in Fig. 7(a) is arranged in the order of departure time. As can be

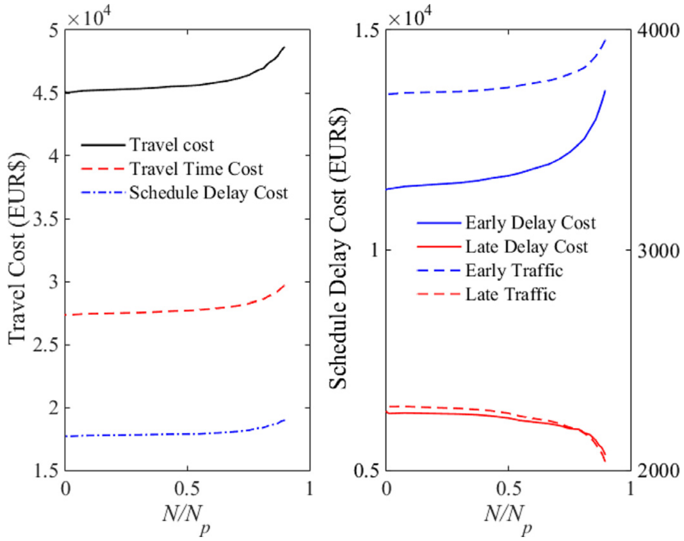


Fig. 8. Costs and early/late arrival traffic vary with parking capacity.

seen in Fig. 7(a), the "estimated" travel time is nearly triangular over the time horizon and reaches its maximum at time t_{μ}= 149.5(min). This means, with our "instantaneous model", the "estimated" travel time exhibits similar pattern (over time horizon) with the analytical travel time. Furthermore, the discrepancies of system efficiency measures (e.g., total travel cost, total travel time, total schedule delay) between the analytical and estimated values are generally small (ranging from $\pm$ 5% to $\pm$ 10%, dependent on the parking capacity).

Besides estimating the user equilibrium solution with departure-time-based accumulation and speed, we can estimate the user equilibrium with arrival-time-based accumulation and speed (the approach is similar to that described in Section 3.3, and this can be achieved by implementing **footnote 7**). In this case, similarly, we can compare the difference between the analytical travel time and "estimated" travel time, which is shown in Fig. 7(b) (note that the travel time cost profile is arranged in the order of arrival time). We can also rearrange both the analytical and "estimated" travel time costs over departure time in Fig. 7(a) according to arrival time, which is also shown in Fig. 7(b). We see that the discrepancy between analytical and "estimated" travel times of late arrival commuters is significant when we use arrival-time-dependent accumulation and speed for estimation (note that the "estimated" travel time cost is almost an increasing function with time as shown in Fig. 7(b)). However, the "estimated" travel time still exhibits similar pattern (arranged in the order of arrival time) with the analytical travel time when we use departure-time-dependent formulation.

5.1.2. System performance under varying parking capacity

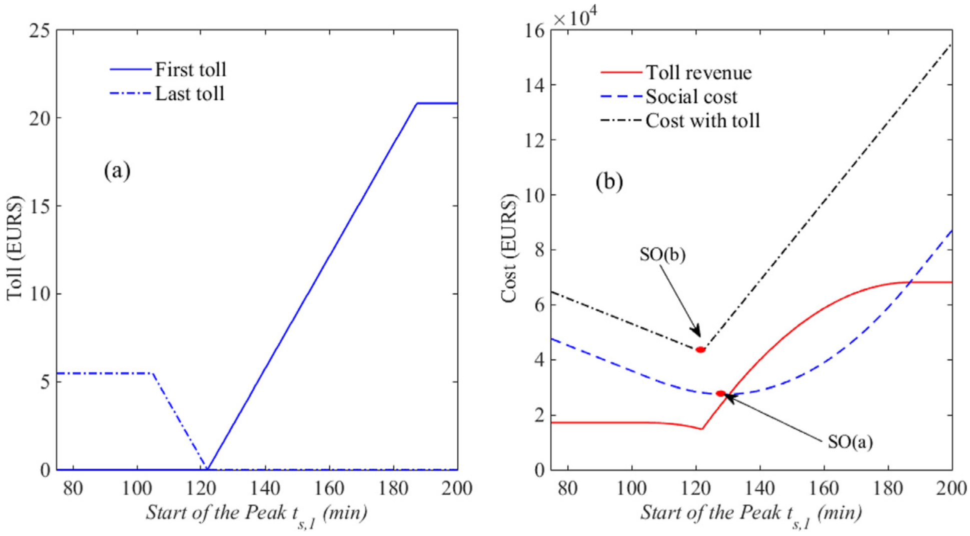
Given $N = 6000$, Fig. 8(a) depicts how total travel cost, travel time cost and schedule delay cost vary with the parking capacity N_p, while Fig. 8(b) shows how the number of early and late arrival traffic and schedule delay cost vary with the parking capacity N_p. The x-axis in Fig. 8 is the ratio of N/N_p, i.e. the parking occupancy rate at the end of the peak.

As N/N_p increases, i.e., parking supply decreases, the total travel cost, travel time and schedule delay will increase. This is because, less parking supply indicates longer trip length and thus smaller network outflow. Fig. 8(b) further shows that the schedule delay cost of early arrival increases more sharply than the total schedule delay cost when parking supply decreases. The reason is that there will be more early arrival traffic under less parking capacity (less late arrival traffic), and the schedule delay cost of late arrival will decrease. In Fig. 8(a), the case with $N/N_p \to 0$ corresponds to the situation with no additional cruising, where the total travel cost reaches its minimum (4.507×10^4 (EUR$)).

5.2. System optimum

5.2.1. Flow pattern and travel cost under system optimum

As discussed in Section 4, an appropriate $t_{s,\,1}$ should be chosen to minimize schedule delay cost, thus minimize total social cost. We start by considering that we can choose different $t_{s,\,1}$, and then develop the time-varying toll described in Eq. (20) such that $n(t) = n_c$. Also, we consider that the time-varying toll should be non-negative (no subsidy), and the minimum toll is set to be zero. This implies that we let $\min\{T_0, T(t_{e,1})\} = 0$. Then, Fig. 9(b) shows how the toll revenue, social cost, and travel cost including toll vary with $t_{s,\,1}$, while Fig. 9(a) shows the first and last tolls (tolls experienced by the first or last traveler respectively) under given $t_{s,\,1}$.



(a) First and Last toll (b) Toll revenue, social cost and total cost

Fig. 9. Toll revenue, social cost and total cost (the minimum toll is zero).

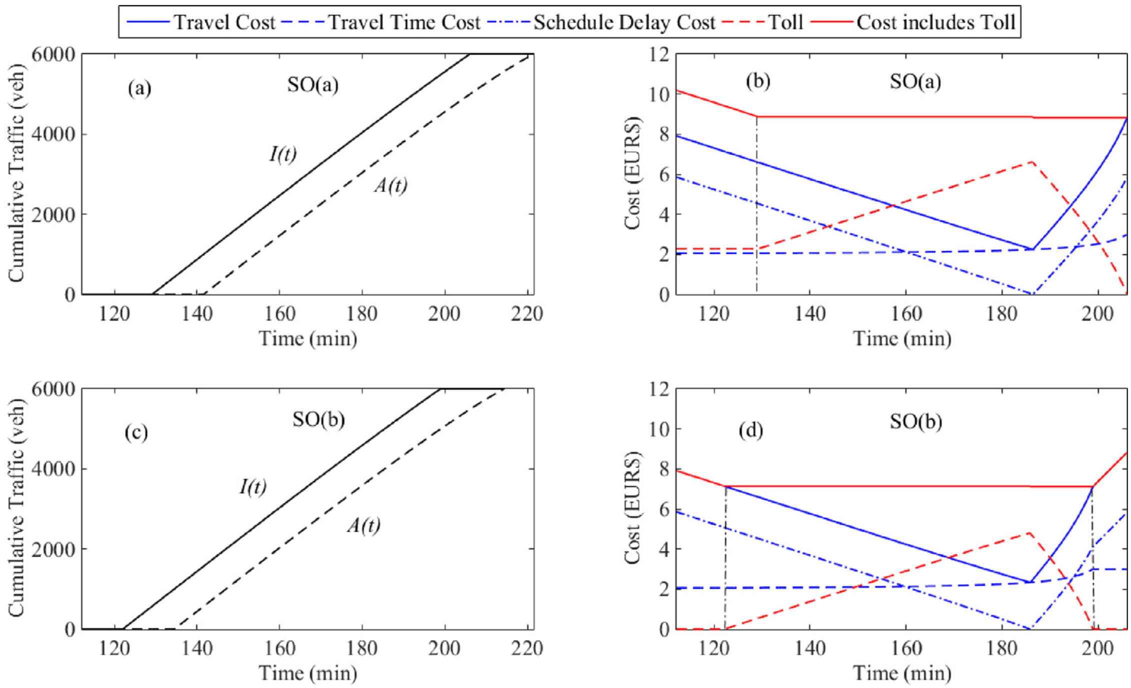


Fig. 10. Flow patterns and costs at SO(a) and SO(b) based on departure time.

As can be seen in Fig. 9(b), the social cost is minimized when $t_{s,1} = 129.3$(min), which is denoted by SO(a). This is exactly the system optimum solution discussed in Section 4 (to minimize total social cost, which consists of travel time cost and schedule delay cost), under which the ratio of early traffic to late traffic, N_e/N_l, is equal to $l/e = 3.1$. The first traveler would experience a higher toll than the last traveler, i.e., 2.28 (EUR$) > 0 (EUR$). The time-varying toll to support SO(a) is shown in Fig. 10.

Besides the SO(a) to minimize total social cost, we here define another optimum in terms of minimizing total cost including both social cost and the toll, which is denoted by SO(b) in Fig. 9(b). This SO(b) is achieved when $t_{s,1} = 122.1$(min).

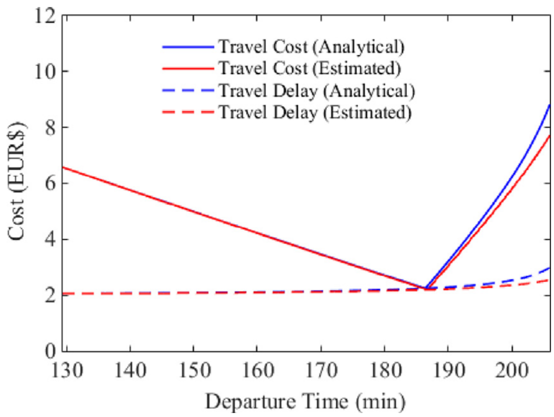


Fig. 11. Travel cost without toll and travel delay cost (analytical vs. estimated).

As can be seen in Fig. 9(b), the toll revenue is minimized at SO(b), and both the first toll (experienced by first traveler) and last toll (experienced by last traveler) are zero.

One can verify that the toll revenue is 25,580 (EUR$) at SO(a), which is 1.74 times of the toll revenue (14,710 (EUR$)) when total cost is minimized (SO(b)). However, by imposing the toll derived for SO(a) instead of SO(b) as shown in Fig. 10, the social cost can only be reduced from 28,060 (EUR$) to 27,490 (EUR$). This reduction is around 1.2% comparing to total travel cost under User Equilibrium (49,955 (EUR$)). This means, we may set a much lower toll other than that for SO(a) to achieve similar efficiency in reducing social cost.[20]

Moreover, the individual travel cost is 8.87 (EUR$) at SO(a), which is higher than the that at User Equilibrium (8.33(EUR$)); while the individual travel cost at SO(b) (7.14 (EUR$)) is less than that at User Equilibrium. This means that, at SO(a), since the toll is relatively high, travelers are worse off compared with the UE, although social cost decreases. To make every traveler better off, the system operators needs to refund travelers. However, the toll to support SO(b) is Pareto-improving although no one receives rebate from the toll revenue.

Fig. 10(b) and Fig. 10(d) respectively show the travel costs and tolls at SO(a) and SO(b), Figs. 10(a) and 10(c) show the cumulative departure and arrival at SO(a) and SO(b). In both SO(a) and SO(b), the cumulative departure and arrival are parallel to each other, thus network accumulation remains at the critical level, i.e., 1000 (veh), and the speed is at its maximum, i.e., 25 (km/h). Also, the cumulative departure/arrival patterns at SO(a) and SO(b) are exactly the same except that they start at different $t_{s,1}$. Furthermore, the slopes of the cumulative departure and arrival, i.e., inflow and outflow of the network, decrease over time (not very significant as traveling speed is at its maximum) as the parking availability decreases and trip length increases over time.

For the System Optimum case, as the accumulation and speed in the network remain constant during the departure/arrival of travelers, approximating the travel time with instantaneous speed (at departure time) does not lead to any inaccuracy. If we again compare the analytical travel delay with "estimated" travel delay (dashed lines in Fig. 11), we see much less discrepancy than that in the User Equilibrium. However, discrepancy still arises as we still approximate the travel time with instantaneous parking availability (which determines trip length).

Fig. 11 also displays the analytical and estimated travel costs without toll (solid lines in Fig. 11). The discrepancy between them follows the same trend as that of travel delay, which approaches zero for most of early arrival traffic, and approaches the maximum (valued at one EUR$ in the example, which is 11% of the individual travel cost shown in Fig. 10(b)) for the last traveler. Note that, the higher discrepancy is only for a small proportion of late arrival travelers (moreover, there are more early arrival travelers than late arrival travelers). Given these, we expect the time-varying toll from our solution to be a good estimate (note that the toll is equal to a constant value, i.e., full trip price, minus the cost without toll). In addition, the discrepancies in system efficiency measures such as total travel time, total travel cost and total schedule delay cost will be less than 3%. It is worth mentioning that, if parking capacity approaches infinity such that p approaches constant, the discrepancy will approach zero for system optimum solution. This is because, the inaccuracy from using instantaneous parking availability to estimate trip length diminishes.

5.2.2. Cases of "underpricing parking" and "underpricing travel"

Fig. 12 further shows the user costs and equilibrium accumulation profile over time under two different time-varying tolls (red dashed lines in Fig. 12(a) and 12 (c)) than the System Optimum toll (the toll to support SO(a), which is shown in

[20] This can be important in practice. Firstly, smaller toll may face less objection from the public, especially when the public can experience much less congestion. Secondly, the government can be wasteful in how it allocates the revenue it collects (thus we try to achieve similar efficiency with less toll revenue).

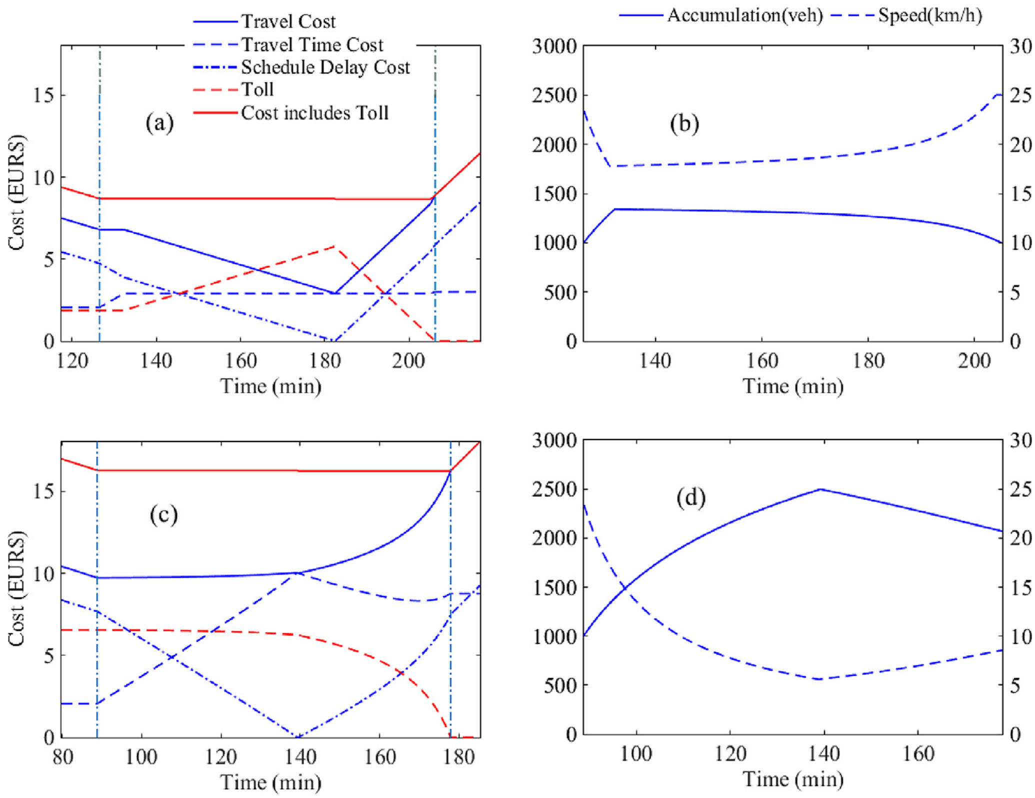


Fig. 12. User costs and accumulations: (a) costs and toll under travel pricing only; (b) accumulation under travel pricing only; (c) costs and toll under parking pricing only; (d) accumulation under parking pricing only.

Table 2
Various efficiency measures for three pricing cases.

	SO(a)	Travel pricing only	Parking pricing only
Social cost (10^4 EUR$)	2.749	3.180	6.191
Toll revenue (10^4 EUR$)	2.558	2.022	3.558
Moving time (10^5 min)	0.747	0.973	2.069
Cruising time (10^4 min)	0.510	0.618	1.741
Schedule (10^4 EUR$)	1.430	1.473	2.486
Early arrival (10^4 EUR$)	1.042	1.081	2.030
Late arrival (10^4 EUR$)	0.388	0.390	0.456
Ratio of N_e/N_l	3.1	3.1	3.1
Departure duration (min)	76.8	79.8	99.2

Fig. 10(b)). Specifically, the two different tolls are: a toll considering only the travel pricing (and neglecting parking), which corresponds to Fig. 12(a) and Fig. 12(b); and a toll considering only parking pricing (and neglecting travel pricing), which corresponds to Fig. 12(c) and Fig. 12(d). As can be seen in Fig. 12(a) and Fig. 12(c), travelers will depart between the two time points marked by the two vertical dash-dot lines (otherwise they will encounter a larger cost, see the red solid lines). Fig. 12(b) and Fig. 12(d) display the corresponding accumulation and speed profiles for the duration with departures (i.e., the duration between the two time pointes marked by the two vertical dash-dot lines in Fig. 12(a) and Fig. 12(c)). Moreover, Table 2 summarizes the relevant efficiency measures under different tolls.[21] By doing so, we aims to show the inefficiency due to either underpricing parking or underpricing travel.

[21] Note that to have a "fair" comparison, we adopt different peak start times for different tolls, i.e., the peak start time is chosen in the way that the ratio of early arrival traffic to late arrival traffic will be $N_e/N_l = l/e = 3.1$, which is also presented in Table 2.

Table 3
Various efficiency measures for different cases.

	UE with cruising	UE (no cruising)	SO(a) with cruising	SO(b) with cruising	SO (no cruising)
Social cost (10^4 EUR$)	4.996	4.507	2.749	2.806	2.553
Toll revenue (10^4 EUR$)	0	0	2.558	1.471	1.309
Moving time (10^5 min)	1.732	1.657	0.747	0.747	0.748
Cruising time (10^4 min)	1.128	0	0.510	0.510	0
Schedule (10^4 EUR$)	1.949	1.770	1.430	1.487	1.318
Early arrival (10^4 EUR$)	1.448	1.137	1.042	1.306	0.996
Late arrival (10^4 EUR$)	0.501	0.633	0.388	0.181	0.322
Ratio of N_e/N_l	3.7	2.4	3.1	5.2	3.1
Departure duration (min)	97.2	92.9	76.8	76.8	74.7

The toll (the red dashed lines) in Fig. 12(a) is obtained by letting $T_p(t) = 0$ (travel pricing only or underpricing parking) in Eq. (20), thus is piece-wise linear (just similar as the fine toll in the standard bottleneck model, which is due to the constant β and γ schedule cost setting). The peak start time is 126.45(min) thus the ratio of $N_e/N_l = l/e = 3.1$. Fig. 12(b) then displays the corresponding time-varying accumulation at equilibrium (for the duration with departures only). As can be seen, without consideration of parking in the toll, the accumulation during the peak can go beyond the critical values (approximately for 80% of the peak duration accumulation is 250 vehicles more than the critical value 1000). This leads to larger moving and cruising delays ($0.973 > 0.747$ and $0.618 > 0.510$), and larger social cost ($3.180 > 2.749$), as shown in Table 2.

In contrast, we also can only keep the parking pricing, i.e., $T_p(t)$ in Eq. (21), and ignore the pricing for travel, which corresponds to $e \cdot (t - t_{s,1})$ or $-l \cdot (t - t_{s,1})$ in Eq. (20). This toll is shown in Fig. 12(c), and the corresponding time-varying accumulation (for the duration with departures) is shown in Fig. 12(d). As implementing such a “parking-pricing-only” toll would lead to congestion, when calculating $T_p(t)$, we need to replace the speed $v(n_c)$ in Eq. (21) by $v(n(t))$. Similarly, the peak start time 88.74(min) is chosen thus $N_e/N_l = l/e = 3.1$. As can be seen, the “parking-pricing-only” toll is non-linear, and decreasing over time, mainly due to the nonlinearly increasing trip length over time. We have to point out that, this toll is very inefficient, thus social cost, toll revenue, moving time, cruising time (fourth column in Table 2) are all significantly larger than those under SO(a). This is expected as the toll pattern over time is quite different from that for SO(a). A very important implication from these results is that, partial pricing (parking pricing only) can lead to very inefficient situation, which indeed highlights the importance of the proposed joint pricing (of travel and parking) model in this paper, especially when cruising for parking cannot be neglected.

5.3. Comparison of different cases

Table 3 further summarizes different efficiency measures for five cases: i) User Equilibrium with cruising-for-parking; ii) User Equilibrium without cruising-for-parking (parking capacity approaches infinity); iii) SO(a) with cruising-for-parking; iv) SO(b) with cruising-for-parking; and v) System Optimum without cruising-for-parking. Note that cases (i) and (ii) are described in Section 5.1, and cases (iii) and (iv) are described in Section 5.2, and case (v) is the system optimum under Vickrey’s bottleneck model when parking capacity is infinity.

By comparing the UE with and without cruising (the first two columns in Table 3), we see that cruising-for-parking lead the total social cost, moving time, cruising time, and schedule delay to increase. However, the schedule delay cost of late arrival decreases, which is due to the fact that travelers are departing earlier to enjoy less cruising (there is a sharp increase in schedule delay cost of early arrival). By comparing the UE with SO(a) (both with cruising), we see huge reduction in travel cost (44.98%), moving time (56.87%), cruising time (54.79%), and schedule delay (26.63%) from implementing the pricing.

Besides, total cost including the toll at SO(a) is larger than social cost under UE with cruising, i.e., $2.749 + 2.558 > 4.996$, which means all individual travelers are worse off (as mentioned before, in SO(b) all travelers are better off, and in Table 3 we have $2.806 + 1.471 < 4.996$). However, for the cases without cruising, total cost including toll at SO ($2.553 + 1.309$) will be less than social cost at UE (4.507). This is consistent with Geroliminis and Levinson (2009).

Due to cruising-for-parking, there are more early arrival traffic, i.e., $N_e/N_l = 3.7$ for UE with cruising, which is larger that $N_e/N_l = 2.4$ for UE without cruising. The toll to support SO(a) then prevents early departure of travelers and N_e/N_l reduces to 3.1, which is equal to the ratio of late arrival penalty to early arrival penalty, i.e., l/e. This is consistent with our analytical results in Section 4. The peak starts later under SO(a) when compared to UE with cruising, and the peak duration shortens as well, i.e., the departure duration under SO(a) (76.8) is shorter than that under UE with cruising (97.2). Furthermore, the reduction of departure duration from UE with cruising to SO(a) or SO(b) (20.4) is larger than the reduction from UE without cruising to SO without cruising (18.2). This is because, besides the moving traffic, the cruising traffic also benefits from the reduced traffic congestion in the network.

6. Conclusion and discussion

In this paper, we construct a model to capture interactions between cruising and traffic congestion in the context of dynamic user equilibrium. Specifically, we formulate and analyze the morning commute equilibrium solution in a congested downtown network with a focus on cruising-for-parking. During the morning peak, as curbside parking vacancy decreases with time, the cruising distance and time for finding a vacant parking space is higher. Due to travelers' competition for smaller cruising distance (as well as trip length), the peak starts earlier; and due to reduced outflow when considering cruising-for-parking, the peak lasts longer.

A dynamic model of pricing for the network is then developed to reduce social cost, including cruising time cost, moving time (or in-transit time) cost, and schedule delay cost. It is shown that at the system optimum, the network should be operating at the critical accumulation with maximum production and highest traveling speed. However, the network outflow still decreases over time as parking vacancy decreases and trip length increases. The optimal time-dependent pricing gives the first commuter a higher toll than the last commuter, which prevents travelers from departing earlier due to competition for less cruising, and thus reduces total schedule delay of all travelers. Furthermore, it is proved that the total schedule delay is minimized when the ratio of early traffic to late traffic is equal to the ratio of late arrival penalty to early arrival penalty, even if the network outflow is not constant over time.

While this paper considers driving as the only travel mode, the analysis in the paper can be extended to the bi-modal transportation system with public transit as the alternative mode. In this case, besides departing earlier to enjoy less cruising time, travelers can take public transit to avoid cruising-for-parking. It is conjectured that similar dynamic toll can be introduced to reduce traffic inefficiency due to cruising-for-parking and roadway congestion, and achieve the bi-modal system optimum. However, if we consider that public transit service (e.g., fare, frequency) is responsive to the operating state of roadway network (similar to those in e.g., Zhang et al., 2014, Zhang et al., 2016), the problem will become more complicated, which is under our consideration for future research.

This study considers that the distribution of congestion over the network or region is homogeneous. If this is not the case, the MFD for the whole region might experience significant scatter or hysteresis. Recent studies (e.g., Geroliminis and Sun, 2011) have identified the spatial distribution of vehicle density as one of the important features that affect the scatter and the shape of the network MFD. However, the concept of an MFD and modeling framework in the paper might still be applied for the heterogeneously loaded downtown network if it can be partitioned into a small number of homogeneous regions. Recent work created clustering algorithms for heterogeneous transportation networks (e.g., Ji and Geroliminis, 2012).

Besides, the spatial distribution of parking over the network can be an important element to be taken into account as it affects drivers' cruising. In this case, one possible way could be classifying the parking spaces into groups where the parking is similarly distributed. Then we study how spatial distribution of parking can influence travelers' parking choices as well as traffic congestion. Also, we may consider garage parking in the extensions of the model.

The current study is from a long-term perspective, and relies on the recurrent behavior of travelers. However, in reality, travelers' travel choices and traffics are uncertain over time even if they are recurrent. Therefore, it is of our interest to develop a dynamic congestion/parking pricing system based on both the information of recurrent commuting behavior (e.g. distribution of travel demand over time which might be used for prediction of future traffics) and real dynamic traffic, which can maintain the downtown running at or at least near its optimum.

In practice, many travelers may have employee-based parking spaces, or contract-based parking spaces, or reservation-based parking spaces (for parking or highway capacity reservation, see, e.g., Liu et al., 2014a, 2015a) and they do not have to cruise for curbside spaces. In this case, the travelers can be classified into two categories: those with and without a guaranteed space. We then have to treat these two classes of travelers differently to study the commuting equilibrium with cruising for parking. Similar consideration has already been given in Yang et al. (2013), and we expect similar results as Yang et al. (2013).

Acknowledgment

The authors would like to thank Prof. Richard Arnott from University of California, Riverside and two anonymous referees for their helpful comments to improve this paper. This research was supported by ERC Starting Grant “METAFERW: Modeling and controlling traffic congestion and propagation in large-scale urban multimodal networks”.

Appendix A. Computing User Equilibrium[22,23]

Step 0: initialize t_s as $t_s = t^* - \hat{\tau}(t_s)$, and go to **Step 1**.
Step 1: (check convergence of t_s, and update t_s)
(i) take the initial value from **Step 0**, and go to **Step 2; or**
(ii) if $|(N - I(t_e))/N| > \varepsilon$, update t_s by $t_s = t_s - \Delta t_s$, where the Δt_s is based on the gap $N - I(t_e)$, i.e., $\Delta t_s = \frac{e}{e+l} \cdot \frac{N-I(t_e)}{I(t_e)} \cdot (t_e - t_s)$, and then go to **Step 2**; otherwise, the equilibrium solution of t_s is achieved.
Step 2: given t_s from **Step 1**, estimate time-dependent variables $n(t)$, $\hat{v}(t)$, $I(t)$, $p(t)$, $A(t)$, $\tilde{p}(t)$, and $o(t)$ for every $t \in [t_s, t_e]$, where t_e has to be determined during the estimation, based on $n(t_e) = n_c$.
Step 2–0: (check if t reaches t_e, and if not, update clock time t)
(i) update the current (clock) time by $t = t_s$ from **Step 1**, and go to **Step 2–1; or**
(ii) if $|(n(t) - n_c)/n_c| > \varepsilon$, update $t = t + \delta t$, and go to **Step 2–1**; otherwise, let $t_e = t$, go to **Step 1(ii)**.
Step 2–1: (given t from **Step 2–0**, solving a series of equations together, which are equilibrium condition in Eq. (14), traffic dynamics in Eq. (8), traffic conservation in Eq. (15), and parking dynamics in Eq. (2) and Eq. (7))
Step 2–1–0: (check convergence of $I'(t)$, and update $I'(t)$)
(i) let $k = 0$; initialize $I'(t)$ by $I'(t)^{(k)} = I_0$, and go to **Step 2–1–1; or**
(ii) if $|(I'(t)^{(k+1)} - I'(t)^{(k)})/I'(t)^{(k)}| > \varepsilon$, let $k = k + 1$, and then go to **Step 2–1–1**; otherwise, let $I'(t) = I'(t)^{(k+1)}$, and then go to **Step 2–0(ii)**.
Step 2–1–1: (update $p(t)$)
With $I'(t) = I'(t)^{(k)}$ from **Step 2–1–0**, calculate $I(t)$, then compute $p(t) = p(t)^{(k)}$ with Eq. (2) (**Parking dynamics**), and go to **Step 2–1–2**.
Step 2–1–2: (update $n(t)$)
With $p(t) = p(t)^{(k)}$ from **Step 2–1–1**, compute $n(t) = n(t)^{(k)}$ through Eq. (14) (**Equilibrium condition**) where $\hat{\tau}(t)$ is from Eq. (12), and go to **Step 2–1–3**.
Step 2–1–3: (update $\hat{o}(t)$ and $A(t)$)
Step 2–1–3–0: (check convergence of $\hat{o}(t)$ and update $\hat{o}(t)$)
(i) let $i = 0$, and initialize $\hat{o}(t)$ by $\hat{o}(t) = \hat{o}(t)^{(k)(i)} = o_0$, and go to **Step 2–1–3–1; or**
(ii) if $|(\hat{o}(t)^{(k)(i+1)} - \hat{o}(t)^{(k)(i)})/\hat{o}(t)^{(k)(i)}| > \varepsilon$, let $i = i + 1$, then go to **Step 2–1–3–1**; otherwise, let $\hat{o}(t) = \hat{o}(t)^{(k)(i+1)}$, and then go to **Step 2–1–4**;
Step 2–1–3–1: with $\hat{o}(t)$ from **Step 2–1–3–0**, update $A'(t)$ and $A(t)$, and then compute $\tilde{p}(t)$ with Eq. (7) (**Parking dynamics**), and go to **Step 2–1–3–2**;
Step 2–1–3–2: with $\tilde{p}(t)$ from **Step 2–1–3–1**, and $n(t) = n(t)^{(k)}$ from **Step 2–1–2**, compute the outflow $\hat{o}(t) = \hat{o}(t)^{(k)(i+1)}$ through Eq. (8). (**Traffic dynamics**), and go to **Step 2–1–4**;
Step 2–1–4: with $\hat{o}(t) = \hat{o}(t)^{(k)(i+1)}$ from **Step 2–1–3–2**, compute $I'(t)^{(k+1)}$ with Eq. (15) (**Traffic conservation**), and then go to **Step 2–1–0**.

Appendix B. System Optimum for a Single-region network

Given our “instantaneous model” and assumptions (A1-A10), we now show that the system optimum for a single-region network (in terms of minimizing social cost consists of only queueing delay cost and schedule delay cost) must occur at when the transport network is operating at the maximum production of its MFD, i.e., $n(t) = n_c$ and $\hat{v}(t) = v(n_c)$, and production $P = n_c \cdot v(n_c)$, during the departure and arrival intervals of travelers. Besides maintaining $n(t) = n_c$, as shown in Section 4, the peak start $t_{s,1}$ should be appropriately chosen such that the schedule delay can be minimized. While this is a standard finding in traffic engineering and control community, where drivers experience only congestion delay (and not schedule delay) cost, note that under different assumptions than A1-A10, the situation might be different.

First, we show that travel time cost will minimized when $\hat{v}(t) = v(n_c)$. For ease of presentation, we arrange the travelers in the order of their departure, then for the x-th traveler, the distance travelled will be $L(1 - \frac{x}{N_p} - p_0)$ (based on Eq. (2) and Eq. (5)). Suppose the x-th traveler has a traveling speed of v_x, the total travel time is then

$$TT = \int_0^N \frac{L\left(1 - \frac{x}{N_p} - p_0\right)}{v_x} dx. \tag{B1}$$

As $v_x \le v(n_c)$, it follows that

$$TT = \int_0^N \frac{L\left(1 - \frac{x}{N_p} - p_0\right)}{v_x} dx \ge \int_0^N \frac{L\left(1 - \frac{x}{N_p} - p_0\right)}{v(n_c)} dx. \tag{B2}$$

We now explain that to minimize total schedule delay cost, we must have $n(t) = n_c$ during the peak. We now focus on travelers' arrival at destination, as schedule delay depends on the arrival. We arrange the travelers in the order of their arrival, and at system optimum, the x-th traveler arrives at time t_x. The outflow at time t_x then is given by

$$\hat{o}_x = \hat{o}_x(t_x) = \frac{P(n(t_x))}{L\left(1 - \frac{x}{N_p} - p_0\right)}. \tag{B3}$$

22 The initial value $t_s = t^* - \hat{\tau}(t_s)$ is relatively large (the first traveler just arrives on time), and the corresponding $I(t_e)$ would be smaller than N in the beginning. As can be seen in Step 1 of the estimation procedure (or Loop 1 in Fig. 4), we gradually reduce t_s to achieve $I(t_e) \to N$. As $I(t_e) \to N$, the adjustment of t_s, i.e., Δt_s will approach zero, thus $t_s + \Delta t_s \to t_s$, which is numerically shown in Appendix C.

23 In Vickrey's model, as the early arrival travelers are l/e times of late arrival travelers, a more efficient step size Δt_s would be $\frac{l}{e+l} \cdot \frac{N-I(t_e)}{I(t_e)} \cdot (t_e - t_s)$. Here we have a more conservative step size, i.e., $\frac{e}{e+l} < \frac{l}{e+l}$, because we try to ensure that $I(t_e)$ will not go beyond N.

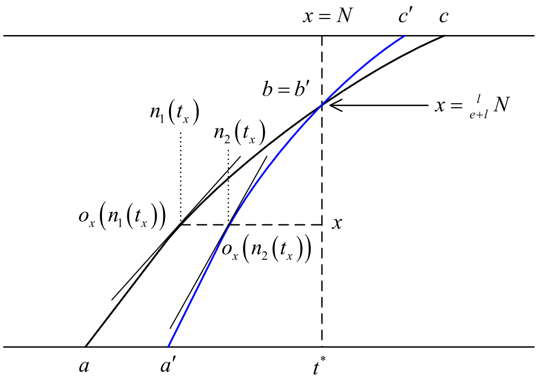


Fig. B.1. Cumulative arrival pattern over the peak.

Based on the analysis in Section 4, we know that at system optimum, the ratio of early traffic to late traffic should be $N_e/N_l = l/e$, i.e., $N_e = \frac{l}{e+l}N$. This means that the $\frac{l}{e+l}N$–th will arrive at time $t*$. This is shown in Fig. B.1 that the cumulative arrivals (no matter how it exactly looks like) would intersect with the vertical line representing $t*$ at $x = \frac{l}{e+l}N$. Then, we have

$$|t^* - t_x| = \begin{cases} \int_x^{N_e} \frac{1}{o_y} dy & x \le N_e \\ \int_{N_e}^{x} \frac{1}{o_y} dy & x > N_e \end{cases}. \tag{B4}$$

The total schedule delay cost would be

$$TS = \int_0^{N_e} e \cdot \int_x^{N_e} \frac{1}{o_y} dydx + \int_{N_e}^{N} l \cdot \int_{N_e}^{x} \frac{1}{o_y} dydx. \tag{B5}$$

As $P(n(t_x)) \le n_c \cdot v(n_c)$, with Eq. (B3) and let

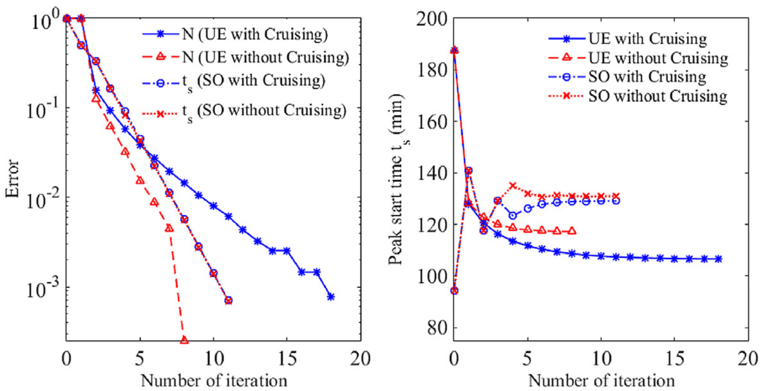
$$o_x(n_c) = o\left(n_c, 1 - \frac{x}{N_p} - p_0\right), \tag{B6}$$

where $o(n, \tilde{p})$ is defined in Eq. (8), we have $\hat{o}_x \le o_x(n_c)$. Then,

$$TS \ge \int_0^{N_e} e \cdot \int_x^{N_e} \frac{1}{o_y(n_c)} dydx + \int_{N_e}^{N} l \cdot \int_{N_e}^{x} \frac{1}{o_y(n_c)} dydx. \tag{B7}$$

Appendix C. Convergence of estimation procedures

Fig. C.1 (a) depicts the errors defined in the estimation procedures for User Equilibrium, i.e., $|(I(t_e) - N)/N|$, and for System Optimum, i.e., $|(t^u_{s,1} - t^l_{s,1})/t^u_{s,1}|$, against the number of iterations, and Fig. C.1 (b) depicts how the peak start time t_s



(a) Errors against iteration (b) The peak start time t_s against iteration

Fig. C.1. Convergences of peak start time for both UE and SO.

evolves over iteration. In Fig. C.1, for the UE and SO with cruising, $N = 6000$ and $N_p = 6500$ are applied, while for the UE and SO without cruising, $N = 6000$ and $N_p = 6 \times 10^{10}$. (Fig. C.1 is illustrative, we have tested different levels of demand and parking capacities, and observe similar trends.)

As can be seen in Fig. C.1, it takes more iterations for the UE with cruising (number of iteration: 18) than the UE without cruising (number of iteration: 8) to achieve $|(I(t_e) - N)/N| \le 10^{-3}$. As shown in Fig. C.1 (b), for estimating UE with or without cruising, by utilizing the information of the gap $|I(t_e) - N|$, and the average departure rate, i.e., $I(t_e)/(t_e - t_s)$, in the first few iterations, t_s is sharply reduced and becomes much closer to the final converged solution. The estimation of $t_{s,\,1}$ for SO with and without cruising is bi-section based, thus the $t_{s,\,1}$ goes up and down over iterations, and gradually converge (11 iterations).

References

Arnott, R., 2013. A bathtub model of downtown traffic congestion. J. Urban Econ. 76, 110–121.
Anderson, S.P., de Palma, A., 2004. The economics of pricing parking. J. Urban Econ. 55 (1), 1–20.
Arnott, R., de Palma, A., Lindsey, R., 1991. A temporal and spatial equilibrium analysis of commuter parking. J. Public Econ. 45, 301–335.
Arnott, R., Inci, E., 2006. An integrated model of downtown parking and traffic congestion. J. Urban Econ. 60, 418–442.
Arnott, R., Inci, E., 2010. The stability of downtown parking and traffic congestion. J. Urban Econ. 68, 260–276.
Axhausen, K.W., Polak, J.W., Boltze, M., Puzicha, J., 1994. Effectiveness of the parking guidance system in Frankfurt/Main. Traffic Eng. Control 35 (5), 304–309.
Arnott, R., Rowse, J., 1999. Modeling parking. J. Urban Econ. 45, 97–124.
Arnott, R., Rowse, J., 2009. Downtown parking in auto city. Reg. Sci. Urban Econ. 39, 1–14.
Bifulco, G.N., 1993. A stochastic user equilibrium assignment model for the evaluation of parking policies. Eur. J. Oper. Res. 71, 269–287.
Daganzo, C.F., 2007. Urban gridlock: Macroscopic modeling and mitigation approaches. Transp. Res. Part B 41 (1), 49–62.
Daganzo, C.F., Geroliminis, N., 2008. An analytical approximation for the macroscopic fundamental diagram of urban traffic. Transp. Res. Part B 42 (9), 771–781.
Fosgerau, M., 2015. Congestion in the bathtub. Econ. Transp. 4 (4), 241–255.
Fosgerau, M., de Palma, A., 2013. The dynamics of urban traffic congestion and the price of parking. J. Public Econ. 105, 106–115.
Fosgerau, M., Small, K., 2013. Hypercongestion in downtown metropolis. J. Urban Econ. 76, 122–134.
Geroliminis, N., 2015. Cruising-for-parking in congested cities with an MFD representation. Econ. Transp. 4 (3), 156–165.
Geroliminis, N., Boyaci, B., 2012. The effect of variability of urban systems characteristics in the network capacity. Transp. Res. Part B 46 (10), 1607–1623.
Gonzales, E.J., Daganzo, C.F., 2012. Morning commute with competing modes and distributed demand: User equilibrium, system optimum, and pricing. Transp. Res. Part B 46 (10), 1519–1534.
Geroliminis, N., Daganzo, C.F., 2008. Existence of urban-scale macroscopic fundamental diagrams: some experimental findings. Transp. Res. Part B 42 (9), 759–770.
Geroliminis, N., Levinson, D., 2009. Cordon pricing consistent with the physics of overcrowding. In: Lam, W., Wong, S., Lo, H. (Eds.), Transportation and Traffic Theory 2009, pp. 219–240.
Glazer, A., Niskanen, E., 1992. Parking fees and congestion. Reg. Sci. Urban Econ. 22 (1), 123–132.
Geroliminis, N., Sun, J., 2011. Properties of a well-defined macroscopic fundamental diagram for urban traffic. Transp. Res. Part B 45 (3), 605–617.
Haddad, J., Ramezani, M., Geroliminis, N., 2013. Cooperative traffic control of a mixed network with two urban regions and a freeway. Transp. Res. Part B 54, 17–36.
Inci, E., 2015. A review of the economics of parking. Econ. Transp. 4 (1-2), 50–63.
Ji, Y., Geroliminis, N., 2012. On the spatial partitioning of urban transportation networks. Transp. Res. Part B 46 (10), 1639–1656.
Leclercq, L., Chiabaut, N., Trinquier, B., 2014. Macroscopic Fundamental Diagrams: A cross-comparison of estimation methods. Transp. Res. Part B 62, 1–12.
Leclercq, L., Geroliminis, N., 2013. Estimating MFDs in simple networks with route choice. Transp. Res. Part B 57, 468–484.
Liu, W., Yang, H., Yin, Y., 2014a. Expirable parking reservations for managing morning commute with parking space constraints. Transp. Res. Part C 44, 185–201.
Liu, W., Yang, H., Yin, Y., 2015a. Efficiency of a highway use reservation system for morning commute. Transp. Res. Part C 56, 293–308.
Liu, W., Yang, H., Yin, Y., Zhang, F., 2014b. A novel permit scheme for managing parking competition and bottleneck congestion. Transp. Res. Part C 44, 265–281.
Liu, W., Yin, Y., Yang, H., 2015b. Effectiveness of variable speed limits considering commuters' long-term response. Transp. Res. Part B 81, 498–519.
Mahmassani, H., Herman, R., 1984. Dynamic user equilibrium departure time and route choice on idealized traffic arterials. Transp. Sci. 18 (4), 362–384.
Qian, Z., Rajagopal, R., 2014. Optimal dynamic parking pricing for morning commute considering expected cruising time. Transp. Res. Part C 48, 468–490.
Qian, Z., Rajagopal, R., 2015. Optimal dynamic pricing for morning commute parking. Transp. A: 11 (4), 291–316.
Qian, Z., Xiao, F., Zhang, H.M., 2011. The economics of parking provision for the morning commute. Transp. Res. Part A 45, 861–879.
Qian, Z., Xiao, F., Zhang, H.M., 2012. Managing morning commute traffic with parking. Transp. Res. Part B 46, 894–916.
Ramezani, M., Haddad, J., Geroliminis, N., 2015. Dynamics of heterogeneity in urban networks: aggregated traffic modeling and hierarchical control. Transp. Res. Part B 74, 1–19.
Shoup, D.C., 2006. Cruising for parking. Transp. Policy 13 (6), 479–486.
Small, K.A., Chu, X., 2003. Hypercongestion. J. Transp. Econ. Policy 37, 319–352.
Tseng, Y.Y., Ubbels, B., Verhoef, E.T., 2005. Value of time, schedule delay and reliability-estimation based on choice behaviour of Dutch commuters facing congestion. Paper presented at the 45th ERSA Congress. VU University Amsterdam.
Vickrey, W.S., 1969. Congestion theory and transport investment. Am. Econ. Rev. 59, 251–261.
Yildirimoglu, M., Geroliminis, N., 2013. Experienced travel time prediction for congested freeways. Transp. Res. Part B 53, 45–63.
Yildirimoglu, M., Geroliminis, N., 2014. Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams. Transp. Res. Part B 70, 186–200.
Yang, H., Huang, H.J., 1997. Analysis of the time-varying pricing of a bottleneck with elastic demand using optimal control theory. Transp. Res. Part B 31 (6), 425–440.
Yang, H., Liu, W., Wang, X., Zhang, X.N., 2013. On the morning commute problem with bottleneck congestion and parking space constraints. Transp. Res. Part B 58, 106–118.
Zhang, X.N., Huang, H.J., Zhang, H.M., 2008. Integrated daily commuting patterns and optimal road tolls and parking fees in a linear city. Transp. Res. Part B 42 (1), 38–56.
Zhang, X.N., Yang, H., Huang, H.J., 2011. Improving travel efficiency by parking permits distribution and trading. Transp. Res. Part B 45 (7), 1018–1034.
Zhang, F., Lindsey, R., Yang, H., 2016. The Downs-Thomson paradox with imperfect mode substitutes and alternative transit administration regimes. Transp. Res. Part B 86, 104–127.
Zhang, F., Yang, H., Liu, W., 2014. The Downs-Thomson Paradox with responsive transit service. Transp. Res. Part A 70, 244–263.