Analysis and Design of Markets for Tradable Mobility Credit Schemes

Siyu Chen\textsuperscript{a,}\textsuperscript{*}, Ravi Seshadri\textsuperscript{b}, Carlos Lima Azevedo\textsuperscript{c}, Arun P. Akkinepally\textsuperscript{d}, Remming Liu\textsuperscript{c}, Andrea Araldo\textsuperscript{e}, Yu Jiang\textsuperscript{c}, Moshe E. Ben-Akiva\textsuperscript{a}

\textsuperscript{a}Civil and Environmental Engineering Department, Massachusetts Institute of Technology, Cambridge, MA, US
\textsuperscript{b}Singapore-MIT Alliance for Research and Technology, Singapore
\textsuperscript{c}Department of Technology, Management and Economics, Technical University of Denmark, Denmark
\textsuperscript{d}Caliper Corporation, United States
\textsuperscript{e}Télécom SudParis - Institut Polytechnique de Paris, France

Abstract

 Tradable mobility credit (TMC) schemes are an approach to travel demand management that have received significant attention in the transportation domain in recent years as a promising means to mitigate the adverse environmental, economic and social effects of urban traffic congestion. In TMC schemes, a regulator provides an initial endowment of mobility credits (or tokens) to all potential travelers. In order to use the transportation system, travelers need to spend a certain amount of tokens (tariff) that could vary with their choice of mode, route, departure time etc. The tokens can be bought and sold in a market that is managed by and operated by a regulator at a price that is dynamically determined by the demand and supply of tokens.

This paper proposes and analyzes alternative market models for a TMC system (focusing on market design aspects such as allocation/expiration of credits, rules governing trading, transaction costs, regulator intervention, price dynamics), and develops a methodology to explicitly model the dis-aggregate behavior of individuals within the market. Extensive simulation experiments are conducted within a departure time context for the morning commute problem to

\textsuperscript{*}Corresponding author
Email address: siyuc@mit.edu (Siyu Chen)
compare the performance of the alternative designs relative to congestion pricing and a no control scenario. The simulation experiments employ a day to day assignment framework wherein transportation demand is modeled using a logit-mixture model and supply is modeled using a standard bottleneck model. The results indicate that when the actual network equilibrium does not deviate from the forecasted equilibrium (used by the regulator to design the toll structure, which varies by time-of-day but is not day-to-day adaptive), the optimized TMC system can achieve an identical social welfare as congestion pricing in the absence of transaction costs (and only a marginally lower welfare in the presence of transaction costs). On the other hand, when the forecasted and actual equilibria are different, the TMC system can yield efficiency gains over congestion pricing due to the price adjustment mechanism of the market (with and without transaction costs). The TMC system yields similar efficiency gains also when step tolls are used (common in practice) and further, is more robust to supply/demand shocks when the allocation of tokens occurs in continuous time. Finally, the results highlight the importance of transaction costs and decaying prices until expiration in mitigating undesirable behavior in the market.

The paper addresses a growing and imminent need to develop methodologies to realistically model TMCs that are suited for real-world deployments and can help us better understand the performance of these systems – and the impact in particular, of market dynamics.

*Keywords:* Tradable Mobility Credits; Demand Management; Human Behavior; Traffic Management; Simulation

1. Introduction

Historically, transportation network inefficiencies such as congestion and vehicular emissions have been addressed through road pricing, which although used in several cities worldwide, is plagued by issues of inequity and political and public acceptability (Tsckeris and Voß, 2009; de Palma and Lindsey, 2011). An alternative approach to travel demand management that has received increas-
ing attention in the transportation domain in recent years is quantity control – in particular, tradable mobility credit (TMC) schemes (Fan and Jiang, 2013; Grant-Muller and Xu, 2014; Dogterom et al., 2017). Within a TMC system, a regulator provides an initial endowment of mobility credits to all potential travelers. In order to use a transportation system, users need to spend a certain number of permits (i.e.: tariff) that could vary with the conditions/performance of the specific mobility alternative used. The permits can be bought and sold in a market that is monitored by the regulator at a price that is determined by demand and supply interactions.

In principle, TMC schemes are appealing since they offer a means of directly controlling quantity (important when the elasticity of demand to prices in the short term may be low), they are revenue neutral in that there is no transfer of money to the regulator, and they are viewed as being less vertically inequitable than congestion pricing (de Palma and Lindsey, 2020). Despite these promises, several important questions remain with regard to the design and functioning of the market within TMC schemes, an aspect critical to the effective operationalization of these schemes. For instance, how should the allocation and expiration of tokens be designed? What rules should govern trading behavior in the market so as to avoid undesirable speculation and trading (see Brands et al. (2020) for more on this), and yet ensure efficiency and revenue neutrality? How should the regulator intervene in the market in the presence of special or non-recurrent events? What is the role and impact of transaction costs? Despite the large body of literature on TMCs, issues of market design, market dynamics and behavior of individuals in the market has received relatively less attention.

This paper aims to address these issues and contributes to the existing literature in several respects. First, we propose alternative market models (focusing on all aspects of market design including allocation/expiration of credits, rules governing trading, transaction costs, regulator intervention, price dynamics) for a TMC system, and develop a methodology that explicitly models the dis-aggregate behavior of individuals within the market. Second, we conduct extensive simulation experiments within a departure time context for the morning
commute problem to compare the performance of the alternative designs relative to congestion pricing and a no control scenario. The simulation experiments employ a day to day assignment framework wherein transportation demand is modeled using a logit-mixture model and supply is modeled using a standard bottleneck model. The experiments yield insights into market design and the comparative performance of the TMC system relative to congestion pricing.

The results indicate that when the actual network equilibrium does not deviate from the forecasted equilibrium (used by the regulator to design the toll structure, which varies by time-of-day but is not day-to-day adaptive), the optimized TMC system can achieve an identical social welfare as congestion pricing in the absence of transaction costs (and a marginally lower welfare in the presence of transaction costs). On the other hand, when the forecasted and actual equilibria are different, the TMC system can yield efficiency gains over congestion pricing due to the price adjustment mechanism of the market (with and without transaction costs). The TMC system yields similar efficiency gains also when step tolls are used (common in practice) and further, is more robust to supply/demand shocks when the allocation of tokens occurs in continuous time. Finally, the results highlight the importance of transaction costs and decaying prices until token expiration in mitigating undesirable behavior in the market.

The TMC system that we propose (termed Trinity hereafter) comprises three main components, an online bi-level optimization model, a market model and a smartphone app (see Figure 1). From the user’s perspective, Trinity is a smartphone app which includes (1) a personalized trip planner and (2) interfaces for the user’s token account and (3) trading. Users are endowed with a token budget, which is obtained through a subscription or free allocation by the regulator. Prior to a trip, travelers open the trip planner, which presents a menu with different travel alternatives along with their predicted attributes. Each alternative is also associated with a token tariff, which can be charged depending on the alternative-specific contribution to the system’s congestion. The Trinity app tracks and verifies realized trips for token charging. Moreover, Trinity learns individual user’s preferences from previous choices and presents
personalized menus, which increases the user’s benefit (Song et al., 2018).

Figure 1: Trinity Architecture Flowchart

The second component of Trinity is the bi-level optimization module, which is responsible for setting the token charges or tariff for each travel alternative in real time (‘system-level’ optimization) and providing personalized ‘user-optimal’ menus to travelers (‘user-level’ optimization). The system-level optimization utilizes a simulation-based predictive system that uses real-time data from the market and from sensors in the transportation system (Araldo et al., 2019). The overall policy objectives for Trinity in terms of congestion, emissions, network performance, quality of service and sustainability is defined via the system-level optimization.

The third component of Trinity is the Token Market in which users can sell or buy tokens. If a user chooses a travel alternative associated with a certain token amount and her/his token budget is insufficient, she/he can buy the remaining needed tokens. On the other hand, a user can sell excess tokens in her/his budget at any time. The token market price at which these exchanges occur adjusts dynamically based on demand and supply of tokens. If demand
exceeds supply, the price increases and vice-versa. The market enables Trinity to achieve revenue neutrality, which means the system avoids taxes, user charges or incentive funding programs. The operator can also intervene in the market, reducing or increasing the number of tokens available and thus allowing for a better management of non-recurrent situations.

As noted previously, our key focus in this paper is on the analysis and design of the market within Trinity. Future research will focus on other aspects of the Trinity system including the online bi-level optimization framework. The rest of the paper is organized as follows.

2. Review of Literature

Although early work on the use of tradable mobility credits (TMCs; also termed TCS or Tradable Credit Schemes in the literature) in transportation date back several years (Verhoef et al., 1997; Raux, 2007; Goddard, 1997), formulations of the market and network equilibrium for TMCs is more recent, pioneered by the work of Yang and Wang (2011) who proposed a user equilibrium variant for a TMC. Their work, along with advancements in technology and the widely recognized limitations of congestion pricing, has spurred interest in TMCs for transportation network management. Extensive reviews may be found in Grant-Muller and Xu (2014); Fan and Jiang (2013); Dogterom et al. (2017). We provide a brief summary of existing literature, limiting our attention to that of mobility management (in the context of both entire networks and single bottlenecks) although applications may also be found in parking.

In the model of Yang and Wang (2011), the regulator distributes a prespecified number of credits to travelers, charges a link-specific credit tariff and allows trading of credits within a market. They demonstrate that for a given set of credit rates in a general network, the user equilibrium (UE) link flow pattern is unique under standard assumptions and identify additional conditions (relatively mild) to ensure uniqueness of the credit price at the market equilibrium. Extensions to their model have been proposed to incorporate heterogeneity in
the value of time (Wang et al., 2012) and multiple user classes (Zhu et al., 2015) using variational inequality formulations to establish existence and uniqueness properties of the network and market equilibrium. He et al. (2013) employ a similar equilibrium approach considering allocations of credits to not just individual travelers, but to transportation firms such as logistics companies and transit agencies; the effect of transaction costs in a TMC scheme with two types of markets (auction-based and negotiated) is considered by Nie (2012). In contrast with the aforementioned TMC schemes, Kockelman and Kalmanje (2005); Gulipalli and Kockelman (2008) proposed a system of credit-based congestion pricing (termed CBCP) where credits are allowances used to pay tolls.

While the studies discussed thus far have examined the application of TMCs typically in a route choice setting at the network level, several researchers have proposed TMC schemes in the context of managing congestion at a single bottleneck (or simple two route networks) by achieving peak spreading. Nie and Yin (2013) model a tradable credit scheme that manages commuters’ travel choices and attempts to persuade commuters to spread their departure times evenly within the rush hour and between alternative routes (see also Nie (2015)) whereas Tian et al. (2013) investigate the efficiency of a tradable travel credit scheme for managing bottleneck congestion and modal split in a competitive highway/transit network with a continuously distributed value of time. Along related lines, Xiao et al. (2013) study a tradable credit system (consisting of a time-varying credit charged at the bottleneck wherein the credits can be traded and the price is determined by a competitive market) to manage morning commute congestion with both homogenous and heterogeneous users. More recently, Bao et al. (2019) studied the existence of equilibria under tradable credit schemes using different models of dynamic congestion and Akamatsu and Wada (2017) proposed a tradable bottleneck credit scheme where the regulator issues link- and time-specific credits permitting passage through a certain link or bottleneck in a pre-specified time period. They develop a model to describe time-dependent flow patterns at equilibrium under a system of tradable bottleneck permits for general networks and show that the equilibrium obtained
under this system is efficient in that it minimizes the social transportation cost.

In contrast with the previously described literature that largely focus on variants of the standard user equilibrium under TMC schemes, a related stream of research examines the design of the TMC schemes/network design using bi-level optimization formulations in different contexts (Wu et al., 2012; Bao et al., 2017; Wang et al., 2014). On the other hand, the comparison of efficiency properties of tradable credits and congestion pricing has received relatively lesser attention. de Palma et al. (2018) performed a comparative analysis of the two instruments in a simple transportation network and showed that as long as there is no uncertainty, price and quantity regulation are equivalent as in the regular market case studied by Weitzman (1974). In the presence of uncertainty and strongly convex congestion costs, the TMC instrument outperforms the pricing instrument in efficiency terms. Akamatsu and Wada (2017) reached similar conclusions (see also Shirmohammadi et al. (2013)), demonstrating the equivalence of the tradable permit system and a congestion pricing system when the road manager has perfect information of transportation demands. On the behavior side, several stated preference studies have highlighted the importance of key behavioral economics and cognitive psychology factors towards TMC (Dogterom et al. 2017) (see also Brands et al. (2020) for an interesting real-world experiment with tradable credits).

In summary, despite the large body of research on TMCs, the modeling of the market has received little attention and almost all the studies employ an equilibrium approach to model the credit market (with the notable exception of Ye and Yang (2013) who model the price and flow dynamics of a tradable credit scheme). Further, the literature has –to the best of our knowledge– thus far not attempted to model realistically the disaggregate behavior of individuals within the market that could enable the consideration of empirically observed phenomena such as loss aversion, endowment effects, mental accounting, day-to-day learning (Dogterom et al. 2017). Finally, there is also the imminent need to examine design aspects of the credit market including features such as token allocation/expiration, trading, intervention, and transaction costs and the
impact of these on behavior of individuals in the market. This paper aims to address some of these gaps.

3. Market design

Traditional road pricing which charges money for traveling (along certain routes or at certain departure times) can be thought of in a TMC perspective, as a TMC system in which only buying is allowed. Specifically, instead of charging a time-of-day toll (in units of dollars), the regulator charges the same toll but in electronic tokens and maintains a fixed token market price of $1. Further, the regulator does not distribute any token endowment to travelers and is able to satisfy all buying demand. In order to avoid quantity buildup and market manipulation, travelers can only buy tokens for immediate travel use and any extra tokens after travel will expire immediately. Thus, the toll cost (in dollars) under road pricing is the same as the toll cost under TMC scheme, which is equal to the product of the toll in electronic tokens and the $1 token price. Such a TMC scheme is clearly as efficient as road pricing in internalizing congestion externalities. However, similar to road pricing, this TMC scheme will be perceived as unfair or just another flat tax. Under the TMC scheme, revenue neutrality can potentially be achieved by providing a token endowment and allowing 1) token selling and 2) token price adjustment through a market.

To overcome shortcomings of road pricing and maintain its efficiency, in this study, the Trinity regulator 1) maintains the same time-of-day toll in tokens across days, 2) distributes a token endowment to all travelers (more on the nature of this endowment in the following paragraph), 3) satisfies all buying demand, and 4) is able to intervene in the token market within day by controlling token market price, token allocation, and transaction costs to accommodate non-recurrent events. Travelers can trade tokens in a market place at which token price is adjusted day to day considering token demand and supply to ensure revenue neutrality.

Trinity adopts a ‘continuous time’ approach wherein tokens are acquired
(provided by the regulator) at a certain rate over the entire day and each token has a lifetime (i.e. it expires after a certain period also specified by the regulator). The expiration of tokens will avoid undesirable consequences of the TMC system that can compromise public acceptability such as speculative behavior and hedging in the market. Although price is adjusted daily in this study, ‘continuous’ allocation provides more degrees of freedom for regulator to intervene than that of ‘lump sum’ allocation. Numeric experiments of comparing two allocation approaches will be presented in Section 5.

As a result, each individual acquires tokens at a constant rate $r$ over entire day and each token has a lifetime $L$ to avoid speculation. Let $x_{dn}^d(t)$ denote traveler $n$’s token account balance at time $t$ on day $d$ and a Bernoulli indicator $I_{dn}^d(t)$ denote whether the traveler $n$ is at a full wallet (FW) state or not. A full wallet state indicates that the number of tokens in the wallet has reached a maximum and, in the absence of travelling or selling, does not change since the acquisition of new tokens is balanced by an expiry of old tokens. Thus, a full wallet implies that the oldest token in her account has an age $L$. In contrast, when the account is not in a full wallet state, it increases by an amount $r \Delta_t$ in a time interval $\Delta_t$.

In this study, in order to avoid quantity buildup and market manipulation, 1) travelers can only buy tokens from regulator only at time of traveling for immediate use if they are short of tokens, 2) they can sell tokens to regulator but they have to sell all, and 3) buying and selling cannot happen at the same time, i.e. travelers can sell all tokens anytime except at time of traveling. Let $T(t)$ denote toll in electric tokens of travel alternative at time $t$ and $td_{dn}^d$ represents departure time of traveler $n$ on day $d$. At time $t$ on day $d$, traveler $n$ can perform one and only one of the following actions:

1. Perform a trip if $t = td_{dn}^d$.
   - If a traveler $n$ has enough tokens to perform the trip at time $td_{dn}^d$, i.e., $x_{dn}^d(td_{dn}^d) \leq T(td_{dn}^d)$, she consumes $T(td_{dn}^d)$ from her account. Her account balance at time $td_{dn}^d + \Delta_t$ is equal to $x_{dn}^d(td_{dn}^d + \Delta_t)$ and can
be written as:

\[ x^d_n(td^d_n + \Delta t) = x^d_n(td^d_n) - T(td^d_n) + r\Delta t \]  

(1)

- Otherwise, she needs to buy \( T(td^d_n) - x^d_n(td^d_n) \) tokens at the price \( p^d(td^d_n) \) plus additional transaction cost from the market. Her account balance \( x^d_n(td^d_n + \Delta t) \) becomes:

\[ x^d_n(td^d_n + \Delta t) = r\Delta t \]  

(2)

2. Does nothing. Her account balance \( x^d_n(t + \Delta t) \) becomes:

\[ x^d_n(t + \Delta t) = x^d_n(t) + I^d_n(t)r\Delta t \]  

(3)

3. Sells all tokens \( x^d_n(t) \). Traveler \( n \) chooses to sell all her tokens and her account balance becomes:

\[ x^d_n(t + \Delta t) = r\Delta t \]  

(4)

The buying price of tokens at time \( t \) on day \( d \) is equal to token market price \( p^d(t) \). However, the selling price of tokens decays linearly as tokens expire in order to account time values of tokens and avoid quantity buildup. In addition, there are two-part (fixed and proportional) transaction costs for both buying and selling transaction. Let \( c_s \) and \( c_b \) denote proportional part of selling and buying transaction costs; \( TC_s \) and \( TC_b \) denote fixed part of selling and buying transaction costs. Therefore, selling revenue of \( y \) tokens at time \( t \) can be written as

\[ S^d(y,t) = \min(y,FW)p_s^d(t) - \frac{1}{2} p_s^d y^2 + \frac{1}{2} p_s^d y^2 - TC_s \]  

(5)

where \( p_s^d(t) = p^d(t)(1-c_s) \) and the quadratic term is to account for decaying time values of tokens.

Buying cost of \( y \) tokens at time \( t \) on day \( d \) can be written as

\[ B^d(y,t) = yp_b^d(t) + TC_b \]  

(6)
where \( p_d^d(t) = p_d(t)(1 + c_b) \).

At time \( t \) on day \( d \), assume traveler \( n \) has plan about upcoming travel alternative, she is able to calculate profit of selling now assuming no further selling until departure as follows

\[
\text{profit}^d_n(t) = \min(S^d(x_n(t), t), S^d(FW, t)) - \mathbb{I}(T(td_n^d) > FA)B^d(T(td_n^d) - FA, t)
\]

(7)

where \( FA \) represents future allocation and is equal to \( \min((td_n^d - t)r, FW) \); \( T(td_n^d) \) represents toll in tokens of travel alternative at departure time \( td_n^d \). Buying cost only occurs if toll is greater than traveler \( n \)'s future allocation \( T(td_n^d) \geq \min((td_n^d - t)r, FW) \).

At time \( t \) on day \( d \), traveler \( n \) considers selling only if profit value is positive \( (\text{profit}^d_n(t) > 0) \). She further decides selling now or wait based on the derivative of profit function with respect to time as follows:

1. When \( I(T(td_n^d) > FA) = 1 \) (need to buy), it can be written as

\[
\frac{d\text{profit}^d_n(t)}{dt} = \begin{cases} 
-rp_b^d(t) & x_n^d(t) = FW \\
rp_b^d(t) - x_n^d(t)\frac{p_b^d(t)}{24} - rp_b^d(t) & \text{otherwise}
\end{cases}
\]

which is always negative if \( p_d^d(t) \leq p_b^d(t) \). This implies that future profit is guaranteed to be less than current profit. She should sell now if profit is positive.

2. When \( I(T(td_n^d) > FA) = 0 \) (no need to buy), it can be written as

\[
\frac{d\text{profit}^d_n(t)}{dt} = \begin{cases} 
0 & x_n^d(t) = FW \\
rp_b^d(t) - x_n^d(t)\frac{p_b^d(t)}{24} & \text{otherwise}
\end{cases}
\]

which implies that current profit increases until reaching full wallet. Note that profit function is not continuous so it will stop increasing once toll becomes equal to future allocation.
In summary, selling strategy can be summarized as follows.

**Algorithm 1: Selling Rule**

**Input:** $d, t, n, p^d(t), t_n^d, x_n^d(t)$

1. Calculate $\text{profit}_n(t)$:
   - If $\text{profit}(t) > 0$ then
     - If $T(t_d) > \min((t_d - t)r, FW)$ then
       - Sell now;
     - Else if $T(t_d) < \min((t_d - t)r, FW)$ then
       - If $x(t) = FW$ then
         - Sell now;
       - Else
         - Do nothing;
     - Else
       - Sell now;
   - Else
     - Do nothing;
2. The market place dictates the token price $p^d$ on day $d$ and the policy to establish it is defined a priori by the regulator. The price $p^d$ is modified daily with a deterministic rule considering previous day’s regulator revenue $R^{d-1}$ as follows:

   $p^d = \begin{cases} 
   p^{d-1} & R^{d-1} \in [-RBTD, RBTD] \\
   p^{d-1} + \Delta p & R^{d-1} < -RBTD \\
   p^{d-1} - \Delta p & R^{d-1} > RBTD 
   \end{cases} \quad (8)$

   where $\Delta p$ currently is a constant parameter representing price change amount. $RBTD$ ensures that price will not fluctuate for small regulator revenue. Price is ensured to be positive and below a certain cap $p^{max}$ as follows:
\[ p^d = \max(0, \min(p^d, p^{\text{max}})) \] (9)

Although token price is constant within day for the most of time, regulator is able to control token price for a period time in one day to accommodate unusual events. For example, if road capacity drops because of a large concert or extreme weather, regulator is able to increase token price in peak hour to discourage travel in peak period. Numerical experiments are conducted to study this in section 5.

4. Experiments

4.1. Overview

In this study, we conduct simulation-based experiments to assess three scenarios: No Toll (NT), Congestion Pricing (CP), and Trinity. As we can see from Figure 2, for a given scenario, travelers use forecasted information, including travel time, departure time, and account balance, to make mobility decision in the beginning of a day (Preday). For the sake of simplicity, travelers are homogeneous and each one performs one morning trip per day without cancellation allowed in this study. The network considered is a single OD pair connected by a single link. Thus the mobility decision is departure time choice only, which is modeled by a discrete logit model. Determined departure time choice will be simulated in the network within day along with rule-based trading decisions described in previous section. Congestion is modeled by a point queue model, in which a queue develops once flow exceeds capacity. The demand and supply interaction is modeled by a dynamic process, which explicitly simulates the evolution of the system state considering day to day learning. Exponential smoothing filter is adopted to update forecasted information with realized costs if system state (e.g. travel time and flow) is not stationary across days. We compute social welfare at stationary state and compare it across scenarios.

In next sections, we first describe formulations for a system model (travelers and network components in Fig.1) of commuters’ departure time choice in a
Figure 2: Flowchart of stochastic simulations. Travelers’ departure time choices determined by discrete logit model using forecasted information in the beginning of the day are simulated on a single link with a finite capacity along with trading activities. If system is not stationary (flow and travel time differ across days), forecasted information will be updated using realized information. Social welfare will be computed at the end and compared across scenarios.

single corridor and welfare computation for NT, CP and Trinity. Then, we present a variant Trinity in which regulator adopts a strict token quantity control to control flow pattern. Finally, we propose different experiments to analyze efficiency and robustness of Trinity. Important notation is shown in Table 1.
| Variables | Description |
|-----------|-------------|
| \( h \)  | Time interval |
| \( t \)  | Simulation time step |
| \( d \)  | Day \( d \) |
| \( t_h \) | Start time of interval \( h \) |
| \( \Delta_h \) | Duration of time interval |
| \( \Delta_t \) | Time step |
| \( \alpha \) | Value of time |
| \( \beta_E, \beta_L \) | Value of schedule delay early/late |
| \( \beta_D \) | Value of delay due to postpone |
| \( \mu \) | Scale parameter of random component |
| \( e^d(h) \) | Random utility component on day \( d \) |
| \( p \) | Fixed market price as $1 |
| \( T(h) \) | Toll in tokens in \( h \) |
| \( \tilde{\tau}^d(h) \) | Forecasted travel time in \( h \) on day \( d \) |
| \( \tilde{\omega}^d(h) \) | Forecasted delay time in \( h \) on day \( d \) |
| \( \tilde{\omega}^d(h) \) | Forecasted delay time in number of intervals on day \( d \) |
| \( x_n(t) \) | Account balance of individual \( n \) at time \( t \) |
| \( L \) | Token lifetime |
| \( r \) | Token allocation rate |
| \( t_0 \) | Free flow travel time |
| \( t_w(t) \) | Waiting time in queue at \( t \) |
| \( D(t) \) | Number of drivers in queue at \( t \) |
| \( \tau^d(t) \) | Travel time at \( t \) on day \( d \) |
| \( \tau^d(h) \) | Averaged travel time of interval \( h \) on day \( d \) |
| \( \theta_\tau/\theta_{\omega}/\theta_\psi \) | Weights on previous day’s forecasts |
4.2. System model

The setting we consider involves a single origin-destination pair connected by a path containing a bottleneck of finite capacity. Unlike the classical bottleneck model proposed by Vickrey (1969), users wish to arrive at the destination within a certain “preferred arrival time window” in the morning, and can adjust their departure times to avoid congestion (similar to the model in Ben-Akiva et al. (1984), which is a dynamic extension of De Palma and Lefevre (1983)). In addition, the system is modeled using a stochastic process approach that incorporates day-to-day and can be viewed as a simplification of the model in Cascetta and Cantarella (1991), which considers the stochastic assignment problem in general networks. Day to day adjustment is modeled using suitable learning and forecasting filters and a logit model is used to model within-day departure time decisions. We refer to Cantarella and Cascetta (1995) for a nuanced discussion of terminology and a detailed description of deterministic and stochastic process models (with probabilistic assignment or a probabilistic model for users’ choice behavior). They propose conditions for existence and uniqueness of fixed-point attractors of the deterministic process, which extend results for the traditional user and stochastic user equilibrium. In case of the stochastic process, conditions for regularity are proposed which ensure existence and uniqueness of a stationary distribution of network states. It is noted that the model of Ben-Akiva et al. (1984) may be viewed as a deterministic process model with probabilistic assignment.

The travel behavior model, network model, and demand-supply interactions are discussed in detail next.

4.2.1. Travel behavior model

In this study, traveler’s only travel behavior —departure time choice— at a unit time interval $h$ being chosen within a set of mutually exclusive possible departure time intervals $\{1, \ldots, H\}$ is modeled by a discrete logit model (similar to continuous logit departure time choice model considered in Ben-Akiva and Watanatada (1981)). According to random utility theory, let $V^d(h)$ be a sys-
tematic monetary utility (utility in unit of dollars) for departing at time interval $h$ on day $d$ and $\epsilon^d(h)$ be independently and identically distributed as Gumbel random variables with zero mean and scale parameter $\frac{1}{\mu}$. The probability of departing in a time interval $h$ on day $d$ is written as

$$P^d(h) = \frac{\exp(V^d(h)\mu)}{\sum_{h=1}^{H} \exp(V(h)\mu)}$$

(10)

in which a time interval $h$ contains several simulation time steps $t$ based on simulation setting and travelers will randomly select a time point to depart within a chosen time interval $h$.

Systematic monetary utility $V^d(h)$ consists of four cost components that affect the choice of departure time: forecasted travel time, schedule delay early, schedule delay late and forecasted toll ($0$ for NT). For a traveler, the marginal utility of an additional unit of travel time is $\alpha$. For simplicity, we assume travelers have common knowledge of forecasted travel time. Denote her desired time period for arrival as $[t^*-\Delta,t^*+\Delta]$, where $t^*$ represents the center of the period and $\Delta$ represents arrival flexibility. If she arrives outside of desired time period, she suffers a schedule delay. The marginal utility of an additional unit of schedule delay early is $\beta_E$ and an additional unit of schedule delay late is $\beta_L$, which $\beta_E \leq \alpha \leq \beta_L$ according to empirical results (Small (1982)). Under CP scheme, forecasted toll cost incurred at time interval $h$ is equal to toll in dollars charged at $h$, $p^{CP}(h)$. For Trinity, forecasted toll cost $\tilde{c}^d_n(h)$ is based on individual forecasted account balance on day $d$. Thus, the utility of an individual $n$ can be written as

$$U^d_n(h) = V^d_n(h) + \epsilon^d_n(h)$$

(11)

$$= -\alpha \tilde{t}^d(h) - \beta_E SDE(h,t^*) - \beta_L SDL(h,t^*)$$

(12)

$$- \tilde{c}^d_n(h) + \epsilon(h)$$

(13)
where

\[ SDE(h,t^*) = \max(0, t^* - \Delta - (h + \hat{r}^d(t_h))) \]  \hspace{1cm} (14) \]

\[ SDL(h,t^*) = \max(0, (h + \hat{r}^d(t_h)) - t^* - \Delta) \]  \hspace{1cm} (15) \]

\( \hat{c}^d_n(h) \) represents forecasted toll cost and depends on individual forecasted account balance as follows:

\[
\hat{c}^d_n(h) = \begin{cases} 
\max(S^d(T(h), t), 0) & \hat{x}_n^d(h) \geq T(h) \\
B^d(T(h) - \hat{x}_n^d(h), t) + \max(S^d(\hat{x}_n(h)), t), 0) & \text{otherwise}
\end{cases} \]  \hspace{1cm} (16) \]

If forecasted account balance is greater than toll in tokens \((T(h))\), the opportunity cost of paying toll is equal to selling revenue of those tokens; otherwise, the opportunity cost is equal to selling revenue of current balance plus buying cost of additional tokens. If selling price does not decay and transaction costs are zero, toll cost reduces to toll in tokens times token market price on day \(d\)

\[ \hat{c}^d_n(h) = T(h)p^d. \]

4.2.2. Network model

The network is assumed to be a single origin-destination pair connected by a single path containing a bottleneck of fixed capacity \(s\) (Arnott et al., 1990). A first-in-first-out (FIFO) queue develops once the flow of traveler exceeds \(s\). The free flow travel time is \(t_0\) and the waiting time for a traveler at time step \(t\) is \(t_v(t)\). Thus, the total travel time for a traveler at time \(t\) is:

\[ \tau(t) = t_v(t) + t_0 \]  \hspace{1cm} (17) \]

Let \(D(t)\) be the number of travelers in the queue at time \(t\). The waiting time at time \(t\) is derived from the deterministic queuing model as follows:

\[ t_v(t) = \frac{D(t)}{s} \]  \hspace{1cm} (18) \]
where $D(t) = 0$ and $t_v(t) = 0$ when there is no congestion.

Note that a time interval contains several time steps and travel time of a time interval is equal to averaged travel time of all travelers departing in that time interval.

### 4.2.3. Demand supply interactions

Let $d$ denote the index for the current day and $\tau^{d-1}(h)$ be the experienced travel time on day $d - 1$ at time interval $h$. As we specified in mobility model, travelers are assumed to make their choices of departure time according to forecasted travel time $\tilde{\tau}^d(h)$ from their memory and learning. In this study, we use an exponential smoothing filter, a type of homogeneous filter Cantarella and Cascetta (1995), to model the learning and forecasting process by weighting actual and forecasted costs of previous day as follow:

$$
\tilde{\tau}^d(t) = \theta_{\tau} \tilde{\tau}^{d-1}(t) + (1 - \theta_{\tau}) \tau^{d-1}(t), \; \forall \theta_{\tau} \in [0, 1] \tag{19}
$$

where $\theta_{\tau}$ is learning weight for previous day’s forecasted travel time.

We apply similar filter on individual forecasted departure time $t\tau^d_m$ on day $d$ also as follows:

$$
\tilde{t}\tau^d_m = \theta_{t\tau} \tilde{t}\tau^{d-1}_m + (1 - \theta_{t\tau}) t\tau^{d-1}_m, \; \forall \theta_{t\tau} \in [0, 1] \tag{20}
$$

With forecasted departure time, we can apply rule-based trading model to get forecasted individual account balance.

### 4.3. Trinity with Token Supply Cap

In addition to generic Trinity described in previous section, we also consider a variant Trinity with token supply cap in this study, named as Trinity TSC. For Trinity TSC, travelers are not able to sell tokens. Also, there is no token price adjustment and price is fixed as $1$. However, regulator controls the maximum number of tokens can be bought —token supply cap— as another measure of
control, in order to control the maximum number of vehicles on road (desired flow cap). Regulator would set the token supply cap equal to the product of the desired flow cap and forecasted average number of tokens purchased. As a result of token supply cap, if a traveler cannot buy extra tokens to travel, she has to postpone and wait in a queue to buy tokens to travel. Essentially, Trinity TSC manages travel flow with a strict quantity control. Changes of travel behavior model, network model, and demand supply interactions are discussed in following sections.

4.3.1. Travel behavior model

Due to token supply cap controlled by regulator, some travelers have to postpone their departures if they cannot buy enough tokens to pay toll. Thus, we need to consider forecasted delay due to postponement ($\tilde{\omega}^d(h)$) for each time interval $h$ in mobility model. For a traveler, the marginal utility of an additional unit of delay due to postponement is $\beta_D$. Based on the intuition that travelers can spend delay due to postponement at home or doing work, we expect $\beta_D \leq \beta_E$. In addition, the effect of delay due to postponement on other components (forecasted travel time, schedule delay and toll cost) should also be considered. With the fixed $1$ token price and selling disabled, the monetary cost incurred at time interval $h + \tilde{\omega}^d(h)$ is equal to $T(h + \tilde{\omega}^d(h))$ times the fixed market price $p = 1$. Thus the utility of an individual $n$ can be written as:

$$U^d_{\text{H}}(h) = V^d_{\text{H}}(h) + \epsilon^d(h)$$

$$= -\alpha\tilde{\tau}^d(h + \tilde{\omega}^d(h)) - \beta_E SDE(h + \tilde{\omega}^d(h), t^*) - \beta_L SDL(h + \tilde{\omega}^d(h), t^*)$$

$$- \beta_D \tilde{\omega}^d(h) - T(h + \tilde{\omega}^d(h))p + \epsilon(h + \tilde{\omega}^d(h))$$

(21)

(22)

(23)

where forecasted delay time is converted to the number of time intervals delayed as

$$\tilde{\omega}^d(h) = \left\lfloor \frac{\tilde{\omega}^d(h)}{\Delta h} \right\rfloor$$

(24)
SDE and SDL incorporating the effect of delay can be written as:

\[
SDE(h, t^*) = \max(0, t^* - \Delta - (t_{h+\hat{\omega}^d(h)} + \hat{\tau}^d(h + \hat{\omega}^d(h))))
\] (25)

\[
SDL(h, t^*) = \max(0, (t_{h+\hat{\omega}^d(h)} + \hat{\tau}^d(h + \hat{\omega}^d(h))) - t^* - \Delta)
\] (26)

### 4.3.2. Network model

For Trinity TSC, one additional FIFO queue develops once the number of tokens bought in a time interval exceeds the token supply cap set by regulator. In next time interval, delayed travelers will be processed first than those who planned to depart.

### 4.3.3. Demand supply interactions

For Trinity TSC, let \(\omega^{d-1}(h)\) be the experienced delay due to postponement on day \(d - 1\) at time interval \(h\), and \(\psi^{d-1}(h)\) be the realized average number of tokens purchased at time interval \(h\). As we specified in mobility model, travelers are assumed to make their choices of departure time according to forecasted travel time \(\hat{\tau}^d(h)\) and delay \(\hat{\omega}^d(h)\) from their memory and learning. In addition, regulator sets token supply cap as the product of the desired flow cap and forecasted averaged number of tokens purchased. We apply similar exponential smoothing filters to model the learning and forecasting process as follow:

\[
\hat{\omega}^d(h) = \theta_\omega \hat{\omega}^{d-1}(h) + (1 - \theta_\omega)\omega^{d-1}(h), \forall \theta_\omega \in [0, 1]
\] (27)

\[
\hat{\psi}^d(h) = \theta_\psi \hat{\psi}^{d-1}(h) + (1 - \theta_\psi)\psi^{d-1}(h), \forall \theta_\psi \in [0, 1]
\] (28)

where \(\theta_\omega\) and \(\theta_\psi\) are learning weights on previous day’s forecasted delay and average number of tokens purchased.
4.4. Simulation-based optimization

In this study, the social welfare is adopted to measure the performance of NT, CP and Trinity. For NT, the social welfare is equal to the consumer surplus, which is the sum of individual experienced utilities. For CP, the social welfare is equal to the sum of consumer surplus and regulator revenue. Regulator revenue is the sum of users’ out-of-pocket costs, which cancels out with toll costs considered in experienced utilities. Thus, the social welfare of CP is equivalent to the sum of travel time cost, schedule delay cost and unobserved attributes.

The social welfare of Trinity is equal to the sum of consumer surplus, regulator revenue, user revenue and monetary endowment. Regulator revenue is the sum of users’ out-of-pocket payment minus cost of regulator paying for users’ selling. User revenue is the sum of individual selling revenue. Monetary endowment is the sum of toll paid by endowment tokens because users get endowment tokens for free. Because money transfer has no net impact on social welfare, welfare equals to the combination of experienced travel time cost, schedule delay cost, and unobserved attributes in the end. It can be written as

\[
SW^d = \sum_{n=1}^{N} \left[ -\alpha \tau_n^d(t_{d,n}^d) - \beta E\text{SD}(t_{d,n}^d, t^*) - \beta L\text{SDL}(t_{d,n}^d, t^*) + \epsilon_n^d(t_{d,n}^d) \right]
\]

For Trinity with token supply cap, its welfare also includes experienced delay due to postponement cost.

Optimization is conducted to find the optimal toll (in dollars for CP and in tokens for Trinity) which leads to best social welfare. To facilitate optimization, we represent toll profile by a (mixture) of Gaussian functions and optimize function parameters instead. Noted that our optimization problem has no closed-form objectives for the stochastic dynamic simulation, but can be formulated as a simulation-based optimization problem. Our simulation captures the system model presented in 4.2 including all detailed traveler and regulator states and actions along with the resulting network and market conditions.
To solve our simulation-based optimization, a Bayesian optimization (BO) approach is adopted as it can approximate the simulation-based objective function using a few evaluations.

### 4.4.1. Bayesian optimization

Bayesian optimization essentially has two iterative steps. First, a model that approximates a complex map from the input points (i.e., the parameters to be evaluated) to the output (i.e., objective function value) is updated by adding new pairs of input and output. Second, the new input point is determined by optimizing an acquisition function.

In this paper, we assume that the objective function values with respect to different input points are joint distributed, and adopt a Gaussian Process (GP) to model our social welfare as follows,

\[
U(X) \sim \mathcal{GP}(\mu(X), k(X, X'))
\]

where \( \mu(X) \) is the mean function and \( k(X, X') \) is the covariance kernel function of the GP. It is worth noting the mean function is defined as the mean of the input points, and the Matern kernel is used. \( X \) is the input points which consist of vectors of parameters to be evaluated. The toll (or token) profile in CP (Trinity) is assumed to be a Gaussian curve with one peak, which has three parameters: mean, variance and peak value. Specifically, each input point in CP, denoted as \( X_P \), consists of these three parameters, while the input point in Trinity, \( X_T \), has another (market) parameter \( \Delta p \).

Besides, upper confidence bound (UCB) is used as the acquisition function, which has a good performance in practice. To further improve the efficiency and reduce the number of runs of simulation, we use 'Latin Hypercube Sampling' to generate the initial sample points of \( X \).

### 4.5. Experiments design

In this study, we propose different sets of numerical experiments using system model described in previous section to assess three scenarios (NT, CP and
Trinity) and demonstrate: 1) Trinity functions as expected, 2) the efficiency of Trinity, 3) the robustness (adaptiveness) of Trinity and 4) market behavior of Trinity with transaction costs and decaying selling price. In this section, we describe experiment settings and experiment results are reported in section 5. Simulation time step is set to 1-min and departure time interval considered is 5-min. Choice-set of every individual is the same, which include 157 5-min time intervals ranging from midnight to 1PM. Other parameter values we use are shown in Table 2.

Table 2: Simulation variables

| Variables | Description                     | Values        |
|-----------|---------------------------------|---------------|
| $N$       | Population                      | 10,000        |
| $\alpha$  | Value of time                   | $15/hr        |
| $\beta_E$ | Value of schedule delay early   | $9/hr         |
| $\beta_L$ | Value of schedule delay late    | $36/hr        |
| $\mu$     | Scale parameter of mobility model| 0.36          |
| $t_0$     | Free flow time                  | 15 mins       |
| $t^*$     | Center of on-time arrival period| 8:15AM        |
| $\Delta$  | Range of on-time arrivals       | 30 mins       |
| $s$       | Bottleneck capacity             | 95/min        |
| $\theta_{\tau}/\theta_{\omega}/\theta_{\psi}$ | Learning weights | 0.9          |
| $L$       | Token lifetime                  | 1440 mins     |
| $\Delta_t$| Simulation step                 | 1 min         |
| $\Delta_h$| Departure time interval         | 5 min         |

First of all, we conduct simulations to inspect the functionality of Trinity and compare its performance with NT and CP. For NT, it does not need any toll profile, while for CP, its toll profile is optimized through Bayesian Optimization assuming actual conditions are the same as anticipated conditions. In other words, social welfare of CP with optimal toll profile is the best we can achieve.
Regarding Trinity, it uses the same toll profile as CP but in unit of tokens. With appropriate token allocation and functional price adjustment mechanism, token price is supposed to reach $1 at equilibrium. Thus, Trinity should perform the same as CP and attain revenue neutrality.

Second, we perform simulations to study the efficiency of Trinity. Assume actual road capacity is lower than anticipated capacity and regulator does not want to change toll profile because of practical difficulties, we are interested in how different instruments perform with toll profile based on anticipated capacity. Clearly, CP with toll profile based on anticipated road capacity cannot achieve the system optimum anymore. For lower capacity, the best social welfare we can have is CP with toll profile based on lower capacity. We are interested in if Trinity with toll profile based on anticipated capacity can recover some welfare loss with price adjustment. In addition, because of practical difficulties such as continuously changeable charges, step-toll road pricing schemes are adopted in Singapore and Stockholm. Assume regulator in this study adopts a five-step toll profile similar to that of Singapore, we compare CP and Trinity with step-tolls to they with continuous tolls. We also look at the effect of optimal step-toll versus sub-optimal step-toll on CP and Trinity.

Third, we evaluate the robustness and adaptiveness of Trinity. We look into how Trinity and CP respond to an unusual event, such as a sudden and temporary road capacity reduction because of weather or accidents. In particular, we assume road capacity drops 25% from 6AM to 10AM and regulators can intervene but they cannot change toll profile. Regulator of CP can broadcast road capacity drop information and forthcoming travelers are able to plan their trips accordingly with greater perceived travel time. Regulator of Trinity can additionally adjust token market price and if regulator adopts continuous allocation, she is able to adjust allocation rate too.

Finally, we study the effect of transaction costs. The introduction of transaction costs can reduce undesired speculations (Brands et al., 2020), but it diverts the system from the desired equilibrium regardless of its magnitude (Bao et al., 2014; Nie, 2012). In this study, we define behaviors of selling tokens which would
be bought back later as ‘undesired speculation’ and investigate the effect of two different forms of transaction cost, including 1) a fixed fee for every transaction and 2) a fee per unit of trading credit. Specifically, we look into the trade-off of social welfare and undesired speculation as transaction costs increase.

5. Results and Discussion

In this section, we report results of extensive experiments we conduct. First, we examine efficiency of Trinity TSC and Trinity when 1) anticipated road capacity is the same as actual road capacity, 2) actual capacity is lower than anticipated capacity by 25% and 3) using 5-step toll profile (as commonly used in practice) for Trinity only. Second, we evaluate robustness of Trinity under a sudden capacity reduction and usefulness of continuous allocation. Third, we look into undesired market behavior and analyze the role of the transaction costs and decaying selling price.

5.1. Efficiency

To begin with, we inspects the functionality of Trinity TSC assuming anticipated road capacity is the same as actual road capacity. Recall that Trinity TSC utilizes a strict token quantity control which only allows buying tokens. Regulator is able to control token supply cap per time interval in order to control desired flow cap (DFC) per time interval. Also, it uses the same toll profile in tokens as CP and has a fixed token price as $1. Thus, it should perform the same as CP when regulator does not control token supply cap and achieves the system optimum with toll profile based on anticipated road capacity. Regardless of magnitude, allocation rate has no effect on social welfare as long as desired flow cap is not constraining. However, it applies an effect once desired flow cap becomes constraining as shown in Figure 3. This is because flow is controlled through token supply cap, which depends on average number of tokens purchased.

Then, we inspects the efficiency of Trinity TSC assuming actual road capacity is lower by 25% than anticipated capacity in planning phase and regulator
Figure 3: Effect of AR and DFC on social welfare of Trinity TSC when anticipated capacity is the same as actual capacity. When DFC is not constraining and AR is large enough, Trinity TSC performs the same as CP, which is the system optimum; the performance of Trinity TSC decreases as DFC and AR decrease.

does not want to update toll profile. CP with toll profile based on anticipated capacity does not perform the best and CP with toll profile based on actual capacity is the system optimum we can achieve. For Trinity TSC with toll profile based on anticipated capacity, among all flat desired flow caps we have tried, it cannot perform better than CP with toll profile based on anticipated capacity as shown in Figure 4. The best Trinity TSC can perform is the same as CP with toll profile based on anticipated capacity when desired flow cap is unconstraining.

In order to better understand why additional control, token supply cap, does not work, we assume regulator of Trinity TSC has some knowledge about optimal flow pattern from CP with toll profile based on lower road capacity (SO) and sets token supply cap based on knowledge of it. Indeed, Trinity TSC
Figure 4: Effect of AR and DFC on social welfare of Trinity TSC when actual capacity is lower than anticipated capacity. When DFC is not constraining, Trinity TSC performs the same as CP with toll profile based on anticipated capacity, which is lower than system optimum; the performance of Trinity TSC decreases as DFC decreases.

achieves almost the same flow pattern as Benchmark. As shown in cumulative flow pattern (Figure 5), cumulative departure and arrival of Trinity TSC overlap with cumulative departure and arrival of Benchmark almost completely. However, as shown in welfare plot (Figure 5), Trinity TSC has lower welfare than that of Benchmark. From components of welfare, we observe that Trinity TSC has similar experienced travel time cost and schedule delay cost as Benchmark because of similar flow pattern and 0 value of delay due to postponement; but Trinity Step 1 has lower unobserved attributes values compared to Benchmark. This implies that travelers in Trinity Step 1 are constrained in decision making to choose departure intervals with lower unobserved attributes and suffer from less freedom of making their departure time choice.

Assume anticipated capacity is the same as actual capacity and ignore trans-
Figure 5: Cumulative flow and social welfare comparisons of Benchmark and Trinity TSC. Cumulative actual departure and arrival of Trinity TSC are the same as those of Benchmark. Shaded area is desired time period for arrival. Trinity TSC has lower welfare value because of restrictions in decision making.

action costs and selling price decaying, Trinity with selling enabled and day to day price adjustment can achieves the system optimum if using toll profile in tokens based on anticipated capacity. As shown in Figure 6, different allocation rates lead to different equilibrium prices and social welfare. We are able to find a particular allocation rate, which is 78% of maximum toll in tokens charged, to achieve equilibrium price as $1 and optimal social welfare. In addition, regardless of allocation rate, Trinity is able to achieve revenue neutrality because of functional price adjustment.

Now assume actual road capacity is lower than anticipated capacity by 25% and regulator does not change toll profile. CP with toll profile based on anticipated capacity achieves sub-optimal social welfare while CP with toll profile based on lower capacity achieves optimal social welfare. However, Trinity with toll profile based on anticipated capacity is able to perform better than CP with toll based on anticipated capacity because of price adjustment as shown in Figure 7. As we can see, with allocation rate as 70% of maximum toll in tokens charged, Trinity reaches equilibrium price as $1.45 and equilibrium social welfare closer to the system optimum. In addition, Trinity achieves revenue
Day to day price adjustment of different allocation rates

Equilibrium social welfare of different allocation rates

Equilibrium regulator revenue of different allocation rates

Figure 6: Equilibrium price, social welfare, regulator revenue of Trinity by allocation rates. Ignore transaction costs and selling price decaying, Trinity has a functional market place, in which users can buy and sell to regulator. Assume anticipated capacity is the same as actual capacity, Trinity with toll profile based on anticipated capacity and a suitable allocation rate performs the same as CP and achieves the system optimum. Trinity also achieves revenue neutrality because of price adjustment mechanism.

5.2. Robustness

Regarding evaluating the robustness of Trinity, we adopt a symmetric 5-step toll similar to that of Singapore. Without optimizing step-toll profile, Trinity with step toll can perform better than CP with step toll because of price adjustment as shown in Figure 8. With optimal step-toll, Trinity and CP
ignore transaction costs and selling price decaying, trinity has a functional market place, in which users can buy and sell to regulator. assume actual capacity is the lower than anticipated capacity by 25%, trinity with toll profile based on anticipated capacity and a suitable allocation rate performs almost the same as cp with toll based on lower capacity (so). trinity also achieves revenue neutrality because of price adjustment mechanism.

have similar performance as shown in figure 9.

next, we look into the robustness and adaptiveness of trinity assume there is road capacity drop (because of weather or major accidents) from 6am to 10am on the 10th day when simulation has already reached an stationary state. trinity regulator has flexibility to regulate token market price, allocation rate, and transaction cost while users who have not traveled can update their plan according to new information. users would also expect a travel time increase during
Figure 8: Equilibrium price, social welfare, regulator revenue of Trinity by allocation rates. Ignore transaction costs and selling price decaying, Trinity with a sub-optimal step-toll and a suitable allocation rate performs better than CP with the same step-toll. Trinity also achieves revenue neutrality because of price adjustment mechanism.

peak period between 6AM and 10AM. From optimization, the scale of travel time increase is 1.4 and the period of travel time increase is from 6:40AM to 9:35AM. For Trinity with continuous allocation without selling price decaying, we optimize token market price, allocation rate, and transaction costs that can be controlled by regulator. It can perform much better than no intervention except broadcasting news, which makes travelers re-plan with increased perceived travel time for a period as shown in Figure 10. For Trinity with lump sum allocation, regulator can only regulates token market price. The optimum it can
Ignore transaction costs and selling price decaying, Trinity with a optimal step-toll and a suitable allocation rate performs similar to CP with the same step-toll. Trinity also achieves revenue neutrality because of price adjustment mechanism.

In intuition, this is because toll cost term in utility specification of continuous allocation additionally depends on allocation rate and transaction costs, which provides regulator more degree of freedom to regulate.

5.3. Market behavior

Finally, we analyze the effect of transaction costs and decaying selling price. In this study, we define undesired transactions as selling tokens which would be bought back later. With proportional transaction costs of buying and selling
fixed at 3% (commonly used by e-commerce platforms), we vary fixed transaction costs of buying and selling together and present results in Figure 11. For each case, we assume regulator does not want to change toll profile but can optimize allocation rate. As we can see, small fixed transaction cost can reduce undesired transactions significantly and maintain social welfare similar to the system optimum. In addition, We find that although small proportional transaction cost does not reduce social welfare too much, it does not reduce undesired transactions too. Last, we find introducing decaying selling price makes Trinity performs worse. Instead of linearly decaying, we should try different decaying in future experiments.
Figure 11: The trade-off between social welfare and undesired transactions. Without changing toll profile, the introduction of a small fixed transaction cost reduces the percentage of undesired transactions from 60% to almost 0% while Trinity still achieves similar social welfare.

6. Conclusions

This study presented a detailed formulation of a tradable mobility credit scheme, Trinity, with a focus on market design aspects including allocation/expiration of credits, trading rules, transaction costs, price dynamics and regulator intervention. We conducted extensive numerical experiments for a system model of commuters’ departure time choice in a single corridor. The system is modeled using a stochastic process approach that incorporates day-to-day dynamics. It consists of discrete logit departure time choice model, rule-based trading model, deterministic queuing model, and exponential smoothing learning and forecast-
We consider Trinity in congestion pricing applications, which uses congestion pricing toll profile in tokens and has a token market place to adjust token price daily. In addition, we consider a variant of Trinity using strict quantity control (Trinity TSC), which only allows buying and fixes market price at $1. Regulator of Trinity TSC is able to control token supply cap to control actual flow. We formulated toll profile optimization as simulation-based optimizations, which were solved by a Bayesian optimization approach. We conducted simulation experiments of NT, CP, and Trinity assuming homogeneous commuters, which led to following insights:

1. When actual capacity is the same as anticipated capacity, Trinity with toll profile based on anticipated capacity is able to achieve the system optimum ignoring transaction costs and selling price decaying; Trinity TSC with unconstraining desired flow cap can also achieve the system optimum

2. When actual capacity is lower than anticipated capacity by 25%, Trinity with toll profile based on anticipated capacity is able to perform almost the same as system optimum ignoring transaction costs and selling price decaying; Trinity TSC focusing on controlling flow pattern cannot perform better than CP with toll profile based on anticipated capacity

3. Ignoring transaction costs and selling price decaying, Trinity with sub-optimal step-toll performs better than CP with the same toll profile; Trinity with optimal step-toll performs similar to CP with the same step-toll

4. Trinity with continuous allocation has more flexibility for regulator (through regulating token price, allocation rate, and transaction cost) to accommodate for unusual events better than Trinity with lump-sum

5. Trinity with a small fixed transaction cost can reduce undesired transactions significantly while maintain system performance; Proportional transaction costs do not prevent undesired transactions; Linearly selling price decaying makes Trinity have sub-optimal performance.

Finally, more realistic market operation models, more investigations on mar-
ket design and its dynamics, more tests on different decaying selling prices, including heterogeneity, examining adaptive toll profile, switching to large and real network offer some interesting avenues for further work.

Appendix A. Stationarity Test

In statistics, a stochastic process is stationary if unconditional join probability distribution does not change when shifted in time Gagniue (2017). In intuition, it means mean and variance do not change over time. In this study, since our experiments are stochastic, it is important to make sure simulations become stationary as we compute social welfare. Otherwise, it is meaningless to compare social welfare across different scenarios. The most basic methods for stationarity detection is to plot data and if there is any obvious violation. More rigorous approach to detect stationarity is to perform statistic tests. In this study, two statistical tests are used: Augmented Dickey Fuller (ADF) test and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test. ADF test is used to determine the presence of unit root in the series. The existence of unit root means 1 is a root of the process’s characteristic equation and process is not stationary. The null hypothesis of ADF test is that the series has a unit root while the alternate hypothesis is that the series has no unit root. Therefore, if the null hypothesis is failed to be rejected, it provides evidence that the series is non-stationary. KPSS is another test for checking the stationarity. The null hypothesis of KPSS test is that the process is trend stationary while the alternate hypothesis is that the series has a unit root (non-stationary). The null and alternate hypothesis for the KPSS are opposite that of the ADF test and KPSS test complements ADF test. The process is truly stationary if both tests conclude it is stationary. If KPSS indicates stationarity and ADF indicates non-stationarity, it means the process is trend stationary; if KPSS indicates non-stationarity and ADF indicates stationarity, it means the process is difference stationary.
References

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

Araldo, A., Gao, S., Seshadri, R., Azevedo, C.L., Ghafourian, H., Sui, Y., Ayaz, S., Sukhin, D., Ben-Akiva, M., 2019. System-level optimization of multi-modal transportation networks for energy efficiency using personalized incentives: Formulation, implementation, and performance. Transportation Research Record, 0361198119864906.

Arnott, R., De Palma, A., Lindsey, R., 1990. Economics of a bottleneck. Journal of urban economics 27, 111–130.

Bao, Y., Gao, Z., Xu, M., Yang, H., 2014. Tradable credit scheme for mobility management considering travelers' loss aversion. Transportation Research Part E: Logistics and Transportation Review 68, 138–154.

Bao, Y., Gao, Z., Yang, H., Xu, M., Wang, G., 2017. Private financing and mobility management of road network with tradable credits. Transportation Research Part A: Policy and Practice 97, 158–176.

Bao, Y., Verhoef, E.T., Koster, P., 2019. Regulating dynamic congestion externalities with tradable credit schemes: Does a unique equilibrium exist? Transportation Research Part B: Methodological 127, 225–236.

Ben-Akiva, M., Cyna, M., De Palma, A., 1984. Dynamic model of peak period congestion. Transportation Research Part B: Methodological 18, 339–355.

Ben-Akiva, M., Watanatada, T., 1981. Application of a continuous spatial choice logit model. Structural analysis of discrete data with econometric applications , 320–343.

Brands, D.K., Verhoef, E.T., Knackaert, J., Koster, P.R., 2020. Tradable permits to manage urban mobility: market design and experimental implementation. Transportation Research Part A: Policy and Practice 137, 34–46.
Cantarella, G.E., Cascetta, E., 1995. Dynamic processes and equilibrium in transportation networks: towards a unifying theory. Transportation Science 29, 305–329.

Cascetta, E., Cantarella, G.E., 1991. A day-to-day and within-day dynamic stochastic assignment model. Transportation Research Part A: General 25, 277–291.

De Palma, A., Lefevre, C., 1983. Individual decision-making in dynamic collective systems. Journal of Mathematical Sociology 9, 103–124.

Dogterom, N., Ettema, D., Dijst, M., 2017. Tradable credits for managing car travel: a review of empirical research and relevant behavioural approaches. Transport Reviews 37, 322–343.

Fan, W., Jiang, X., 2013. Tradable mobility permits in roadway capacity allocation: Review and appraisal. Transport Policy 30, 132–142.

Gagniuc, P.A., 2017. Markov chains: from theory to implementation and experimentation. John Wiley & Sons.

Goddard, H.C., 1997. Using tradeable permits to achieve sustainability in the world’s large cities: policy design issues and efficiency conditions for controlling vehicle emissions, congestion and urban decentralization with an application to Mexico City. Environmental and Resource Economics 10, 63–99.

Grant-Muller, S., Xu, M., 2014. The role of tradable credit schemes in road traffic congestion management. Transport Reviews 34, 128–149.

Gulipalli, P.K., Kockelman, K.M., 2008. Credit-based congestion pricing: A Dallas-Fort Worth application. Transport Policy 15, 23–32.

He, F., Yin, Y., Shirmohammadi, N., Nie, Y.M., 2013. Tradable credit schemes on networks with mixed equilibrium behaviors. Transportation Research Part B: Methodological 57, 47–65.
Kockelman, K.M., Kalmanje, S., 2005. Credit-based congestion pricing: a policy proposal and the public’s response. Transportation Research Part A: Policy and Practice 39, 671–690.

Nie, Y.M., 2012. Transaction costs and tradable mobility credits. Transportation Research Part B: Methodological 46, 189–203.

Nie, Y.M., 2015. A new tradable credit scheme for the morning commute problem. Networks and Spatial Economics 15, 719–741.

Nie, Y.M., Yin, Y., 2013. Managing rush hour travel choices with tradable credit scheme. Transportation Research Part B: Methodological 50, 1–19.

de Palma, A., Lindsey, R., 2011. Traffic congestion pricing methodologies and technologies. Transportation Research Part C: Emerging Technologies 19, 1377–1399.

Raux, C., 2007. Tradable driving rights in urban areas: their potential for tackling congestion and traffic-related pollution. The Implementation and Effectiveness of Transport Demand Management Measures, 95.

Shirmohammadi, N., Zangui, M., Yin, Y., Nie, Y., 2013. Analysis and design of tradable credit schemes under uncertainty. Transportation Research Record 2333, 27–36.

Small, K.A., 1982. The scheduling of consumer activities: work trips. The American Economic Review 72, 467–479.
Song, X., Danaf, M., Atasoy, B., Ben-Akiva, M., 2018. Personalized menu optimization with preference updater: a boston case study. Transportation Research Record 2672, 599–607.

Tian, L.J., Yang, H., Huang, H.J., 2013. Tradable credit schemes for managing bottleneck congestion and modal split with heterogeneous users. Transportation Research Part E: Logistics and Transportation Review 54, 1–13.

Tsekeris, T., Voß, S., 2009. Design and evaluation of road pricing: state-of-the-art and methodological advances. NETNOMICS: Economic Research and Electronic Networking 10, 5–52.

Verhoef, E., Nijkamp, P., Rietveld, P., 1997. Tradeable permits: their potential in the regulation of road transport externalities. Environment and Planning B: Planning and Design 24, 527–548.

Vickrey, W.S., 1969. Congestion theory and transport investment. The American Economic Review 59, 251–260.

Wang, G., Gao, Z., Xu, M., 2014. An mpec formulation and its cutting constraint algorithm for continuous network design problem with multi-user classes. Applied Mathematical Modelling 38, 1846–1858.

Wang, X., Yang, H., Zhu, D., Li, C., 2012. Tradable travel credits for congestion management with heterogeneous users. Transportation Research Part E: Logistics and Transportation Review 48, 426–437.

Weitzman, M.L., 1974. Prices vs. quantities. The review of economic studies 41, 477–491.

Wu, D., Yin, Y., Lawphongpanich, S., Yang, H., 2012. Design of more equitable congestion pricing and tradable credit schemes for multimodal transportation networks. Transportation Research Part B: Methodological 46, 1273–1287.

Xiao, F., Qian, Z.S., Zhang, H.M., 2013. Managing bottleneck congestion with tradable credits. Transportation Research Part B: Methodological 56, 1–14.
Yang, H., Wang, X., 2011. Managing network mobility with tradable credits. Transportation Research Part B: Methodological 45, 580–594. URL: http://dx.doi.org/10.1016/j.trb.2010.10.002, doi:10.1016/j.trb.2010.10.002.

Ye, H., Yang, H., 2013. Continuous price and flow dynamics of tradable mobility credits. Transportation Research Part B: Methodological 57, 436 – 450.

Zhu, D.L., Yang, H., Li, C.M., Wang, X.L., 2015. Properties of the multiclass traffic network equilibria under a tradable credit scheme. Transportation Science 49, 519–534.