Traffic & Mobility

TA2: Traffic Equilibrium

Thursday 8:30 – 10:30 AM

Session Chair: Carlo Prato

8:30 **Analyzing Departure Time Choice in a Bottleneck with Stochastic Service Time**
Gege Jiang, Hong Lo*
Hong Kong University of Science and Technology

9:00 **An Algorithm for Transit Assignment Problem with Flow-Dependent Dwell Times**
Yufeng Zhang, Alireza Khani*
University of Minnesota

9:30 **Statistical Inference of Probabilistic Origin-destination Demand Using Day-To-Day Traffic Data**
Wei Ma, Zhen Qian*
Carnegie Mellon University

10:00 **Threshold-Based Stochastic User Equilibrium Models**
*[1]David Watling, [2]Thomas Rasmussen, [2]Otto Nielsen, [3]Carlo Prato**
[1]University of Leeds, [2]Technical University of Denmark, [3]The University of Queensland

Analyzing Departure time Choice in a Bottleneck with Stochastic Service Time

Gege JIANG , Hong K. LO*
E-mail: cehklo@ust.hk
Department of Civil and Environmental Engineering,
The Hong Kong University of Science and Technology, Hong Kong, China

1. Introduction

The bottleneck model has interested researchers for many years since its original development by Vickery (1969). This model describes commuters' trip scheduling profile during the morning peak, and models congestion as queuing behind a single bottleneck, providing simplicity yet the ability to reasonably capture the performance of a bottleneck.

Recently attention has turned to analyzing the bottleneck model under stochastic conditions. Li (2008) calculated the departure time distribution under stochastic conditions and concluded that the presence of random delay can significantly shift the departure time. Fosgerau (2010a) studied the equivalence between the scheduling model and mean-variance model under travel time uncertainty. Xiao et.al (2015) investigated the properties of the stochastic bottleneck model and proposed two schemes for congestion toll. Siu and Lo (2009) linked the departure sequence to the Within Budget Travel-time Reliability (WBTR) requirement. Compared with deterministic bottleneck models (Arnott et. al, 1990), indeed our understanding of bottleneck models under a stochastic context can be much improved.

Previous stochastic bottleneck studies usually calculated the departure profile to equalize the mean cost of each traveler. Few studies have established the connection between travel cost variability and departure time choice. Moreover, they cannot distinguish between people who choose to depart in the more congested period and those who depart in the period with less congestion. In other words, the motivation or incentive for a traveler to choose a specific departure time has not been considered extensively. To fill this gap, this study makes three contributions.

First of all, this study introduces travel cost variation as an additional cost. The equilibrium condition is not only based on mean travel cost but also cost variation. Travelers trade between their cost mean and variance. When user equilibrium is achieved, we can infer their attitudes towards cost variability based on their chosen departure times. Secondly, the randomness of service time and travelers' sensitivity towards variability are investigated. We can conclude that the departure profile of travelers is influenced by both factors. Thirdly, we apply a one-step coarse toll during a specific time interval to investigate the effects of tolling on queuing time as well as cost variation. Such an understanding will shed light on developing traffic management strategies for stochastic bottlenecks.

2. Problem Formulation

Following Fosgerau (2010b), we assume the bottleneck has a capacity of $1/s$ per time unit, and we normalize the total demand to be 1. Therefore, the time to serve all the travelers is s, where s is a random variable with probability density function (PDF) $\phi(s)$ and cumulative probability function (CDF) $\varphi(s)$. Using this assumption, we can accommodate the variability of capacity and demand simultaneously, because s increases with demand and decreases with the capacity. We further assume that the service time falls within a certain range, expressed as: $s \in [\underline{s}, \overline{s}]$. Every day s takes on a specific realization, so that the ratio of

capacity to demand remains the same during the morning peak hour.

Suppose travelers enter the bottleneck with the time interval $[t_0, t_e]$ every day at equilibrium. The departure rate is $r(t)$ and the cumulative number of travelers having entered the bottleneck at time t is

$$R(t) = \int_{t_0}^{t} r(t')dt' \tag{1}$$

Depending on the realized capacity, travelers may have to experience a queue of

$$Q(t) = \max[R(t) - (t - t_0)/s, 0] \tag{2}$$

The condition for the existence of a standing queue at time t on a particular day is

$$s(t) = (t - t_0)/R(t) \tag{3}$$

$s(t)$ is the service time if the queue is exactly zero at time t. If $s > s(t)$, $Q(t)$ will be greater than zero and vice versa. Thus the queuing time is

$$q(t) = \begin{cases} 0, s \le s(t) \\ sR(t) - (t - t_0), s > s(t) \end{cases} \tag{4}$$

Denote the arrival time of the traveler who enters the bottleneck at time t to be t_{arr},

$$t_{arr}(t) = \begin{cases} t, s \le s(t) \\ sR(t) + t_0, s > s(t) \end{cases} \tag{5}$$

Assume the preferred arrival time t^* for all the travelers is 0 for simplicity, and the cost under the realized service time of s is

$$C_s(t) = \alpha q(t) + \beta SDE(t) + \gamma SDL(t) \tag{6}$$

where $SDE(t)$ and $SDL(t)$, respectively, are the schedule delay early and late at time t.

$$\begin{aligned} SDE(t) &= \max[-t_{arr}(t), 0] \\ SDL(t) &= \max[0, t_{arr}(t)] \end{aligned} \tag{7}$$

The mean travel cost of the traveler who enters the bottleneck at time t is:

$$E[C(t)] = \int_{\underline{s}}^{\bar{s}} C_s(t)\phi(s)ds = \alpha E[q(t)] + \beta E[SDE(t)] + \gamma E[SDL(t)] \tag{8}$$

Because of the day-to-day stochastic capacity and demand of the bottleneck, commuters experience different travel cost on each day. The variance of travel cost can be defined as:

$$\sigma^2[C(t)] = \int_{\underline{s}}^{\bar{s}} \left[E[C(t)] - C_s(t)\right]^2 \phi(s)ds = E[C^2(t)] - E^2[C(t)] \tag{9}$$

The traditional bottleneck model only considers mean travel cost but not cost variation. It cannot answer the question that who will depart early or late. Besides, it cannot capture peoples' risk attitude. In our study, we propose that people consider the variance of travel cost in departure time choice, which can be formulated as

following:

$$G(t) = E[C(t)] + \lambda\sigma[C(t)] \tag{10}$$

$G(t)$ is the generalized cost, a linear combination of mean travel cost and standard deviation of travel cost. λ is the coefficient, interpreted as the sensitivity of cost variation in this study. When equilibrium is achieved, all the travelers share the same generalized cost so they have no incentive to change departure time. Thus, travelers depart during different time intervals may have different mean costs. According to their arrival times, travelers can be classified into three categories:

(i) Travelers who always arrive early and experience no schedule delay late.
(ii) Travelers who may arrive early or late, Depending on the realized capacity on the day.
(iii) Travelers who always arrive late and experience no schedule delay early.

For case (i), travelers depart within $[t_0, t_1]$, their travel costs come from queuing delay and schedule early cost. Their mean travel cost is:

$$E[C(t)] = \alpha\int_{s(t)}^{\bar{s}} \left(sR(t) - t + t_0\right)\phi(s)ds - \beta\left[\int_{\underline{s}}^{s(t)} t\phi(s)ds + \int_{s(t)}^{\bar{s}} \left(sR(t) + t_0\right)\phi(s)ds\right] \tag{11}$$

The boundary condition is $-t_0 = \bar{s}R(t)$.

For case (ii), travelers depart within $[t_1, t_2]$, there is a critical service time $\tilde{s}(t) = -t_0 / R(t)$. If $s > \tilde{s}(t)$, the travelers would be late. So the mean travel cost is

$$E[C(t)] = \alpha\int_{s(t)}^{\bar{s}} \left(sR(t) - t + t_0\right)\phi(s)ds - \beta\int_{\underline{s}}^{\tilde{s}(t)} (t + q(t))\phi(s)ds + \gamma\int_{\tilde{s}(t)}^{\bar{s}} (t + q(t))\phi(s)ds \tag{12}$$

The boundary condition is $t = 0$ if $\underline{s} \le s(t)$ and $-t_0 = \underline{s}R(t)$ otherwise.

For case (iii), travelers depart within $[t_2, t_e]$. Their mean travel cost is

$$E[C(t)] = \alpha\int_{s(t)}^{\bar{s}} \left(sR(t) - t + t_0\right)\phi(s)ds + \gamma\left[\int_{\underline{s}}^{s(t)} t\phi(s)ds + \int_{s(t)}^{\bar{s}} \left(sR(t) + t_0\right)\phi(s)ds\right] \tag{13}$$

3. Numerical Studies

For simplicity, we assume *s* follows a uniform distribution, $s \in [1-\delta, 1+\delta]$.We have:

$$\phi(s) = \frac{1}{\bar{s} - \underline{s}} = \frac{1}{2\delta} \tag{14}$$

For the first traveler who departs at t_0, $q(t_0) = 0$.Thus

$$EC(t_0) = -\beta t_0 \tag{15}$$

$$\sigma^2[C(t_0)] = 0 \tag{16}$$

For the last traveler who departs at t_e, $s(t_e)=1, R(t_e)=1$ and the generalized cost can be expressed as a function of t_e. When equilibrium is achieved, each traveler has the same generalized cost:

$$G(t_0)=G(t_e) \Leftrightarrow E[C(t_0)]+\lambda\sigma[C(t_0)]=E[C(t_e)]+\lambda\sigma[C(t_e)] \quad (17)$$

So we can solve the queuing profile for the whole interval with the same equilibrium generalized cost. Due to the complex form of the generalized cost, no analytical expression of $R(t)$ can be derived. However, we can solve $R(t)$ by numerical methods. According to previous research by Xiao et.al (2015), the length of departure interval, $t_e - t_0$, is only slightly influenced by capacity variation. For calculation simplicity, we assume $t_e - t_0 = 1$ which is calculated based on the mean capacity. This assumption might cause subtle inconsistency but the error range is within 0.1% of the actual length of departure interval based on the numerical experiments.

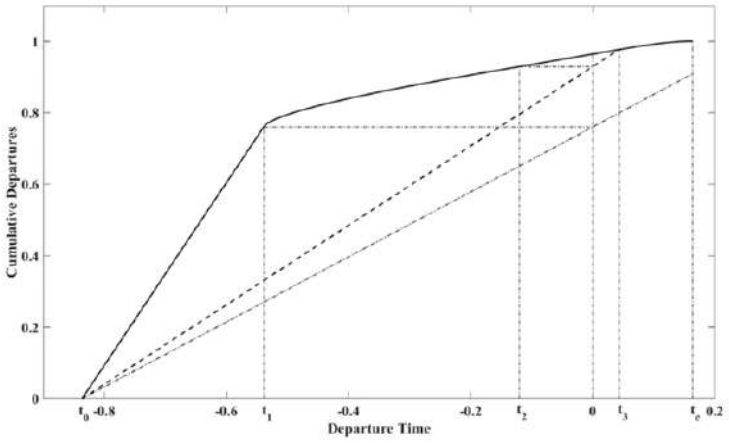


Figure1. The departure profile

Figure 1 shows the departure profile for $\delta=0.1, \lambda=0.5$. The two dashed lines correspond to the maximum and minimum capacity, respectively. t_0, t_e are the start and end times of the departure interval. t_1, t_2 are the watershed points to separate the three cases. Travelers depart after t_3 may or may not experience queue depending on the realization of capacity. If capacity is really large, t_3 can be smaller than t_2.

Figure 2 shows the cost components for travelers entering in different times. The generalized equilibrium cost is 0.5185, the same for each traveler, even though the component costs (variation cost, schedule early/late cost, travel time cost) vary. For travelers departing early, their costs mainly come from schedule early cost. For later departures, the percentage of variation cost gradually increases. The largest variation cost does not synchronize with the maximum queue because we consider both variations of travel time and schedule delay cost. It is easy to understand that travelers departing later are more likely to be influenced. Unlike deterministic case, the mean queuing cost for the last traveler is not zero. That is consistent with the result in Figure 1. When the capacity is low, even the last traveler must endure a queue.

Figure 3 is the departure profile under different δ or service time variability. The case of $\delta=0$ corresponds to the deterministic case. We can see that as δ increases, travelers tend to depart earlier. Since the

capacity/service time is random, we cannot plot out the queue length during the interval, but it can be inferred from Figure 3 that the maximum queue length decreases with δ. We can interpret this result as following. As variation increases, people are more likely to endure schedule late penalty. According to our assumption, $\gamma > \beta$, so they would like to depart earlier to accommodate the potential large cost.

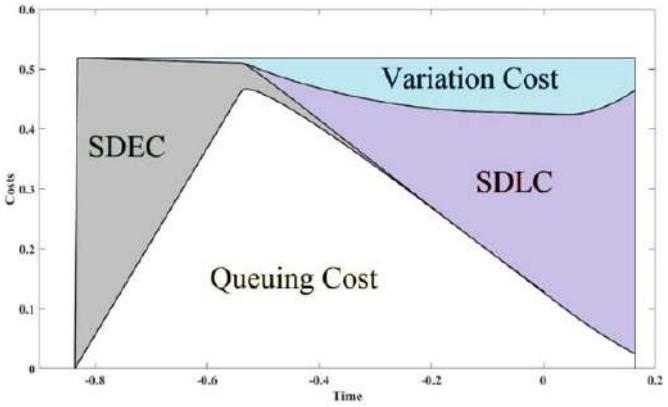


Figure2. Cost Components

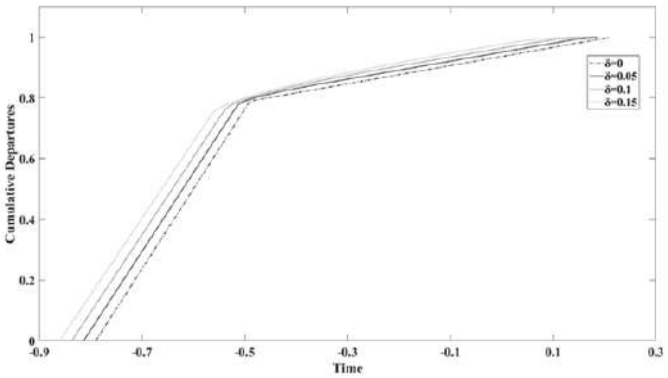


Figure3. The departure profile under different δ

4. Conclusion

This study developed a bottleneck model under stochastic cost variation, in which people's attitude towards cost uncertainty is incorporated in the decision process. From the simple illustrative model, we can see that the stochastic bottleneck model exhibits different characteristics as compared with the classic deterministic models. When equilibrium is achieved, each person has the same generalized cost, but different mean cost and cost variation. The tradeoff between expected cost and uncertain cost is manifested in each person's risk attitude. An interesting point is that this model can produce a mapping between the departure profile and risk attitude of travelers, which does not exist in the original bottleneck model and other stochastic bottleneck models. Although this mapping is not a one-to-one mapping, it can help reveal a person's risk attitudes according to his departure time. This work is only partly done by now. The sensitivity of generalized cost toward cost variability and the effect of one-step toll will be demonstrated in the full paper.

5. Reference

Arnott, R., De Palma, A., & Lindsey, R. (1990). Economics of a bottleneck. Journal of urban economics, 27(1), 111-130.

Fosgerau, M., & Karlström, A. (2010a). The value of reliability. Transportation Research Part B: Methodological, 44(1), 38-49.

Fosgerau, M. (2010b). On the relation between the mean and variance of delay in dynamic queues with random capacity and demand. Journal of Economic Dynamics and Control, 34(4), 598-603.

Li, H., Bovy, P. H., & Bliemer, M. C. (2008). Departure time distribution in the stochastic bottleneck model. International Journal of ITS Research, 6(2).

Siu, B. W., & Lo, H. K. (2009). Equilibrium trip scheduling in congested traffic under uncertainty. In Transportation and Traffic Theory 2009: Golden Jubilee (pp. 147-167). Springer US.

Small, K. A. (1982). The scheduling of consumer activities: work trips. The American Economic Review, 467-479.

Vickrey, W. S. (1969). Congestion theory and transport investment. The American Economic Review, 251-260.

Xiao, L. L., Huang, H. J., & Liu, R. (2015). Congestion behavior and tolls in a bottleneck model with stochastic capacity. Transportation Science, 49(1), 46-65.

AN ALGORITHM FOR TRANSIT ASSIGNMENT PROBLEM WITH FLOW-DEPENDENT DWELL TIMES

Yufeng Zhang
Department of Civil, Environmental, and Geo- Engineering
University of Minnesota Twin Cities
500 Pillsbury Drive SE, Minneapolis MN 55418
zhan4879@umn.edu

Alireza Khani (Corresponding Author)
Department of Civil, Environmental, and Geo- Engineering
University of Minnesota Twin Cities
500 Pillsbury Drive SE, Minneapolis MN 55418
Email: akhani@umn.edu
Phone: +1 (612) 624-4411

Abstract: In a typical transit network, bus dwell times are not constant but are affected by flows of boarding and alighting passengers, whose cumulative property has impacts on how passengers choose their routes towards user equilibrium (UE). This leads to an asymmetric transit assignment with link interactions. In this paper, the problem is formulated as variational inequalities (VI) and a solution algorithm is developed based on a linear constrained generalized Benders decomposition method. Some decent properties unique to transit network are explored, which simplifies the assignment problem by converting the original VI problem to a linear optimization problem. A hypothetical transit network example is included to test the validity and efficiency of the proposed algorithm, which also gives some insights to the proposed method. Computational results show that the proposed algorithm can successfully find UE solutions with high accuracy and a good convergence rate. Simplicity of the algorithm encourages practical applications to schedule-based transit assignment problems with flow-dependent dwell times.

***Keywords*:** Asymmetric Link Interactions, Variational Inequalities, Transit Assignment, User Equilibrium, Variable Dwell Time

INTRODUCTION

The transit equilibrium assignment problem has unique properties compared with general traffic assignment problems. Some complicated and realistic properties such as link interactions make the problem sometimes more difficult to solve. Consider the situation where a bus arrives at a stop, 1) the dwell time is subject to the time consumed by alighting and boarding passengers, and 2) if taking bus capacity into consideration, some of the passengers may be forced to wait for the next bus due to the insufficient capacity on the bus. There exist link interactions in this case, and the extents to which they influence one another is not the same, which is known as asymmetric link interactions. Due to these properties, passengers using one route will exert additional travel cost on this route and other passengers who have multiple route choices may avoid using the excessively used routes to reduce their travel time. Take the transit network shown in figure 1 for example, if there are more passengers boarding at stop B_0 and alighting at stop C_0, then the dwell times on link 10 and 13 increase and at the same time increases the total travel time of the passengers from A_0 to D_0 taking the route consisting of links 7, 2, 10, 3, 13, 4, and 14. Thus, passengers will alternatively use the route comprised of links 5, 1, and 6 to travel from A_0 to D_0. In short, the significant property in this transit assignment problem is that dwell times are functions of flows on boarding and alighting links and this influences how passengers choose their routes. This results in an asymmetric transit assignment problem, which is the problem we aim to solve in this research.

This research develops an algorithm based on Benders decomposition, which basically partitions the network into two separate but interlinked ones. We explored some decent properties of the transit network and further extended the Benders decomposition into asymmetric link interaction transit assignment problems, which converts the problem of solving a generally difficult VI into a linear

optimization problem. Besides, a numerical example is given in the paper to test the validity of the algorithm, explore the efficiency and provide some insights into the algorithm. The computational results show that the algorithm solves the problem effectively and efficiently. The work on this research extended the Benders decomposition to solving a flow dependent bus dwell time characterized transit assignment problem. The methodology is intuitive and easy to implement in practice to solve a more realistic version of the transit assignment problem.

Problem Formulation

A transit network comprised of nodes and arcs is represented as $G(N, A)$, where the set of nodes, *N*, consists of nodes representing both physical bus stops and artificial boarding/alighting nodes, and the elements in the arc set, *A*, represent alighting links, boarding links and dwell-time links. A detailed network representation is introduced as follows.

In this study, three types of links, including alighting links (e.g. link 11), boarding link (e.g. link 12), and dwell-time link (e.g. link 13), are involved in link interactions. Specifically, travel times on dwell-time links are functions of flows on alighting links and boarding links. In contrast, travel time on boarding links and alighting links are functions only of flows on themselves. One can check the asymmetric property that this network exhibits by checking the Jacobian matrix of link travel time functions. The other type of link is in-vehicle links (e.g. link 1, 3) whose travel time is assumed to be flow invariant.

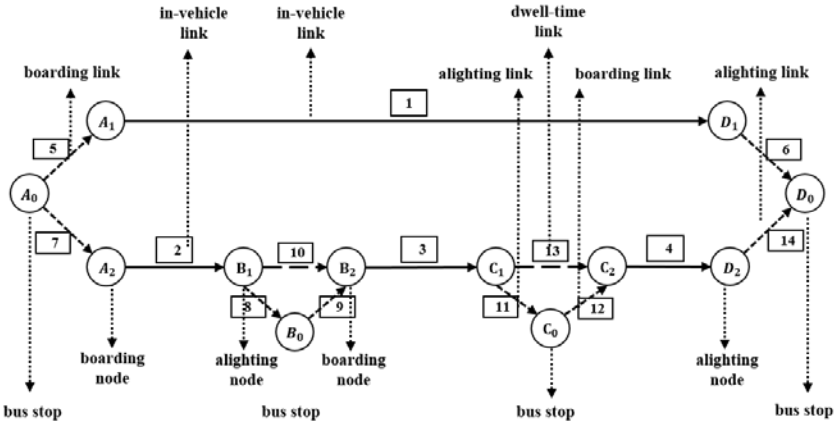


Figure 1. Representation of a transit network with link interactions

Due to variable dwell times sourced from link interactions, passengers' route choice behaviors are subject to the flows (or to be exact, the travel times) on each link. Under the assumptions of UE, all the used routes connecting the same OD pair have the same travel times, which should be less than the travel time on unused routes connecting the same OD pair. We aim to find the flow pattern on the links that satisfies UE. VI is adopted to formulate this transit assignment problem.

The problem is to find $\boldsymbol{x}^* \in S$, such that

$$c(\boldsymbol{x}^*)(\boldsymbol{x} - \boldsymbol{x}^*) \geq 0 \tag{1}$$

s.t.

$$x_{ij} = \sum_{k=1}^{r} x_{ij}^k, \forall ij \in L \tag{2}$$

$$A\boldsymbol{x}^k = d^k, \forall k \in \{1, 2, ..., r\} \tag{3}$$

$$\boldsymbol{x}^k \geq 0, \forall k \in \{1, 2, ..., r\} \tag{4}$$

where $C(\boldsymbol{x})$ is the vector function of link costs, $\boldsymbol{x}$ is the vector of aggregate link flows, *L* is the set of all the links in the network, *A* is the arc-node incidence matrix, *k* represents one OD pair, d^k is the demand of the k-th OD pair, and r is the number of OD pairs in the network.

Literature Review

Variational inequalities were first studied in general traffic assignment problem. Dafermos [1] developed an algorithm for solving VI in the situation where it requires a monotonicity condition and the travel time function must have a special form with a symmetric and positive definite matrix. Simplicial decomposition, another popular technique used to solve VI problems, have been extensively applied in traffic assignment study area. It is developed based on the principle of Dantzig-Wolfe decomposition [2]. An "SDVI" algorithm [3] is proposed for solving asymmetric traffic assignment, which solves the master problem of simplicial decomposition with the projection technique. Nguyen *et al.* [4] applied restricted simplicial decomposition for solving the transit assignment problem with capacity constraints. Wu *et al.* [5] solved the VI with the linearized Jacobi method. They included nonlinear waiting cost and in-vehicle cost in their model, which generalized the model by Spiess *et al.*[6]. There are still handful studies on applying simplicial decomposition to traffic assignment problem, including the research by Larsson and Patriksson on nonlinear column generation [7], and their work on symmetric traffic assignment problem [8]. Most of these works focused on solving the master problem of the simplicial decomposition, which is the key and difficulty in applications of the method. Other methods for solving asymmetric traffic assignment problem take the idea of simplifying the travel cost functions by either making it symmetric or leaving out some unimportant terms, e.g. [9].

METHODOLOGY

The approach of Benders decomposition in [4] transforms the equation (1) to the form shown as follows:

$$C_1(\boldsymbol{x}^*)(\boldsymbol{x}-\boldsymbol{x}^*)+C_2(\boldsymbol{y}^*)(\boldsymbol{y}-\boldsymbol{y}^*)\geq 0,\forall(\boldsymbol{x},\boldsymbol{y})\in S \tag{5}$$

where S is the set of all the feasible flows and is defined by the constraints (2) -(4).

The above formulation divides the VI into two parts. One corresponds to links with symmetric Jacobian matrix, whose link flows are represented by vector $\boldsymbol{x}$, and the other corresponds to the link flows of asymmetric link interactions represented by vector $\boldsymbol{y}$. Correspondingly, $C_1(\cdot)$ and $C_2(\cdot)$ represent the vector functions of travel cost of symmetric part and asymmetric part, respectively.

A typical Benders decomposition method solves a master problem and a sub-problem, which finds the shortest travel time in symmetric part. In a transit network, vector function $C_1(\cdot)$ are costs of in-vehicle links, whose travel time is constant in a schedule-based transit network. Thus the shortest paths in the symmetric part are automatically found. Then, the problem is reduced to solving the problem described in the equation (6).

$$\boldsymbol{\lambda}(\boldsymbol{x}-\boldsymbol{x}^*)+C_2(\boldsymbol{y}^*)(\boldsymbol{y}-\boldsymbol{y}^*)\geq 0,\forall(\boldsymbol{x},\boldsymbol{y})\in S \tag{6}$$

where vector $\boldsymbol{\lambda}$ is the travel times of in-vehicle links, and vector $\boldsymbol{x}$ and $\boldsymbol{y}$ are the link flows of in-vehicle links and the other links involved in link interactions. The solution algorithm for the asymmetric link interaction transit assignment problem is proposed as following:

Step 1 (Initialization): Assign the demands to the routes with the shortest free flow travel times and get an initial flow vector $(\boldsymbol{x}_0, \boldsymbol{y}_0)$. Set iteration label $i = 0$.

Step 2 (Solving the linear program):

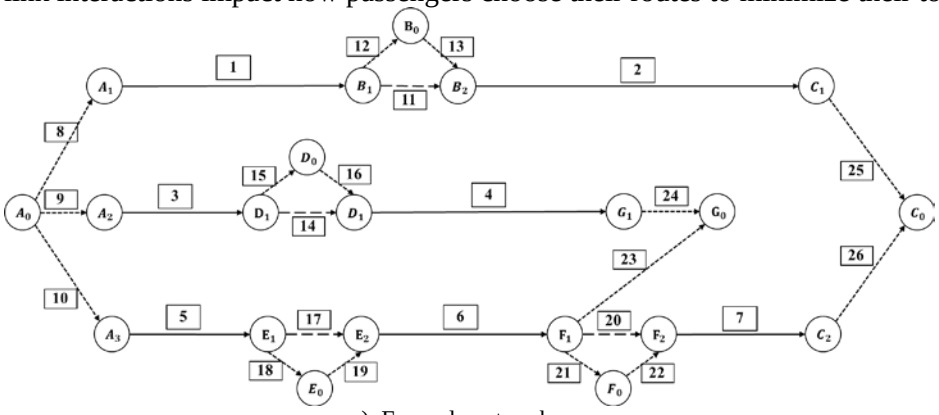
$$\begin{aligned} &Max \quad T_0 \\ s.t. \quad &T_0 \leq \boldsymbol{\lambda}(\boldsymbol{x}_i-\boldsymbol{x})+C_2(\boldsymbol{y}_i)(\boldsymbol{y}_i-\boldsymbol{y}), i=0,1,2,\ldots I \\ &\forall(\boldsymbol{x},\boldsymbol{y})\in S \end{aligned} \tag{7}$$

Step 3 (Convergence check): Find the shortest paths for each OD pair and calculate corresponding shortest travel times κ^k. If $\dfrac{\boldsymbol{\lambda}\boldsymbol{x}+C_2(\boldsymbol{y})\boldsymbol{y}-\sum_{k=1}^{r}\kappa^k d^k}{\sum_{k=1}^{r}\kappa^k d^k}\leq\varepsilon$, stop and $(\boldsymbol{x}, \boldsymbol{y})$ is the optimal solution; otherwise, set $i=i+1$ and go to step 1.

The VI problem is hereto successfully transformed into a linear optimization program.

COMPUTATIONAL STUDIES

A numerical example is given in this section and the algorithm is implemented in Python. The topology of the transit network is shown in Figure 2. In this example, it is assumed that passengers on average take 4 seconds to board the bus and take 2 seconds to alight. Defining $(a, b)_k$ as the alighting flow and boarding flow at the bus stop k, the link cost function for alighting link is $c_a(a) = 2 \times a$, the link cost function for boarding link is $c_b(a, b) = 4 \times b$, and the link cost function for dwell-time link is $c_d(a, b) = 2 \times a + 4 \times b$. Dwell times are functions of boarding flow and alighting flow. Existence of link interactions impact how passengers choose their routes to minimize their total travel times.



a) Example network

O	D
A_0	C_0
A_0	G_0
A_0	F_0
E_0	C_0
E_0	G_0
E_0	F_0
B_0	C_0
D_0	G_0

b) OD pair information

Figure 2. Example transit network

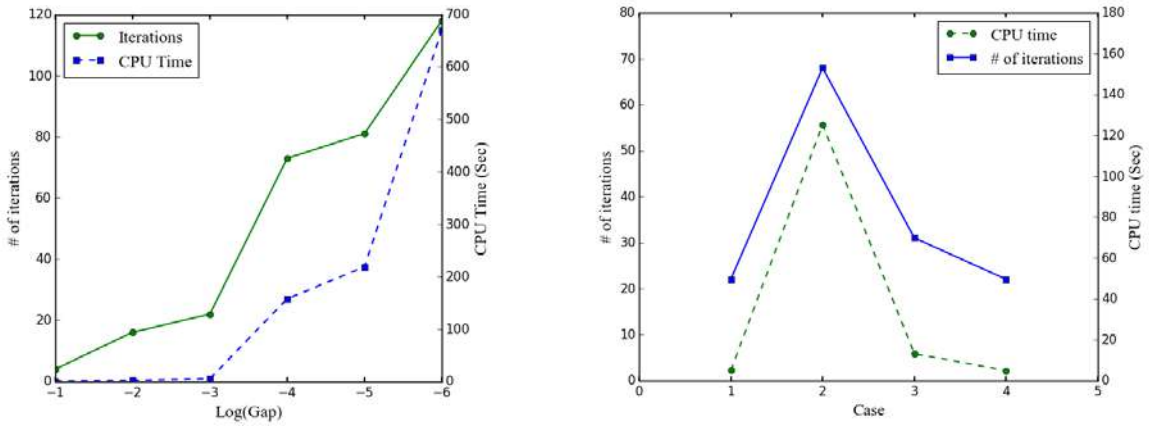
In cases 1 and 2, the travel times of in-vehicle links were fixed, and experiments are conducted by changing the demands between the 8 OD pairs. In contrast, demands were fixed in cases 3 and 4, but in-vehicle times were changed. Detailed information is displayed in Table 1. Terminating error threshold ε in step 3 took the value of 0.1%.

Table 1. Computational results

Case	Travel time (sec)				Computational statistics		
	A-C upper path	A-C lower path	A-G upper path	A-G lower path	Relative gap*	# of iteration	CPU time (sec)
Fixed in-vehicle travel time [90,120,60,90,60,60,90](sec), variable OD demands [A-C, A-G, A-F, B-C, E-C, E-G, E-F, D-G]							
1: [20, 10, 10, 15, 5, 5, 5, 15]	403.69	404.46	271.85	269.03	$7.45\ 10^{-4}$	22	4.85
2: [30, 15, 5, 10, 15, 10, 5, 10]	439.86	439.00	291.14	295.62	$9.82\ 10^{-4}$	68	125.27
Fixed OD demand [20, 10, 10, 15, 5, 5, 5, 15], variable in-vehicle travel times [1, 2, 3, 4, 5, 6, 7] (sec)							
3: [80, 100, 50, 90, 30, 75, 40]	354.73	354.52	265.74	262.77	$9.61\ 10^{-4}$	31	13.11
4: [72, 80, 36, 57, 102, 25, 80]	362.00	368.65	239.76	247.14	$7.81\ 10^{-4}$	22	4.78

Under the user equilibrium assumption, used paths connecting the same OD pair should have the same travel time which should also be less than the travel time on unused paths. In our experiments, the paths connecting the same OD are all used and they exhibited rather close travel times. The "relative gap" defined by the ratio of the total system travel time less the shortest path travel time and the shortest path travel time, were also reported to measure the performance of the algorithm from another perspective. All of the tests reach the gap to 10^{-4}, which is good enough for practical applications.

Another experiment aiming to evaluate the computational times were conducted based on Case 1. Relative gap threshold ε took the value from 10^{-1} to 10^{-6}. Figure 3.a shows that the number of iterations and computational times increase sharply with the increase of accuracy defined by the relative gap. The solution algorithm belongs to the category of cutting plane algorithm [4], which adds a new constraint to the linear program in every iteration. So the computational time increases rapidly with the increase of constraints. Fig 3.b shows the interdependent relationship between number of iterations and CPU times. Usually, a relative gap below 1% can be considered good for practical applications, and according to Figure 3.a, the computational time is acceptable between threshold 10^{-1} and 10^{-3}. So, we can expect a good convergence speed if choosing the above recommended threshold values.



a) Results of case 1 with different relative gap threshold b) number of iterations and CPU time

Figure 3. Computational time and number of iterations

CONCLUSION

Computational results indicate that the solutions obtained converge to user equilibrium. Computational time is related with the number of iterations and due to the property of increasing constraints of the algorithm, time increases sharply but can be controlled within a reasonable range if some good terminating threshold is chosen without a great loss of solution accuracy.

This research further developed Benders decomposition method to the application of solving an asymmetric transit assignment problem which is characterized by the fact that dwell times are functions of boarding and alighting flows. Explorations of some decent properties unique to the transit network allow for the proposal of an easier linear optimization solution algorithm for VI. The solution algorithm performs effectively and efficiently. The results obtained so far are encouraging and the methodology is well worth further development.

Future work will focus on including transfers and congestion and formulating the dual problem, which can convert the addition of constraints to addition of variables to the optimization problem aiming to further reduce the computational time.

REFERENCES

1. Dafermos, S. (1980). Traffic equilibrium and variational inequalities. Transportation science, 14(1), 42-54.

2. Von Hohenbalken, B. (1977). Simplicial decomposition in nonlinear programming algorithms. Mathematical Programming, 13(1), 49-68.

3. Lawphongpanich, S., & Hearn, D. W. (1984). Simplicial decomposition of the asymmetric traffic assignment problem. *Transportation Research Part B: Methodological*, *18*(2), 123-133.

4. Nguyen, S., Pallottino, S., & Malucelli, F. (2001). A modeling framework for passenger assignment on a transport network with timetables. Transportation Science, 35(3), 238-249.

5. Wu, J. H., Florian, M., & Marcotte, P. (1994). Transit equilibrium assignment: a model and solution algorithms. *Transportation Science*, *28*(3), 193-203.

6. Spiess, H., & Florian, M. (1989). Optimal strategies: a new assignment model for transit networks. *Transportation Research Part B: Methodological*, *23*(2), 83-102.

7. Larsson, T., Patriksson, M., & Rydergren, C. (1996). Simplicial decomposition with nonlinear column generation. R*eport, Department of Mathematics, Linköping Institute of Technology, Linköping, Sweden*.
8. Larsson, T., & Patriksson, M. (1992). Simplicial decomposition with disaggregated representation for the traffic assignment problem. *Transportation Science*, *26*(1), 4-17.
9. Wang, S., Yang, H., & He, B. (2000). Solving a class of asymmetric variational inequalities by a new alternating direction method. *Computers & Mathematics with Applications*, *40*(8), 927-937.

Statistical inference of probabilistic origin-destination demand using day-to-day traffic data

Wei Ma, Zhen (Sean) Qian
Department of Civil and Environmental Engineering
Carnegie Mellon University, Pittsburgh, PA 15213

January 4, 2017

The origin-destination (O-D) demand is a critical input to system modeling in transportation planning and management. For decades, O-D demand is deterministically modeled, which leads to deterministic models of link/path flow and travel cost. Recent studies on transportation network uncertainty and reliability call for modeling the stochasticity of O-D demand, namely its spatial correlation and variation. Till now, only a few studies have considered the stochastic features of O-D demand and started to estimate the mean and variance of O-D demand using day-to-day traffic data. However, none of these studies recognize the fact that the variation of traffic data stems not only from O-D variation but also from travelers' independent route choices that vary from day to day. Consequently, those studies are making a biased estimation of O-D demand distribution. In view of this, this paper develops a novel framework for estimating the mean and variance/covariance of probabilistic O-D demand considering the variation induced by travelers' day-to-day random route choices. Furthermore, the statistical explanation, examination and analysis towards the estimated probabilistic O-D demand are also provided.

The classical traffic assignment models [e.g., 1, 2] deterministically map the origin-destination demand $q \in \mathbb{R}_+^{R\times S}$ to link flow $x \in \mathbb{R}_+^N$ or path flow $f \in \mathbb{R}_+^K$. The O-D demand q is assumed to represent the *mean* number of O-D trips in the same peak hour from day to day. Likewise, link (path) flow is also deterministic, representing the *mean* number of vehicles on a link (path) in the same time period of day. The classical traffic assignment model plays a role as the theoretical foundation of most commonly used O-D estimation methods, thus those methods can only estimate the deterministic O-D demand. For example, the entropy maximizing models [3], maximum likelihood models [4, 5, 6], generalized least squares (GLS) models [7, 8, 9], Bayesian inference models [10, 11] and some recent emerging combined models [12, 13, 14].

Both classical traffic assignment models and O-D estimation methods overlook the variation of demand and link/path flow, which is an essential feature for recurrent traffic. Recent studies on network reliability and uncertainty issues start to tackle the stochasticity feature of O-D demand. In these studies, the probability distribution of O-D demand is usually employed as a key input of their methods or models. For example, [15, 16] sampled O-D demand from given multivariate normal distribution (MVN) and evaluated the network performance under equilibrium condition, [17] used similar simulation-based method to evaluate travelers' risk-taking behavior due to probabilistic O-D demand. Both studies indicate the O-D variation is of great importance to network modeling and behavior analysis. Statistical traffic assignment models based on different O-D distribution such as Poisson distribution [18], MVN [19], multinomial distribution [20] are proposed to depict the stochasticity feature of traffic conditions. [21] summarized those formulations and proposed a unified framework for stochastic modeling of traffic flows. Advantages and disadvantages of modeling traffic with those probabilistic distributions are also

discussed by [22]. [23, 24] proposed a reliability-based traffic assignment model (RUE) and extended it to consider different travelers' risk taking behavior. [25] further extended the model to consider the traffic uncertainty and proposed reliability-based statistical traffic equilibrium. [26, 27] proposed a α-reliable mean-excess traffic assignment model which explicitly models the travel time distribution and consider the reliability and uncertainty of the travel time on travelers' route choice behavior.

As a reverse process of statistical traffic assignment models, some of the O-D estimation methods were proposed to estimate the distribution of O-D demand. [28, 29, 30, 31] assumed O-D demand follows independent Possion distribution and formulated a maximum likelihood estimator, whereas the demand covariance between O-D pairs is not considered. [32] proposed a generalized model to estimated the mean and variance of O-D demand by assuming it to be MVN. [33] further extended the model to estimate multi-class O-D and used $L1$ regularizer to shrink the model dimension.

In summary, the above statistical traffic assignment models and O-D estimation methods adopt the same assumption that given Q travelers, the route flow is deterministically determined by assuming the route flow to be the mean of the route choices for all demands, $f = pQ$. Even the route choice probability p is determined by stochastic choice models (such as Probit, Logit, etc.), these models still assume that a fixed number Qp of travelers will choose this specific route everyday, which omits the variation of travelers' probabilistic choices. Implicitly assumed is that the demand is infinitesimal (also known as non-atomic players) where no variation exists. However, recent studies on statistical traffic assignment models indicate that the route flow is the aggregation of random choices of finite number of travelers and thus also random [21, 34, 35, 36, 37]. With a clearly more realistic setting, travelers are atomic. Thus, travelers' route choice follows a multinomial distribution with the probability calculated by the route choice model, $f \sim \text{Multinomial}(Q, p)$. We refer the former assumption as "*fixed portions with stochastic route choice models* " and the latter as "*probabilistic distributions with stochastic route choice models*".

Travelers' route choice variation is an indispensable component of traffic variation. Our previous researches on statistical traffic assignment model show that the travelers' route choice variation takes approximately 20% of the total variation. In behavior science, the variation induced by route choice uncertainty is under intensive exploration. [38] discussed the network toll policy under two extreme travelers risk-taking behavior (risk-averse and risk-prone). [39] built a Markov process to model the day-to-day variation of traffic flow given fixed O-D demand. We show that the ignorance of travelers' route choice variation in O-D estimation method will lead to an overestimate of O-D variation. Various researches [16, 40, 41] show that the variance/covariance matrix of O-D demand have a significant influence on the traffic equilibrium and network conditions. The influence will be particularly significant when several routes share similar costs as their route choice probability are similar, hence the route choice variation is even larger.

To our best knowledge, there is no method that estimates probabilistic O-D demand and considers the travelers' route choice variation simultaneously. To fill up this gap, we first proposed a generalized statistical traffic assignment (GESTA) model incorporating three dimensions of traffic flow variations: O-D demand variation, route choice variation and unobserved errors in our previous research [42]. The hierarchical formulation of GESTA is represented in Equation 1.

$$
\begin{aligned}
Level\ 1: \quad & X_m = X + \epsilon_e && \text{(Unobserved Error)} \\
& \epsilon_e \sim \mathcal{N}(\mathbf{0}, \Sigma_e) && \\
Level\ 2: \quad & X = \Delta F && \\
& F \sim \text{Multinomial}(\tilde{p}Q, \Sigma_f) && \text{(Route choice variation)} \\
Level\ 3: \quad & Q \sim \mathcal{N}(q, \Sigma_q) && \text{(Demend variation)}
\end{aligned}
\tag{1}
$$

where X, F, Q represent the link/path/OD flow random vector, $\Sigma_q, \Sigma_f, \Sigma_e$ represent the covariance matrix of OD/path flow vector and observation errors, x, f, q is the vector of means of link/path/OD flow vector, $\tilde{p} \doteq diag(p)B$ and B is a transition matrix, p represents the route choice probability and Δ is path/link incidence matrix, $\cdot^o$ denotes the observed variables.

In GESTA, level 3 represents the demand variation, which is directly obtained from model assumptions. Level 2 formulation indicates that each traveler makes their route choice independently, thus the path flow vector F follows multinomial distribution. Level 1 represents the unknown errors applied to the observed link flow vector. After simplification and approximation, the path/link flow distributions can be represented by the following proposition.

Proposition 1. *The marginal distribution of F follows,*

$$F \sim \mathcal{N}(f, \Sigma_f)$$

where $f = \tilde{p}q$, $\Sigma_f = \Sigma_{f|q} + \tilde{p}\Sigma_q\tilde{p}^T$, $\Sigma_{f|q}$ is the covariance matrix of path flow conditional on O-D demands $Q = q$, $\Sigma_{f|q} = \Sigma_{f|Q=q}$, the detailed formulation can be found in [42].

Using GESTA as the underlying behavioral model, we estimate probabilistic O-D demand from various data sources. In this paper, a disaggregated O-D demand estimation framework considering travelers' route choice variation is developed to estimate O-D mean vector and variance-covariance matrix iteratively, which decomposes a complex estimation problem into two relatively simpler sub-problems, and therefore the entire solution algorithm is more friendly for practical use. The proposed framework is known as Iterative Generalized Least Square (IGLS)[43]. IGLS contains a large class of statistical algorithms, it is probably one of the most efficient computing tools for estimation problems.

In the sub-problem of estimating O-D demand mean vector q, we extend the statistical interpretation of O-D estimation in [44], both single level and bi-level O-D estimation formulations are discussed. The OD mean estimation with equilibrium constraints formulation is presented in 2.

$$\begin{aligned} \min_{f} \quad & n\left(\Delta^o f - \hat{x^o}\right)^T \Sigma_x^{o-1}\left(\Delta^o f - \hat{x^o}\right) + (q^H - Mf)^T \Sigma_q^{H^{-1}} (q^H - Mf) \\ \text{s.t.} \quad & f \in \Phi^+ \end{aligned} \tag{2}$$

where Φ^+ is the feasible set of means of path flow vector f. The definition of Φ^+ for Deterministic User Equilibrium (DUE) and Stochastic User Equilibrium (SUE) have different formulations but share similar ideas, and we discussed them separately in the paper.

In the sub-problem of estimating O-D demand variance-covariance matrix Σ_q, we utilize the Wishart distributed nature of link flow covariance Σ_x to formulate the basic estimation problem and then apply Lasso regularizer and convex relaxation on the basic formulation. The final formulation as an approximated maximum likelihood estimator of the OD covariance matrix with Lasso regularization is presented in Formulation 3.

$$\begin{aligned} \min_{\Sigma_q} \quad & \|S_x^o - \Sigma_x^o\|_F^2 + \lambda \|\Sigma_q\|_1 \\ \text{s.t.} \quad & \Sigma_x^o = \Delta^o \Sigma_{f|q} \Delta^{oT} + \Delta^o \tilde{p} \Sigma_q \tilde{p}^T \Delta^{oT} \\ & \Sigma_q \in \text{semidefinite}(\Re^{|K_q| \times |K_q|}) \end{aligned} \tag{3}$$

where $\|A\|_F = \sqrt{\text{Tr}(A^T A)}$ and $\|A\|_1 = \sum_{ij} |A_{ij}|$. The former one is known as Frobenius Norm, which is equivalent to the element-wise L_2 norm [45]; the latter one is the element-wise L_1 norm. λ is a penalty scalar, S_x^o is the empirical covariance matrix of the observed link flow.

Another important issue for O-D estimation method is the observability problem [46, 47]. It is well known that the O-D estimation problem is usually a underdetermined system provided with observed link counts only [9]. [48] suggests to use second-order statistical information of the observed link flow to estimated independent Possion distributed O-D demand. By utilizing the second-order information, the Possion distributed O-D demand can be estimated uniquely. [47] proposed to get unique O-D demand by properly select between observed link flow data and history O-D information. [49] recognize that in the statistical O-D estimation problem, as long as the link-path incidence matrix is unimodular, the uniqueness of O-D estimation cannot be achieved. In this paper, the observability of the proposed statistical O-D estimation method is also examined. It can guarantee that the performance of proposed statistical O-D estimation method will be no worse than the conventional deterministic O-D estimation methods.

Our probabilistic OD demand estimation framework is evaluated on both small network and large scale network. We construct probabilistic OD demands based on the previous studies or history OD and then treat it as the "true" OD demand. Then we simulate the observation data by running the (stochastic) user equilibrium model and add some perturbations on the results. The performance of the OD estimation formulations are assessed by comparing the estimated ODs to the "true" ODs [50]. We evaluate our estimation method under different settings: equilibrium conditions, sensor locations, observation noise, and accuracy of the history OD. We also apply the method to a real-world network SR-41 corridor, and the Pittsburgh regional network to demonstrate its scalability and computational efficiency.

References

[1] Dafermos, S. C. and F. T. Sparrow. The traffic assignment problem for a general network. *Journal of Research of the National Bureau of Standards, Series B*, vol. 73, no. 2, 1969, pp. 91–118.

[2] Patriksson, P. *The traffic assignment problem: models and methods*. 1994.

[3] Van Zuylen, H. J. and L. G. Willumsen. The most likely trip matrix estimated from traffic counts. *Transportation Research Part B: Methodological*, vol. 14, no. 3, 1980, pp. 281–293.

[4] Spiess, H. A maximum likelihood model for estimating origin-destination matrices. *Transportation Research Part B: Methodological*, vol. 21, no. 5, 1987, pp. 395–412.

[5] Watling, D. P. Maximum likelihood estimation of an origin-destination matrix from a partial registration plate survey. *Transportation Research Part B: Methodological*, vol. 28, no. 4, 1994, pp. 289–314.

[6] Li, B. Bayesian inference for origin-destination matrices of transport networks using the EM algorithm. *Technometrics*, vol. 47, no. 4, 2005.

[7] Cascetta, E. Estimation of trip matrices from traffic counts and survey data: a generalized least squares estimator. *Transportation Research Part B: Methodological*, vol. 18, no. 4, 1984, pp. 289–299.

[8] Bell, M. G. The estimation of origin-destination matrices by constrained generalised least squares. *Transportation Research Part B: Methodological*, vol. 25, no. 1, 1991, pp. 13–22.

[9] Yang, H., T. Sasaki, Y. Iida, and Y. Asakura. Estimation of origin-destination matrices from link traffic counts on congested networks. *Transportation Research Part B: Methodological*, vol. 26, no. 6, 1992, pp. 417–434.

[10] Maher, M. Inferences on trip matrices from observations on link volumes: a Bayesian statistical approach. *Transportation Research Part B: Methodological*, vol. 17, no. 6, 1983, pp. 435–447.

[11] Tebaldi, C. and M. West. Bayesian inference on network traffic using link count data. *Journal of the American Statistical Association*, vol. 93, no. 442, 1998, pp. 557–573.

[12] Castillo, E., J. M. Menéndez, and P. Jiménez. Trip matrix and path flow reconstruction and estimation based on plate scanning and link observations. *Transportation Research Part B: Methodological*, vol. 42, no. 5, 2008, pp. 455–481.

[13] Li, B. Markov models for Bayesian analysis about transit route origin–destination matrices. *Transportation Research Part B: Methodological*, vol. 43, no. 3, 2009, pp. 301–310.

[14] Castillo, E., J. M. Menéndez, S. Sánchez-Cambronero, A. Calviño, and J. M. Sarabia. A hierarchical optimization problem: Estimating traffic flow using Gamma random variables in a Bayesian context. *Computers & Operations Research*, vol. 41, 2014, pp. 240–251.

[15] Waller, S. T., A. Unnikrishnan, and J. Duthie. Network Evaluation with Uncertain and Correlated Long-Term Demand. In *Transportation Research Board 85th Annual Meeting*, 06-1767. 2006.

[16] Duthie, J. C., A. Unnikrishnan, and S. T. Waller. Influence of demand uncertainty and correlations on traffic predictions and decisions. *Computer-Aided Civil and Infrastructure Engineering*, vol. 26, no. 1, 2011, pp. 16–29.

[17] Chen, A., Z. Ji, and W. Recker. Travel time reliability with risk-sensitive travelers. *Transportation Research Record: Journal of the Transportation Research Board*, , no. 1783, 2002, pp. 27–33.

[18] Hazelton, M. L. Day-to-day variation in Markovian traffic assignment models. *Transportation Research Part B: Methodological*, vol. 36, no. 7, 2002, pp. 637–648.

[19] Castillo, E., J. M. Menéndez, and S. Sánchez-Cambronero. Predicting traffic flow using Bayesian networks. *Transportation Research Part B: Methodological*, vol. 42, no. 5, 2008, pp. 482–509.

[20] Nakayama, S. Effect of providing traffic information estimated by a stochastic network equilibrium model with stochastic demand. *Transportation Research Part C: Emerging Technologies*, 2016.

[21] Nakayama, S. and D. Watling. Consistent formulation of network equilibrium with stochastic flows. *Transportation Research Part B: Methodological*, vol. 66, 2014, pp. 50–69.

[22] Castillo, E., A. Calviño, M. Nogal, and H. K. Lo. On the probabilistic and physical consistency of traffic random variables and models. *Computer-Aided Civil and Infrastructure Engineering*, vol. 29, no. 7, 2014, pp. 496–517.

[23] Shao, H., W. Lam, Q. Meng, and M. Tam. Demand-driven traffic assignment problem based on travel time reliability. *Transportation Research Record: Journal of the Transportation Research Board*, , no. 1985, 2006, pp. 220–230.

[24] Shao, H., W. H. Lam, and M. L. Tam. A reliability-based stochastic traffic assignment model for network with multiple user classes under uncertainty in demand. *Networks and Spatial Economics*, vol. 6, no. 3-4, 2006, pp. 173–204.

[25] Lam, W. H., H. Shao, and A. Sumalee. Modeling impacts of adverse weather conditions on a road network with uncertainties in demand and supply. *Transportation research part B: methodological*, vol. 42, no. 10, 2008, pp. 890–910.

[26] Zhou, Z. and A. Chen. Comparative analysis of three user equilibrium models under stochastic demand. *Journal of Advanced Transportation*, vol. 42, no. 3, 2008, pp. 239–263.

[27] Chen, A. and Z. Zhou. The α-reliable mean-excess traffic equilibrium model with stochastic travel times. *Transportation Research Part B: Methodological*, vol. 44, no. 4, 2010, pp. 493–513.

[28] Vardi, Y. Network tomography: Estimating source-destination traffic intensities from link data. *Journal of the American Statistical Association*, vol. 91, no. 433, 1996, pp. 365–377.

[29] Hazelton, M. L. Estimation of origin–destination matrices from link flows on uncongested networks. *Transportation Research Part B: Methodological*, vol. 34, no. 7, 2000, pp. 549–566.

[30] Hazelton, M. L. Inference for origin–destination matrices: estimation, prediction and reconstruction. *Transportation Research Part B: Methodological*, vol. 35, no. 7, 2001, pp. 667–676.

[31] Parry, K. and M. L. Hazelton. Estimation of origin–destination matrices from link counts and sporadic routing data. *Transportation Research Part B: Methodological*, vol. 46, no. 1, 2012, pp. 175–188.

[32] Shao, H., W. H. Lam, A. Sumalee, A. Chen, and M. L. Hazelton. Estimation of mean and covariance of peak hour origin–destination demands from day-to-day traffic counts. *Transportation Research Part B: Methodological*, vol. 68, 2014, pp. 52–75.

[33] Shao, H., W. H. Lam, A. Sumalee, and M. L. Hazelton. Estimation of mean and covariance of stochastic multi-class OD demands from classified traffic counts. *Transportation Research Part C: Emerging Technologies*, vol. 59, 2015, pp. 92–110.

[34] Watling, D. A second order stochastic network equilibrium model, I: Theoretical foundation. *Transportation Science*, vol. 36, no. 2, 2002, pp. 149–166.

[35] Watling, D. A second order stochastic network equilibrium model, II: Solution method and numerical experiments. *Transportation science*, vol. 36, no. 2, 2002, pp. 167–183.

[36] Nakayama, S. and J. ichi Takayama. STOCHASTIC NETWORK EQUILIBRIUM MODELS CONSIDERING BOTH STOCHASTIC TRAVEL DEMAND AND ROUTE CHOICE. *Doboku Gakkai Ronbunshuu D*, vol. 62, no. 4, 2006, pp. 537–547. doi:10.2208/jscejd.62.537.

[37] Nakayama, S. and J.-i. Takayama. Traffic network equilibrium model for uncertain demands. In *Proceedings of the 82nd Transportation Research Board Annual Meeting*. 2003.

[38] Di, X., H. X. Liu, and X. J. Ban. Second best toll pricing within the framework of bounded rationality. *Transportation Research Part B: Methodological*, vol. 83, 2016, pp. 74–90.

[39] Davis, G. A. and N. L. Nihan. Large population approximations of a general stochastic traffic assignment model. *Operations Research*, vol. 41, no. 1, 1993, pp. 169–178.

[40] Haas, C. N. On modeling correlated random variables in risk assessment. *Risk Analysis*, vol. 19, no. 6, 1999, pp. 1205–1214.

[41] Waller, S., J. Schofer, and A. Ziliaskopoulos. Evaluation with traffic assignment under demand uncertainty. *Transportation Research Record: Journal of the Transportation Research Board*, , no. 1771, 2001, pp. 69–74.

[42] Ma, W. and Z. Qian. On the Variance of Recurrent Traffic Flow for Statistical Traffic Assignment, forthcoming. Working paper.

[43] Goldstein, H. Multilevel mixed linear model analysis using iterative generalized least squares. *Biometrika*, vol. 73, no. 1, 1986, pp. 43–56.

[44] Menon, A. K., C. Cai, W. Wang, T. Wen, and F. Chen. Fine-grained OD estimation with automated zoning and sparsity regularisation. *Transportation Research Part B: Methodological*, vol. 80, 2015, pp. 150–172.

[45] Jennings, A. and J. J. McKeown. *Matrix computation*. Wiley New York, 1992.

[46] Castillo, E., A. J. Conejo, J. M. Menéndez, and P. Jiménez. The observability problem in traffic network models. *Computer-Aided Civil and Infrastructure Engineering*, vol. 23, no. 3, 2008, pp. 208–222.

[47] Yang, Y. and Y. Fan. Data dependent input control for origin–destination demand estimation using observability analysis. *Transportation Research Part B: Methodological*, vol. 78, 2015, pp. 385–403.

[48] Hazelton, M. L. Some comments on origin–destination matrix estimation. *Transportation Research Part A: Policy and Practice*, vol. 37, no. 10, 2003, pp. 811–822.

[49] Hazelton, M. L. et al. Network tomography for integer-valued traffic. *The Annals of Applied Statistics*, vol. 9, no. 1, 2015, pp. 474–506.

[50] Antoniou, C., J. Barceló, M. Breen, M. Bullejos, J. Casas, E. Cipriani, B. Ciuffo, T. Djukic, S. Hoogendoorn, V. Marzano, et al. Towards a generic benchmarking platform for origin–destination flows estimation/updating algorithms: Design, demonstration and validation. *Transportation Research Part C: Emerging Technologies*, 2015.

THRESHOLD-BASED STOCHASTIC USER EQUILIBRIUM MODELS

David P. WATLING [a], Thomas K. RASMUSSEN [b], Otto A. NIELSEN [b], Carlo G. PRATO [c]

[a] Institute for Transport Studies, University of Leeds
[b] Department of Management Engineering, Technical University of Denmark
[c] School of Civil Engineering, The University of Queensland

1. INTRODUCTION AND MOTIVATION

For many decades, the two dominant approaches to represent transport network equilibrium have been the Deterministic User Equilibrium (DUE; Wardrop, 1952) and Stochastic User Equilibrium (SUE; Daganzo & Sheffi, 1977) models. The choice between these two fundamentally leaves us with a choice between two extreme cases, namely that only routes with minimum cost are used in DUE, and that (in its typical implementation) all permitted routes are used in SUE regardless of their costs. The deficiency of the DUE seems unreasonable given the imperfections in drivers' knowledge and the natural variations in real-life systems. The deficiency of the SUE seems also unreasonable, given that solution algorithms for SUE models only identify a subset of the available routes without providing a behavioural mechanism depending on anything fundamental in the model (see, e.g., Bekhor et al., 2008).

We aim to propose a model that - like DUE - addresses equilibration simultaneously with the determination of potential choice sets, but neither excludes all sub-optimal routes (as in DUE) nor includes all available alternatives (as in SUE) and allows flow allocation according to random utility theory (RUT) among used paths. With this aim in mind, we build on our own recent study (Watling et al., 2015), where we compared the relative merits of (a) an SUE model based on bounded random error terms with (b) the Restricted Stochastic User Equilibrium (RSUE) family of models. The RSUE family can be applied with logit and probit models, but admits multiple solutions with different *equilibrated route sets*. We note that the concept of Boundedly Rational User Equilibrium (BRUE) has been proposed, being a space of flow solutions representing drivers' inertia to route-switching where drivers are indifferent to route cost differences within indifference bands (see, e.g., Mahmassani and Chang, 1987; Di and Liu, 2016). However, BRUE models do not generate point estimates of equilibrium, but rather *interval estimates*.

Considering all of these perspectives, we aim at developing approaches that draw on elements of all of the methods above and making a theoretical contribution in terms of formulation of the equilibrium conditions and investigation of their existence/uniqueness. In particular:

- We aim to find a *point estimate* of equilibrium, akin to DUE and SUE, and to establish conditions under which such equilibria can be guaranteed to exist and be unique.
- Like DUE and RSUE, we aim to specify equilibrium conditions that distinguish potentially used from definitely unused routes, in the *equilibrated set* of active routes; that is, to develop an equilibrium prediction of the used/unused routes that is consistent with the travel costs.
- Like SUE, we aim to draw on RUT to represent travellers' misperceptions in travel costs.
- Like BRUE, we incorporate the notion of a threshold (an externally specified "bound").
- A first model specification gives rise to a form of SUE-like model with bounded error terms, but importantly with bounds linked to the threshold through RUT.
- A second model specification gives rise to a stricter version of RSUE, with distinct thresholds on used/unused routes incorporated into the equilibrium conditions.

2. THEORETICAL APPROACH TO CHOICE MODELLING WITH THRESHOLDS

We consider how the notion of a threshold might be incorporated within a Random Utility Model (RUM) setting and we present two alternative candidate formulations and discuss their relative merits. Our focus is on closed-form models, and in particular MNL models and variants where path overlap is accounted for within the systematic utility (see Prato, 2009).

2.1. THRESHOLD CHOICE SET MODEL

This first formulation considers the differences in *systematic utility* for any threshold operation. The particular features we aim to model are:

- a threshold is applied to the difference in *systematic utilities*;
- an option will only be in the choice set of considered alternatives if it is no more than a threshold amount worse in systematic utility than the best alternative available;
- individuals choose based on a RUM among options in the choice set.

This model is inspired by an extension of the concepts presented in Watling et al. (2015, §4), and in general, in any given instance, several choice probability vectors may satisfy the conditions. Let $\mathcal{P}$ denote the set of feasible choice probability vectors for alternatives *i*:

$$\mathcal{P} = \{\mathbf{p} : \mathbf{p} \in \mathbb{R}^{\boldsymbol{n}}, 0 \leq p_i \leq 1 \ (i = 1,2, \dots, n), \textstyle\sum_{i=1}^{n} p_i = 1\}\} \quad (1)$$

Noting that zero probabilities are permitted, any given vector in $\mathcal{P}$ will divide alternatives into chosen ($p_i > 0$) and non-chosen ($p_i = 0$) ones. The index-set of chosen alternatives corresponding to any vector $\mathbf{p}$ is denoted $\tilde{\mathcal{A}}(\mathbf{p}) \subseteq \mathcal{A}$ where $\mathcal{A} = \{1,2, \dots, n\}$ is the universal index set of available alternatives.

We now introduce a function g that includes a threshold parameter $\delta > 0$, which allows us to uniquely define a single *reference utility* for any given set of chosen alternatives:

$$V^* = g(\{V_i : i \in \tilde{\mathcal{A}}(\mathbf{p})\}; \delta) \quad (2)$$

where V_i is the systematic utility of alternative *i* (for example, $V^* = \max\{V_i : i \in \tilde{\mathcal{A}}(\mathbf{p})\} - \delta$).

The behavioural conditions we then aim to satisfy for the set $\tilde{\mathcal{A}}(\mathbf{p})$ of chosen alternatives are:

- Any chosen alternative must have a *systematic utility* no lower than V^* for the set $\tilde{\mathcal{A}}(\mathbf{p})$.
- Any non-chosen alternative must have a *systematic utility* less than V^* for the set $\tilde{\mathcal{A}}(\mathbf{p})$.
- The probabilities for the chosen alternatives satisfy a RUM over the set $\tilde{\mathcal{A}}(\mathbf{p})$.

The conditions are thus that any $\mathbf{p} \in \mathcal{P} \subseteq \mathbb{R}^{\boldsymbol{n}}$, with implied set of chosen alternatives indexed by $\tilde{\mathcal{A}}(\mathbf{p}) \subseteq \mathcal{A}$ and reference utility V^* satisfies the required behavioural rules if:

i) $V_i \geq V^* \ \ (i \in \tilde{\mathcal{A}}(\mathbf{p}))$;

ii) $V_i < V^* \ \ (i \in \mathcal{A} \cap \tilde{\mathcal{A}}(\mathbf{p})^{\mathrm{C}})$;

iii) $p_i = \phi_i(\mathbf{V}; \tilde{\mathcal{A}}(\mathbf{p})) \ \ (i \in \tilde{\mathcal{A}}(\mathbf{p}))$

where for any set $\mathcal{X}$ we denote the complement of that set by $\mathcal{X}^{\mathrm{C}}$, and where $\phi_i(\mathbf{V}; \tilde{\mathcal{A}}(\mathbf{p}))$ denotes a RUM on the set $\tilde{\mathcal{A}}(\mathbf{p})$, with relevant systematic utilities picked out of the *n*-vector $\mathbf{V}$.

2.2. THRESHOLD TRUNCATION MODEL

This second formulation considers any threshold operation being based on *random utilities*. Considering an imaginary reference alternative *A**, the particular features we aim to model are:

- a threshold is applied to the difference in *random utility* between a given alternative and *A**;
- an option will only be in the choice set of considered alternatives if it is no more than a threshold amount worse in random utility than the reference alternative *A**;
- for the subset of considered alternatives, indexed by $\mathcal{J} \subseteq \{1,2, \dots, n\}$, the choice probabilities are given by the odds associated with the binary choice probabilities of A_j versus *A** for $j \in \mathcal{J}$.

Suppose a specification of random utilities for the real and imaginary alternatives of $U_j = \theta V_j + \varepsilon_j \ \ (j = 1,2 \dots, n), U^* = \theta V^* + \varepsilon^*$, where $(\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n, \varepsilon^*)$ are i.i.d. Gumbel. Then, combining RUT with a threshold condition and selecting $V^* = \max(V_1, V_2, \dots, V_n)$ allows us to derive the following choice probability model for alternative A_j from $\{A_1, A_2, \dots A_n\}$ given systematic utilities $\mathbf{V} = (V_1, V_2, \dots, V_n)$ and threshold $\delta > 0$:

$$p_j(\mathbf{V}; \theta, \delta) = \frac{\left(\exp\left(\theta\left(V_j - \max(V_1, V_2, \dots, V_n) + \delta\right)\right) - 1\right)_+}{\sum_{k=1}^{n}\left(\exp\left(\theta\left(V_k - \max(V_1, V_2, \dots, V_n) + \delta\right)\right) - 1\right)_+} \quad (3)$$

where the + operator above is such that $(x)_+ = \max(x, 0)$. It is seen that, in the limit as $\delta \to \infty$, this model approaches the MNL model, as would be expected.

3. NETWORK EQUILIBRIUM PROBLEM

We now consider how the threshold-based models presented above may be implemented in a network equilibrium context where the network is a directed graph consisting of links *a* (*a*=1, 2, …, *A*). Define d_m as the demand for OD-pair *m* (*m*=1, 2, …, *M*), R_m as the index set of all simple acyclic paths for OD-pair *m*, N_m as the number of paths in R_m, and *R* as the union of the sets R_m so

that R = {1,2,..,N}, where $N = \sum_{m=1}^{M} N_m$. Denote with x_{mr} the flow on path $r \in R_m$, $\mathbf{x}$ the N-dimensional flow-vector on the universal choice set across all M OD-pairs, f_a the flow on link a, and $\mathbf{f} = (f_1, f_2, \dots, f_a, \dots, f_A)$ the A-dimensional link-flow vector. The convex set G of demand-feasible non-negative path flows is given by:

$$G = \left\{\mathbf{x}\epsilon\mathbb{R}_+^N : \sum_{r=1}^{N_m} x_{mr} = d_m\right\} \tag{4}$$

where $\mathbb{R}_+^N$ denotes the N-dimensional, non-negative Euclidean space. Define δ_{amr} equal to 1 if link a is part of path r for OD-pair m and 0 otherwise, then the convex set F of demand-feasible link flows is given by:

$$F = \left\{\mathbf{f}\epsilon\mathbb{R}_+^A : f_a = \sum_{m=1}^{M}\sum_{r=1}^{N_m} \delta_{amr} \cdot x_{mr}\,,\ \mathbf{x}\epsilon G\right\} \tag{5}$$

Given the column vectors $\mathbf{x}$ and $\mathbf{f}$, define $\mathbf{\Delta}$ as the A×N-dimensional link-path incidence matrix so that the link-path flow relation may be written as $\mathbf{f} = \mathbf{\Delta x}$. We suppose that the travel cost on path r for OD-pair m is additive in the travel costs of the utilised links:

$$c_{mr}(\mathbf{x}) = \sum_{a=1}^{A} \delta_{amr} \cdot t_a(\mathbf{\Delta x})\,,\ r\epsilon R_m \ and\ \mathbf{x}\epsilon G \tag{6}$$

Define $\mathbf{t}(\mathbf{f})$ ($\mathbf{t}$: $\mathbb{R}_+^A \to \mathbb{R}_+^A$) as the vector of generalised link travel cost functions, and $\mathbf{c}(\mathbf{x})$ ($\mathbf{x}$: $\mathbb{R}_+^N \to \mathbb{R}_+^N$) as the vector of generalised route travel cost functions, then the link-path cost relation may be written as $\mathbf{c}(\mathbf{x}) = \mathbf{\Delta}^{\mathrm{T}}\mathbf{t}(\mathbf{\Delta x})$.

For SUE-style models, we posit random utilities U_{mr} for each route:

$$U_{mr} = -\theta \cdot c_{mr}(\mathbf{x}) + \xi_{mr},\ r\epsilon R_m \tag{7}$$

where ξ_{mr} are continuous random variables following some given joint probability distribution, and $\theta > 0$ is a given parameter. We then define the following functions as the probability relations:

$$P_{mr}\big(\mathbf{c}(\mathbf{x})\big) = Pr(-\theta \cdot c_{mr}(\mathbf{x}) + \xi_{mr} \geq -\theta \cdot c_{ms}(\mathbf{x}) + \xi_{ms}, \forall s\epsilon R_m),\ r\epsilon R_m \tag{8}$$

These relations express the probability that path r between OD-pair m will have a perceived utility greater than or equal to the utilities of all alternative paths in the universal set of routes for that OD-pair. For any non-empty subset $\tilde{R}_m$ of R_m we also define:

$$P_{mr}\big(\mathbf{c}(\mathbf{x})\big|\tilde{R}_m\big) = Pr\big(-\theta \cdot c_{mr}(\mathbf{x}) + \xi_{mr} \geq -\theta \cdot c_{ms}(\mathbf{x}) + \xi_{ms}, \forall s\epsilon\tilde{R}_m\big),\ r\epsilon\tilde{R}_m \subseteq R_m \tag{9}$$

That is to say, whenever such a subset is not specified, we suppose P_{mr} refers to the universal set.

In this equilibrium model formulation, we use the Threshold Truncation Model defined in section 2.2 within a SUE framework where the systematic utility is equal to the negative of the travel cost. In order to accommodate the possibility of either assuming an absolute or relative threshold, we consider two cases:

Relative Cost Model: Variable thresholds $\boldsymbol{\delta}(\mathbf{c}) = (\delta_1(\mathbf{c}), \delta_2(\mathbf{c}), \dots, \delta_M(\mathbf{c}))$ where $\delta_m(\mathbf{c}) = (\tau - 1)\min(c_{ms}: s \in R_m)$ and $\tau > 1$.
Absolute Cost Model: Constant thresholds $\boldsymbol{\delta}(\mathbf{c}) \equiv \boldsymbol{\delta} = (\delta_1, \delta_2, \dots, \delta_M)$ where $\delta_m > 0$.
This yields the following definition:

Definition 1: Truncated Stochastic User Equilibrium (TSUE)
The route flow $\mathbf{x} \in G$ *is a TSUE if and only if it is an SUE:*

$$x_{mr} = d_m P_{mr}\big(\mathbf{c}(\mathbf{x})\big)\,,\ r \in R_m \tag{10}$$

and the choice model is given by the truncated logit form based on thresholds $\boldsymbol{\delta}(\mathbf{c})$ *:*

$$P_{mr}(\mathbf{c}) = \frac{\left(\exp\left(-\theta\left(c_{mr} - \min(c_{ms}: s \in R_m) - \delta_m(\mathbf{c})\right)\right) - 1\right)_+}{\sum_{t \in R_m}\left(\exp\left(-\theta\left(c_{mt} - \min(c_{ms}: s \in R_m) - \delta_m(\mathbf{c})\right)\right) - 1\right)_+}\,,\ r \in R_m \tag{11}$$

We have proven existence and uniqueness of the TMNL SUE (not included in the abstract). Then, we want to formulate a model with distinct thresholds on used/unused routes incorporated into the equilibrium thresholds (a threshold choice set model allowing general reference distributions and general choice models including the TMNL and the MNL). In particular, we extend the definition of the Restricted Stochastic User Conditions and the RSUE model (see Watling et al., 2015) by incorporating two reference functions, Φ and Ω. The former function Φ plays the same role as in RSUE, namely giving a *lower* bound on the actual cost of *unused* routes.

The latter function Ω, on the other hand, defines a second type of reference cost, namely one which gives an *upper* bound on the actual cost of *used* routes.

Definition 2: Restricted Stochastic User Equilibrium with Threshold (RSUET(Φ ,Ω))
The route flow $\mathbf{x} \in G$ *is a RSUET*(Φ ,Ω) *if and only if for all* $r \in R_m$*:*

$$x_{mr} > 0 \;\Rightarrow r \in \tilde{R}_m \wedge x_{mr} = d_m \cdot P_{mr}\left(\mathbf{c}(\mathbf{x})\middle|\tilde{R}_m\right) \wedge c_{mr}(\mathbf{x}) \leq \Omega\left(\{c_{ms}(\mathbf{x}): s \in \tilde{R}_m\}; \boldsymbol{\varsigma}_m\right) \quad (12)$$

$$x_{mr} = 0 \;\Rightarrow r \notin \tilde{R}_m \wedge c_{mr}(\mathbf{x}) \geq \Phi\left(\{c_{ms}(\mathbf{x}): s \in \tilde{R}_m\}; \boldsymbol{\xi}_m\right) \quad (13)$$

where $P_{mr}\left(\mathbf{c}(\mathbf{x})\middle|\tilde{R}_m\right)$ is given by equation (9).

Focusing on the properties of a particular 'absolute threshold' formulation RSUET(min+τ, min+τ) where $\Phi\left(\{c_{ms}(\mathbf{x}): s \in \tilde{R}_m\}; \boldsymbol{\xi}_m\right) = \Omega\left(\{c_{ms}(\mathbf{x}): s \in \tilde{R}_m\}; \boldsymbol{\varsigma}_m\right) = \tau + \min\left(c_{ms}: s \in \tilde{R}_m\right)$, we have proven (i) equivalence of TMNL(τ) SUE and TMNL(τ) RSUET(min+τ, min+τ) and (ii) that the limit case of TMNL(τ) SUE when τ approaches infinity is a MNL RSUET(min+τ, min+τ) and a MNL SUE (not included in the abstract).

4. NUMERICAL EXPERIMENTS

We present the results of applying the TMNL(τ) SUE and MNL RSUET(min+τ, min+τ) to two cases, namely a parallel-route network with a single OD-relation and the Sioux Falls network.

4.1. PARALLEL ROUTE NETWORK

Consider a network with three independent routes linking a single OD-pair with a demand of 200 and link cost-flow relation $c_i(\mathbf{x}) = c_i(x_i) = t_{0i} \cdot (1 + 0.3 \cdot (x_i/100)^4)$ for *i=1,2,3.*
In the following we evaluate the effect of varying the free-flow travel cost t_{01} on the solution of the TMNL(τ) SUE and the MNL RSUET(min+τ, min+τ). Assume t_{02}=18, t_{03}=20, θ=0.2 and τ=4. Figure 1 illustrates the equilibrated flows on the three routes as a function of t_{01}. A unique TMNL(4) SUE solution exists on the whole range of t_{01}. Paths 2 and 3 are 'activated' (allocated flow) when $c_1(\mathbf{x}) + \tau$ increases to t_{02} and t_{03}, respectively. For t_{01} > 19.75 path 1 is the costliest (allocated least flow), and for t_{01} > 28.6 then $c_1(\mathbf{x}) > \min(c_2(\mathbf{x}), c_3(\mathbf{x})) + \tau$, and path 1 is thus not used. Furthermore, figure 1 highlights a weakness of the MNL RSUET(min+4, min+4), namely that there are some ranges of t_{01} for which no equilibrium solution exists.

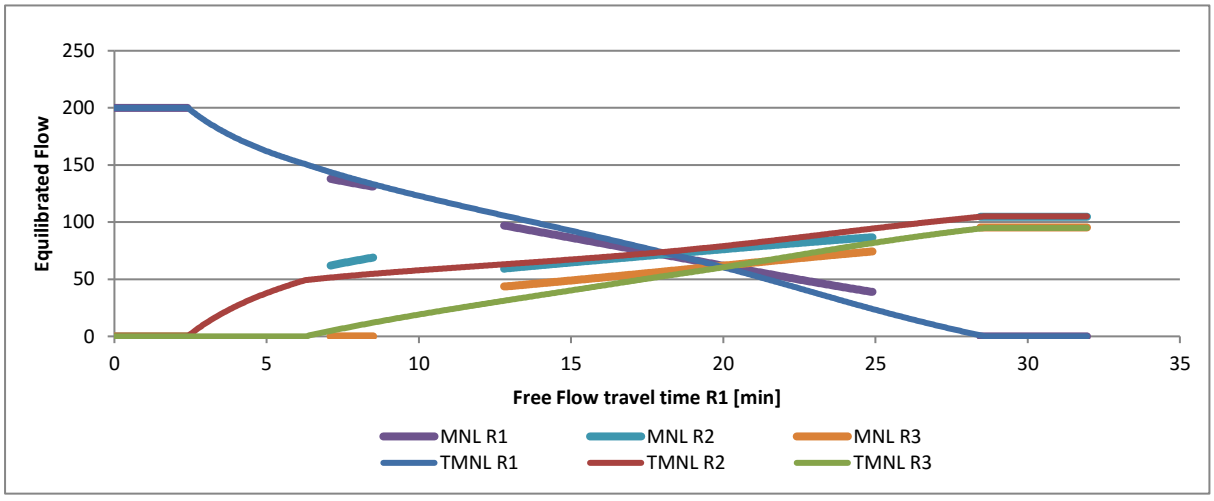


Figure 1. Equilibrated flow share across three paths as function of free-flow travel time on path 1 (t_{01}), MNL RSUET(min+4, min+4) and TMNL(4) SUE

4.2. SIOUX FALLS NETWORK

The Sioux Falls network contains 76 links and 528 OD-pairs. In the following, assume θ=0.2 and τ=15. We demonstrate the applicability of the TMNL(τ) SUE and MNL RSUET(min+τ, min+τ), and we have verified convergence of the solutions.

On an aggregate level, the results indicate similarity of the solutions across the two model formulations; the maximum choice set size is 18 for both applications, the average choice set size is 4.27/4.52 for the MNL RSUET/TMNL SUE, and the average and maximum link flow difference is 0.01% and 0.17% respectively. On a disaggregate level, a comparison of choice set composition

and link flow allocation for a single OD-relation also indicate similar and reasonable results (Figure 2). No unexpected links are used, and links on the 'main direction' from the origin to the destination is allocated the largest share of the flow.

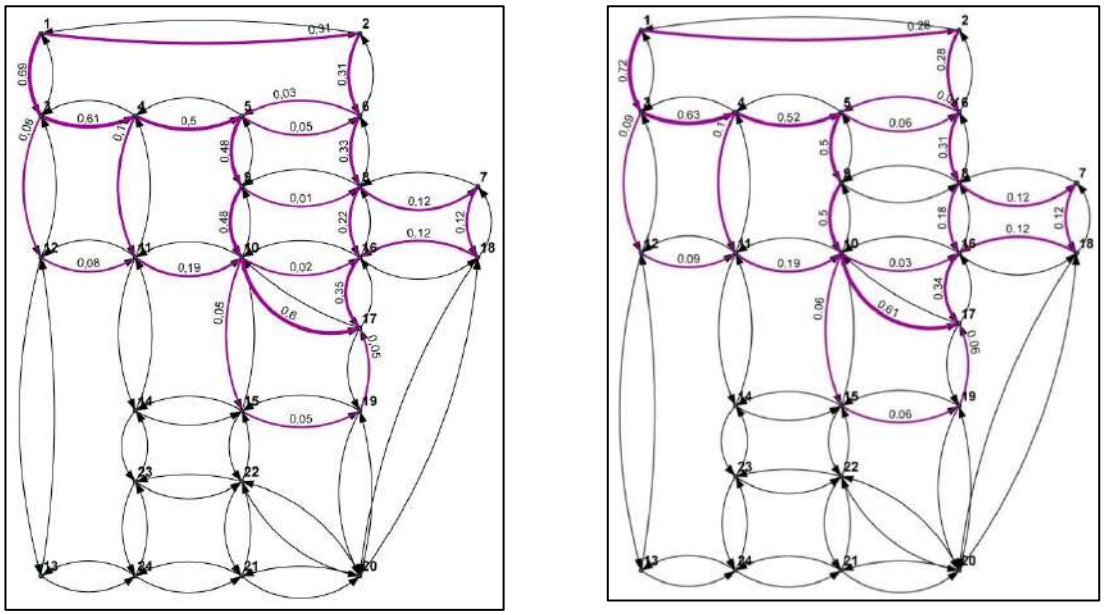


Figure 2 – Left: Equilibrated TMNL(15) SUE link flow shares for single OD-relation (Origin: 1, Destination: 17); Right: Equilibrated MNL RSUET(min+15,min+15) link flow shares for single OD-relation (Origin: 1, Destination: 17)

5. Conclusions and Further Research

We have formulated two novel models that address equilibration and choice set generation in a consistent manner. The introduction of cost thresholds allow the inclusion of some sub-optimal paths (unlike DUE) without requiring all available alternatives to be used (as in SUE), and flow is allocated according to RUT among used paths. We prove existence/uniqueness of solutions for one model, and highlight characteristics of the models by applications to two cases.

References

Bekhor S., Toledo T., Prashker, J., 2008. Implementation issues of route choice models in path-based algorithms. *Transportmetrica*, 4 (2), 117-133.

Daganzo, C.F., Sheffi, Y., 1977. On stochastic models of traffic assignment. *Transportation Science*, 11 (3), 351–372.

Di, X., Liu, H.X., 2016. Boundedly rational route choice behavior: A review of models and methodologies. *Transportation Research Part B: Methodological*, 85, 142–179.

Mahmassani, H.S., Chang, G.-L., 1987. On Boundedly Rational User Equilibrium in transportation systems. *Transportation Science*, 21 (2), 89–99.

Prato, C.G., 2009. Route choice modeling: past, present and future research directions. *Journal of Choice Modelling*, 2, 65-100.

Wardrop, J., 1952. Some theoretical aspects of road traffic research. *Proceedings of the Institution of Civil Engineers*, 1, 325–378.

Watling, D.P., Rasmussen, T.K., Prato, C.G., Nielsen, O.A., 2015. Stochastic User Equilibrium with equilibrated choice sets: Part I – Model formulations under alternative distributions and restrictions. *Transportation Research Part B: Methodological*, 77, 146–165.

Traffic & Mobility

TB2: Pricing and Competition

Thursday 1:00 – 2:30 PM

Session Chair: Song Gao

1:00 Competitive Rebalancing in One-Way Car-Sharing
Szymon Albinski, Stefan Minner*
Technical University of Munich

1:30 The Costs and Benefits of Ridesharing: Sequential Individual Rationality and Sequential Fairness
Ragavendran Gopalakrishnan, Koyel Mukherjee, Theja Tulabandhula*
Xerox Research Centre India

2:00 Cooperative Scheme - An Alternative Approach to Equitable and Pareto-Improving System Optimum
Sayeeda Ayaz, Song Gao, Hyoshin Park*
University of Massachusetts Amherst

Competitive Rebalancing in One-Way Car-Sharing

Szymon Albinski* and Stefan Minner

Technical University of Munich, School of Management, Arcisstraße 21, 80333 Munich, Germany

Key words: car-sharing; rebalancing; competition; Nash equilibrium; substitution

1 Introduction

Car-sharing services have experienced a rapid growth in recent years. In an outlook towards the future, the Boston Consulting Group predicts about 35 million car-sharing users worldwide by 2021, generating global revenues of 4.7 billion Euros. The most significant advantage of one-way car-sharing over the two-way system is the higher flexibility, as customers do not have to return the vehicles to the original pick-up point. For providers like Zipcar, DriveNow or car2go, this flexibility often leads to an uneven availability of the cars and thus coverage of the demand, which results in lower customer satisfaction and less usage of the vehicles. Consequently, car-sharing operators have to relocate their cars regularly in order to keep the system profitable. For that purpose, the providers partition their business area into several zones and relocate cars from zones with an expected overage to zones with an expected underage. This rebalancing is typically done at night when most of the cars are sitting idle.

Many authors have addressed this problem of determining optimal relocation strategies, considering different user-related factors like adoption rates, booking behaviour or time and type of day [4]. The problem of optimally rebalancing car-sharing vehicles has been addressed in the literature among others by simulation studies [5, 7], mixed-integer linear programming [3], stochastic programming [9] or rule based methods [6, 12]. All publications up to now have the one assumption in common that the car-sharing provider is operating in a monopoly. This is in contrast with the fact that there are often two or more providers offering their services in the same city. DriveNow and car2go, for example, competed in eleven cities in 2016. Car-sharing providers can be seen as retailers who offer a product at several disjunct locations in a city. This analogy motivates us to consider findings from retail management that show that competition can have a direct influence on the customers' demand as well as the companies' inventories and revenues. In literature, one distinguishes between stock-out based and assortment based substitution [8]. In the former case, customers substitute only when the preferred retailer is out of stock while in the latter case, customers make their choice among the available assortment. The problem of inventory competition under stock-out based substitution was introduced by Parlar. He proves the

*szymon.albinski@tum.de

existence of a unique Nash equilibrium for the case of two players [10]. Van Ryzin and Mahajan propose a Multinomial Logit approach to model assortment based substitution. They assume that each customer associates a certain utility with buying the product at a retailer's store and that this utility can be directly influenced by the offered assortment [11]. However, in reality a car-sharing operator may serve both types of customer groups and thus encounter stock-out based and assortment based substitution simultaneously. Consequently, a hybrid model that combines both types of substitution may be suitable.
The new concept of Mobility as a Service (MaaS) will intensify the competition in the car-sharing sector, as in MaaS different mobility options like public transport, car-sharing or bike-sharing as well as different operators can be used with one access card. This motivates an important research question: How does competition influence the availability of the cars and as as consequence the optimal rebalancing plan? The three main contributions of our work are as follows: (1) Formulation of the problem of rebalancing under competition as a Nash game considering stock-out based and assortment based substitution. (2) Insights into the optimal rebalancing policies and recommendations how car-sharing providers should react to the presence of a competitor. (3) Comparison and evaluation of the model using real-life data.

2 Model formulation

We study rebalancing under competition in a single period. For that purpose, we first introduce the monopolistic rebalancing problem (MRP) in a single period by adapting Allen's model [1]: A car-sharing company offers cars at price s in n zones. The demands at the zones can be expressed as random variables D_i with the corresponding probability density functions f_i and its cumulative F_i. The expected revenues $\pi_i(y_i) = s_i\big(y_i - \int_0^{y_i} F_i(d)\mathrm{d}d\big)$ depend on the inventory levels y_i in each location. To provide the optimal service level at all zones, the company can rebalance the initial car distribution y_i^0 between the zones at cost c before the demands are realized. The additional cars received at a zone are denoted by x_i. Thus, we obtain the problem:

$$\begin{aligned} (MRP) \qquad & \max\ \Pi(y,x) & \\ s.t. \qquad & x_i + y_i^0 \geq y_i & i = 1,...,n \\ & \textstyle\sum_{i=1}^n y_i^0 \geq \sum_{i=1}^n y_i & \\ & x_i, y_i \geq 0 & i = 1,...,n \end{aligned}$$

The objective is to maximize the expected revenue minus the rebalancing cost summed over all zones $i = 1,...,n$: $\Pi(y,x) := \sum_{i=1}^n \big(\pi_i(y_i) - c\,x_i\big)$. The first constraint ensures that from each zone no more than the available car stock can be taken, the second that the total number of cars after rebalancing corresponds to its initial number. Having stated the (MRP), we derive the corresponding Karush-Kuhn-Tucker conditions that allow us to partition the zones into sending, dormant and receiving zones. Thus, y_i and x_i can be determined without explicitly solving the underlying optimization problem. Using the Lagrangian, we show that a unique solution exists when the revenue function is strictly concave.

Next, we extend the (MRP) to the competitive case. For that purpose, we consider two players l and k who offer their car-sharing services in the same n zones. Due to the presence of the competitor k, player l has to take the influence of k's inventory levels on his own revenue function $\pi_{il}(y_{il}, y_{ik})$ into account in each zone i, and vice versa. As a consequence, we have to adapt the players' revenue functions while all constraints remain valid. We denote the resulting problem as competitive rebalancing problem (CRP). To cope with different customer behaviours, we look at stock-out based as well as assortment based substitution. In the case of stock-out based substitution, customers are willing to switch at a given substitution rate $b_{il} \in [0,1]$. That is, b_{il} of k's customers will substitute in the same zone from provider l when k has no cars in that zone. Analogous to Parlar's model [10], each player maximizes his profit by serving the own demand plus the competitor's unsatisfied one. We obtain the following profit function, denoting by B_{il} the maximum excess demand l can serve: $\pi(y_{il}, y_{ik}) = s_{il}\Big(y_{il} - \int_0^{y_{il}} F_{il}(d_l)\mathrm{d}d_l + \int_0^{y_{il}} \int_{y_{ik}}^{B_{il}} b_{il}(d_k - y_{ik}) f_{il}(d_l) f_{ik}(d_k)\mathrm{d}d_k\mathrm{d}d_l + \int_0^{y_{il}} \int_{B_{il}}^{\infty} (y_{il} - d_l) f_{il}(d_l) f_{ik}(d_k)\mathrm{d}d_k\mathrm{d}d_l\Big)$. Since we consider only one product, a higher assortment corresponds to a higher availability of cars. That is, we can translate assortment based competition into availability based competition. Thus, we assume that each potential customer chooses between both players after having observed – for instance online – the proximity of cars to his position. To determine q_{im}, i.e. the share of demand that goes to company $m \in \{k; l\}$ in zone i, we apply a Discrete Choice model. Using Multinomial Logit [2], we assume that each customer relates a certain utility to the observed proximity of cars. That is, at each location the demand for car-sharing services is split among the two players according to $q_{im}(y_{im})$, which is determined by the inventory levels y_{im}. The corresponding revenue function reads as follows: $\pi_{im}(y_{im}) = s_{im}\left(y_{im} - \int_0^{y_{im}/q_{im}(y_{im})} q_{im}(y_{im}) F_i(d_i)\mathrm{d}d_i\right),\ m \in \{k; l\}$. Having defined the objective functions for both cases of competition, we prove in a lemma that, given the competitor's inventory level, there exists a unique solution to the (CRP). Furthermore, we show that, for differentiable and strictly monotone probability density functions, the lemma holds for stock-out based as well as availability based competition. As both players solve the (CRP) simultaneously, we model this as a non-cooperative newsvendor game and show that a unique Nash equilibrium exists for both cases of competition.

3 Computational results

To analyse the effects of competition on the rebalancing policy, we conduct a numerical study. For that purpose, we consider several scenarios with different amounts of cars per company, as well as different mean demands and standard deviations for stock-out based (SB) and availability based (AB) substitution. In each scenario, we vary the substitution rates and MNL parameters, rebalancing cost as well as initial distribution of the stock. To demonstrate the effects of competition, we also consider the case of no competition in each scenario. That is, both players ignore the competitor's presence. The results in Table 1

Table 1: Test results showing the profit increase.

Case	Player 1		Player 2		Profit increase SB		Profit increase AB	
	# cars	Mean	# cars	Mean	P1	P2	P1	P2
1	300	200	300	200	14.16%	14.16%	8.00%	8.00%
2	300	300	300	300	14.13%	14.13%	9.50%	9.50%
3	300	450	300	450	13.36%	13.36%	3.01%	3.01%
4	300	300	100	100	8.01%	21.82%	11.31%	19.66%
5	500	200	500	200	12.38%	12.38%	7.75%	7.75%
6	500	300	500	300	14.31%	14.31%	9.81%	9.81%

show that ignoring the competition comes at high costs, as the players can increase their profits on average by up to 14 percent when customers substitute stock-out based and up to 15 percent when they substitute availability based. Naturally, the players are reacting to the competitor's actions only as long as it is still profitable for them. This is, for example, the case when the spread between revenues and relocation cost is sufficiently high or when most of the cars are not initially concentrated in one zone. We then observe for stock-out based substitution that the providers mostly focus on fulfilling the competitors' unserved demand, thus sharing the market. We see the opposite result for availability based substitution: to gain a larger market share, each player increases the number of cars in a zone when his competitor does so, too. This also leads to a higher number of rebalancing operations. When the total expected demand exceeds the providers' number of cars, the companies serve the demand in the zones where the cars are initially located and thus significantly reduce their rebalancing operations.
We further present empirical evidence for these results from car-sharing data in Munich, where car2go and DriveNow compete. The DriveNow bookings in the second half of 2016 were 70 percent higher than the car2go bookings, while the average number of bookings per car were comparable for both operators. The fact that DriveNow operated a larger fleet in that period indicates that the customers made their choices based on availability. A first analysis of the providers' car positioning leads to the assumption that both of them ignore the presence of a competitor.

4 Managerial insights

In this work, we provide insights into optimal rebalancing policies for car-sharing providers who face competition. Our results show that under stock-out based substitution the providers tend to divide the market while under availability based substitution they will compete. This competition leads to a higher number of rebalancing operations. Furthermore, the profitability and the initial car distribution y_i^0 influence the players' willingness to react. y_i^0, so far considered as given, can be made endogenous by formulating a Markov game, where both players solve the (CRP) in each stage and thus influence the car distribution in the next period. Additionally, the providers' optimal fleet sizes can be determined by adapting a two-stage capacity game, where the optimal fleet size decisions in the first stage depend on the (CRP) solved in the second stage.

References

[1] S. C. Allen. "Redistribution of total stock over several user locations". In: *Naval Research Logistics Quarterly* 5.4 (1958), pp. 337–345.

[2] M. Ben-Akiva and M. Bierlaire. "Discrete Choice Methods and their Applications to Short Term Travel Decisions". In: *Handbook of Transportation Science*. Ed. by R. W. Hall. Boston, MA: Springer US, 1999, pp. 5–33. ISBN: 978-1-4615-5203-1.

[3] B. Boyacı, K. G. Zografos, and N. Geroliminis. "An optimization framework for the development of efficient one-way car-sharing systems". In: *European Journal of Operational Research* 240.3 (2015), pp. 718–733.

[4] G. Brandstätter, C. Gambella, M. Leitner, E. Malaguti, F. Masini, J. Puchinger, M. Ruthmair, and D. Vigo. "Overview of Optimization Problems in Electric Car-Sharing System Design and Management". In: *Dynamic Perspectives on Managerial Decision Making: Essays in Honor of Richard F. Hartl*. Ed. by H. Dawid, K. F. Doerner, G. Feichtinger, P. M. Kort, and A. Seidl. Cham: Springer International Publishing, 2016, pp. 441–471. ISBN: 978-3-319-39120-5.

[5] E. M. Cepolina and A. Farina. "A new shared vehicle system for urban areas". In: *Transportation Research Part C: Emerging Technologies* 21.1 (2012), pp. 230–243.

[6] D. Jorge, G. H. Correia, and C. Barnhart. "Comparing optimal relocation operations with simulated relocation policies in one-way carsharing systems". In: *IEEE Transactions on Intelligent Transportation Systems* 15.4 (2014), pp. 1667–1675.

[7] A. G. Kek, R. L. Cheu, Q. Meng, and C. H. Fung. "A decision support system for vehicle relocation operations in carsharing systems". In: *Transportation Research Part E: Logistics and Transportation Review* 45.1 (2009), pp. 149–158.

[8] A. G. Kök, M. L. Fisher, and R. Vaidyanathan. "Assortment Planning: Review of Literature and Industry Practice". In: *Retail Supply Chain Management: Quantitative Models and Empirical Studies*. Ed. by N. Agrawal and S. A. Smith. Boston, MA: Springer US, 2015, pp. 175–236. ISBN: 978-1-4899-7562-1.

[9] R. Nair and E. Miller-Hooks. "Fleet management for vehicle sharing operations". In: *Transportation Science* 45.4 (2011), pp. 524–540.

[10] M. Parlar. "Game theoretic analysis of the substitutable product inventory problem with random demands". In: *Naval Research Logistics (NRL)* 35.3 (1988), pp. 397–409.

[11] G. v. Ryzin and S. Mahajan. "On the relationship between inventory costs and variety benefits in retail assortments". In: *Management Science* 45.11 (1999), pp. 1496–1509.

[12] S. Weikl and K. Bogenberger. "A practice-ready relocation model for free-floating carsharing systems with electric vehicles–Mesoscopic approach and field trial results". In: *Transportation Research Part C: Emerging Technologies* 57 (2015), pp. 206–223.

The Costs and Benefits of Ridesharing: Sequential Individual Rationality and Sequential Fairness

Ragavendran Gopalakrishnan, Koyel Mukherjee, Theja Tulabandhula
Xerox Research Centre India (XRCI), Bangalore, India.

1 Introduction

Ridesharing[1] has emerged as a popular solution to combat ever-increasing congestion along road networks around the world. The resulting decrease in the number of vehicles can reduce the carbon footprint significantly, making ridesharing a mechanism that is all the more desirable from a sustainability perspective. While these benefits are undoubtedly good for the society, many individual commuters are reluctant to embrace ridesharing, despite the cost savings that would result. Several factors such as reliability, privacy, security, and delays contribute to the inconveniences due to ridesharing. Therefore, for the average commuter, it may be easier to stick with their existing routine that they are comfortable with, when faced with an often challenging decision process to determine whether the cost savings is worth the inconvenience.

In order to increase the adoption of ridesharing, the routing, pricing, and cost sharing schemes must be persuasive to the passenger by addressing the trade-off between the additional delay and cost savings in a way that ultimately incentivizes them to participate in ridesharing. Traditionally, this notion is captured in the mechanism design literature by means of a concept called *individual rationality*, that is, every passenger is better off (according to some utility function) having participated in ridesharing than not. However, from a practical perspective, this concept falls short of ensuring that passengers are satisfied *during* the ride. A frequent source of frustration are the detours taken to pick up and/or drop off additional passengers, which inconvenience existing passengers. In order to address these pain points, we propose imposing a stronger condition called *sequential individual rationality*, which requires that existing passengers are progressively better off every time an additional passenger is picked up. Such a property would also ensure some robustness, e.g., in a dynamic/online setting, passengers would remain satisfied even if a future pickup is canceled.

Our central goal in this work is to develop a framework for cost sharing in ridesharing that explicitly models the "inconvenience cost" experienced by the passengers due to the detours, and explore the properties and consequences of imposing sequential individual rationality and fairness on the routing and cost sharing schemes.

1.1 Summary of our Contributions

First, we introduce our cost sharing framework, in which we model the disutility to an existing passenger at any stage of the ride as the sum of the passenger's monetary payment for the ride (as determined by the cost sharing scheme), and an inconvenience cost term (as a function of the detour due to ridesharing), assuming that there are no more passengers to be picked up from there on. Sequential individual rationality then requires this disutility to be non-increasing throughout the ride, for all passengers. In other words, every time a new passenger is picked up, the resulting additional detour must be worth the additional cost savings to the existing passengers.

Then, we provide an exact characterization for any route to be "SIR-feasible", that is, there exists some budget-balanced cost sharing scheme that is sequentially individually rational on that route. Such a characterization would be useful in designing "cost-aware" routing algorithms that suggest SIR-feasible

[1]The term "ridesharing" in popular culture has become a buzzword that refers to any *ride-booking* or *ride-hailing* service such as Uber and Lyft, even if there is only one passenger taking the ride and there is no sharing involved. Recently, the Associated Press has criticized this abuse of the term [8]. In this work, we use the term "ridesharing" to denote only those services that allow two or more passengers to share rides, such as UberPool and LyftLine, in addition to community carpooling.

routes (when grouping ridesharing requests and assigning them to vehicles). These SIR-feasibility constraints are necessarily complex, so, we consider a simplified scenario where all the passengers are travelling to a common destination. For this "single dropoff" scenario, we show that the SIR-feasibility constraints simplify to natural upper bounds on the incremental detours, that keep shrinking as the ride progresses towards the destination and as more passengers are picked up. We also show, in a series of theorems, that these bounds on incremental detours can be aggregated to establish upper and lower bounds (that are *sublinear* in the number of passengers in realistic scenarios) on the total detour endured by a passenger as a fraction of their direct distance to their destination.

Next, we observe that budget-balanced cost sharing schemes that are SIR can be alternately viewed as *benefit sharing schemes*, where the benefit being shared is the total decrease in the disutilities of all the passengers, every time a new passenger is picked up. This "duality" enables a natural definition of *sequential fairness* that requires a portion of the incremental benefit to be distributed among the existing passengers in proportion to the inconvenience costs they suffer due to picking up the new passenger. We then present an exact characterization of sequentially fair cost sharing schemes for the single dropoff scenario, which exposes several useful structural properties of such schemes, including a strong requirement that *passengers must compensate each other for the detour inconveniences that they cause*.

Finally, we explore some important algorithmic questions motivated by sequential individual rationality, most of which are left open for future work. In particular, it is unknown, even for the single dropoff scenario, whether there exists a polynomial time algorithm to check for the existence of SIR-feasible routes, when restricted to a metric space (we show that it is NP-hard otherwise). Even if so, we show that optimizing for total distance traveled over SIR-feasible routes is NP-hard (through a reduction from a variant of Metric-TSP). We then consider a variant of the vehicle routing problem where passengers are allocated to vehicles such that the total "vehicle-miles" traveled is minimized. While this problem is known to be NP-hard in general, we show that it can be solved in polynomial time given a fixed ordering on the pickup points. Analyzing the impact of imposing SIR-feasibility on the resulting vehicle-routes is left open. We conclude with more open directions—connections to the online mechanism design literature and extensions to multiple dropoff scenarios.

1.2 Related Work

The cost sharing problem for ridesharing has garnered relatively little attention in literature (compared to the ride matching and route optimization problems) – in most existing schemes, individual passengers are either asked to post what they are willing to pay in advance [5], share the total cost proportionately among themselves according to the distances travelled [2, 10], or negotiate their cost shares on their own during/after the ride. Such methods ignore the real-time costs and delays incurred during the ride (as in the first instance), are insensitive to the disproportionate delays encountered during the ride (as in the second instance), or lead to a complicated and often uncomfortable negotiation process between possible strangers (as in the third instance).

Recent work has studied cost sharing when passengers have significant autonomy in choosing their rides or forming their own ridesharing groups, e.g., cost sharing schemes based on the concept of kernel in cooperative game theory [4], second-price auction based solutions [14], and market based ridematching models with deficit control [19]. Fair cost sharing in ridesharing has also been studied in [13] under a mechanism design framework, where an individually rational VCG-based payment scheme is modified to recover budget-balance at the cost of incentive compatibility. Our work differs from all the above in that we do not make any assumptions about the mechanics of ride matching; our cost sharing model is independent of the routing framework (static or dynamic), and is applicable to community carpooling and commercial ridesharing providers alike. Our work is also different from the problem of pricing in ridesharing (see, e.g., [3]); our focus is on sharing the resulting cost among the passengers.

To the best of our knowledge, all previous works on ridesharing problems to have addressed individual rationality and/or detour limits have treated them as *independent* constraints, e.g., [13, 17, 18]. In contrast, in our model, requiring (a strong version of) individual rationality *induces* natural bounds on the detours experienced by the ridesharing passengers.

Variations of individual rationality involving temporal aspects are well studied in the economics literature, e.g., *ex-ante*, *interim*, and *ex-post* individual rationality in mechanism design [16], and sequential individual

rationality in bargaining and repeated games [7]. However, to the best of our knowledge, we are the first to explore its applicability to the ridesharing problem[2] and its consequences and fairness properties of the resulting outcomes.

There is an extensive literature on cooperative game theory and fair division [12, 15] that suggest various cost sharing schemes that can be analyzed in our framework. Our view of fairness relies on a different view of how the total incremental benefit due to ridesharing is allocated among the passengers during each stage of the ride (sequential fairness). While we believe the two approaches are not independent, exploring the connections is beyond the scope of this work.

Finally, there is a plethora of work when it comes to optimization problems in ridesharing [1, 9, 17]. While the constraints on detours induced by the sequential individual rationality constraints can augment any routing optimization problem to make it more challenging, in this paper, we focus on the problem of finding an optimal allocation of passengers to vehicles (that minimizes the total vehicle-miles traveled), which is a variant of the vehicle routing problem (VRP) [6].

2 Illustrative Example: A Single Dropoff Scenario

Due to space constraints, in the rest of this abstract, we limit ourselves to describing a simple single dropoff ridesharing scenario using which we illustrate the key concept of sequential individual rationality and its consequences for cost-aware routing. We refer the interested reader to the full version of our work [11].

Consider $n = 3$ passengers, picked up from their sources S_1, S_2, S_3 (in that order), and travelling to a common destination D. Let the set of passengers be $\mathcal{N} = \{1, 2, 3\}$. The progression of the route as the passengers are picked up one by one, is depicted in Fig. 1. Given the final route, the total distances traveled by passengers $1, 2$ and 3 are $d_1(\mathcal{N}) = S_1S_2 + S_2S_3 + S_3D, d_2(\mathcal{N}) = S_2S_3 + S_3D$ and $d_3(\mathcal{N}) = S_3D$. The total distance traveled by the vehicle is $d(\mathcal{N}) = S_1S_2 + S_2S_3 + S_3D$, which is the same as the distance traveled by the first passenger. The operational cost is thus $\mathcal{OC}(\mathcal{N}) = \alpha_{op}(S_1S_2 + S_2S_3 + S_3D)$, where $\alpha_{op} > 0$ is the operating cost per unit distance.

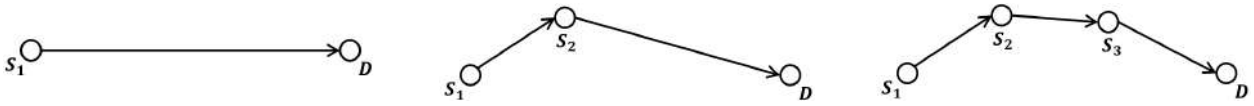


Figure 1: Route progress while picking up passengers traveling to a common destination.

Cost Sharing Scheme: A cost sharing scheme f is a function that specifies, for any subset $S \subseteq \mathcal{N}$, how $\mathcal{OC}(S)$ is distributed among the passengers in S. That is, $f(i, S)$ denotes the portion of $\mathcal{OC}(S)$ allocated to passenger $i \in S$. If f is a *budget-balanced* cost sharing scheme, we have $\sum_{i \in S} f(i, S) = \mathcal{OC}(S)$ for all $S \subseteq \mathcal{N}$.
Inconvenience Cost: In a ride involving a subset of passengers $S \subseteq \mathcal{N}$, for each $i \in S$, the inconvenience cost incurred due to other passengers is proportional to the amount of detour caused by the passengers in $S - \{i\}$ to i. The constant of proportionality, $\alpha_i \geq 0$, is the inconvenience cost of i per unit distance. For example,

$$\begin{aligned}
\mathcal{IC}_1(\{1, 2\}) &= \alpha_1(S_1S_2 + S_2D - S_1D), \\
\mathcal{IC}_2(\{1, 2, 3\}) &= \alpha_2(S_2S_3 + S_3D - S_2D), \\
\mathcal{IC}_3(\{1, 2, 3\}) &= \alpha_3(S_3D - S_3D) = 0.
\end{aligned}$$

Disutility: When a subset of passengers $S \subseteq \mathcal{N}$ share a ride, the *disutility* to a passenger $i \in S$ is defined as the sum of their monetary payment for the ride and their inconvenience cost due to any detours, that is,

$$\mathcal{DU}_i(S) = f(i, S) + \mathcal{IC}_i(S).$$

[2]By extension, we believe that any cost sharing or pricing framework involving online resource allocation where new jobs affect existing jobs should benefit from the concept of sequential individual rationality and fairness.

Individual Rationality: A cost sharing scheme f is *individually rational* (IR) if $\mathcal{DU}_i(\mathcal{N}) \leq \mathcal{DU}_i(\{i\})$, for all $i \in \mathcal{N}$. In our example, when f is budget-balanced, these IR conditions can be written as:

$$\begin{aligned} f(1,\mathcal{N}) + \alpha_1(S_1S_2 + S_2S_3 + S_3D - S_1D) &\leq \alpha_{op} S_1D, \\ f(2,\mathcal{N}) + \alpha_2(S_2S_3 + S_3D - S_2D) &\leq \alpha_{op} S_2D, \\ f(3,\mathcal{N}) &\leq \alpha_{op} S_3D. \end{aligned}$$

Sequential Individual Rationality: IR requires that passengers are better off ridesharing than not, but by taking into account their disutilities *only at the end of their ride*. In that sense, it is a static property. To encourage wider adoption of ridesharing, we need to address the pain points of ridesharing passengers *during* the ride as well. An important class of such pain points is when the vehicle undertakes detours to pick up and drop off other passengers. Thus, we propose a stronger property that requires that IR hold at every stage of the ride, that is, every time a new passenger is picked up. We call this property "sequential" IR (SIR). For example, SIR for passenger 1 would mean $\mathcal{DU}_1(\mathcal{N}) \leq \mathcal{DU}_1(\mathcal{N} \setminus \{3\}) \leq \mathcal{DU}_1(\mathcal{N} \setminus \{2,3\})$, that is, the disutilities are nonincreasing throughout the ride. Thus, the SIR constraints for our example are:

$$\begin{aligned} f(1,\mathcal{N}) + \alpha_1(S_1S_2 + S_2S_3 + S_3D - S_1D) \leq f(1,\mathcal{N} \setminus \{3\}) + \alpha_1(S_1S_2 + S_2D - S_1D) &\leq \alpha_{op} S_1D, \\ f(2,\mathcal{N}) + \alpha_2(S_2S_3 + S_3D - S_2D) \leq f(2,\mathcal{N} \setminus \{3\}) &\leq \alpha_{op} S_2D, \\ f(3,\mathcal{N}) &\leq \alpha_{op} S_3D. \end{aligned}$$

SIR-Feasibility: We say that a route is *SIR-feasible* if there exists a budget-balanced cost sharing scheme f that is SIR on that route. In our example, a set of necessary conditions (which can also be shown to be sufficient) for the final route to be SIR-feasible is therefore obtained by first summing up the above SIR constraints (at each stage), then using budget-balance of f, and simplifying:

$$S_2S_3 + S_3D - S_2D \leq \frac{\alpha_{op}}{\alpha_{op} + \alpha_1 + \alpha_2} S_3D \quad \text{and} \quad S_1S_2 + S_2D - S_1D \leq \frac{\alpha_{op}}{\alpha_{op} + \alpha_1} S_2D.$$

SIR-Feasible Routes and Cost-Aware Routing: Upon closer inspection, we note that the left hand sides of the above SIR-feasibility constraints are nothing but the *incremental* detours due to picking up subsequent passengers. Thus, these constraints can be viewed as imposing upper bounds on the permissible incremental detours at every stage of the ride. An interesting observation is that these bounds decrease with increasing number of passengers as well as increasing proximity to the destination, which means that as more and more passengers are picked up, the permissible incremental detour to pick up yet another passenger keeps shrinking, which is natural to expect. For the passengers in our example, Fig. 2 shows the evolution of the "SIR-feasible region" (points from which the next passenger can be picked up so that the resultant route is SIR-feasible) in Euclidean space, when $\alpha_i = \alpha_{op}$ for $i = 1, 2, 3$.

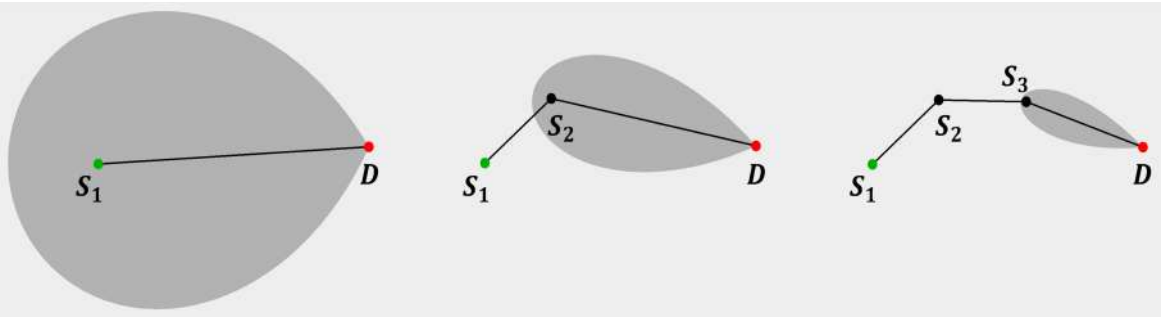


Figure 2: Evolution of the *SIR-feasible region* (darker shade) while picking up passengers that are traveling to a common destination. Note that the region diminishes rapidly with every subsequent pickup.

References

[1] N. Agatz, A. Erera, M. Savelsbergh, and X. Wang. Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, 223(2):295–303, 2012.

[2] N. A. Agatz, A. L. Erera, M. W. Savelsbergh, and X. Wang. Dynamic ride-sharing: A simulation study in metro atlanta. *Transportation Research Part B: Methodological*, 45(9):1450–1464, 2011.

[3] S. Banerjee, R. Johari, and C. Riquelme. Pricing in ride-sharing platforms: A queueing-theoretic approach. In *Proceedings of the Sixteenth ACM Conference on Economics and Computation*, pages 639–639, 2015.

[4] F. Bistaffa, A. Farinelli, G. Chalkiadakis, and S. Ramchurn. Recommending fair payments for large-scale social ridesharing. In *Proceedings of the 9th ACM Recommender Systems Conference (RecSys)*, pages 139–146, 2015.

[5] B. Cao, L. Alarabi, M. F. Mokbel, and A. Basalamah. SHAREK: A scalable dynamic ride sharing system. In *Proceedings of the 16th IEEE International Conference on Mobile Data Management (MDM)*, pages 4–13, 2015.

[6] J.-F. Cordeau, G. Laporte, M. W. Savelsbergh, and D. Vigo. Vehicle routing. In *Handbooks in operations research and management science: Transportation*, volume 14, chapter 6, pages 367–428. 2006.

[7] J. Esteban. *The social viability of money (competitive equilibria and the core of overlapping generations economies)*, volume 372. Springer-Verlag, 1991.

[8] B. Freed. Why You Shouldn't Call Uber and Lyft "Ride-Sharing", 2015.

[9] M. Furuhata, M. Dessouky, F. Ordóñez, M.-E. Brunet, X. Wang, and S. Koenig. Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological*, 57:28–46, 2013.

[10] R. Geisberger, D. Luxen, S. Neubauer, P. Sanders, and L. Volker. Fast detour computation for ride sharing. In *Proceedings of the 10th Workshop on Algorithmic Approaches for Transportation Modelling, Optimization, and Systems, OASIcs-OpenAccess Series in Informatics*, volume 14, pages 88–99. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2010.

[11] R. Gopalakrishnan, K. Mukherjee, and T. Tulabandhula. The costs and benefits of ridesharing: Sequential individual rationality and sequential fairness. *CoRR*, abs/1607.07306, 2016.

[12] K. Jain and M. Mahdian. Cost sharing. In N. Nisan, T. Roughgarden, É. Tardos, and V. V. Vazirani, editors, *Algorithmic Game Theory*, chapter 15, pages 385–410. Cambridge University Press, 2007.

[13] E. Kamar and E. Horvitz. Collaboration and shared plans in the open world: Studies of ridesharing. In *IJCAI*, pages 187–194, 2009.

[14] A. Kleiner, B. Nebel, and V. A. Ziparo. A mechanism for dynamic ride sharing based on parallel auctions. In *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI)*, pages 266–272, 2011.

[15] H. Moulin. *Fair division and collective welfare*. MIT press, 2004.

[16] Y. Narahari, R. Narayanam, D. Garg, and H. Prakash. Foundations of mechanism design. In *Game Theoretic Problems in Network Economics and Mechanism Design Solutions, Part of the Series on Advanced Information and Knowledge Processing*, pages 1–131. Springer, 2009.

[17] D. Pelzer, J. Xiao, D. Zehe, M. H. Lees, A. C. Knoll, and H. Aydt. A partition-based match making algorithm for dynamic ridesharing. *IEEE Transactions on Intelligent Transportation Systems*, 16(5):2587–2598, 2015.

[18] D. O. Santos and E. C. Xavier. Dynamic taxi and ridesharing: A framework and heuristics for the optimization problem. In *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, volume 13, pages 2885–2891, 2013.

[19] D. Zhao, D. Zhang, E. H. Gerding, Y. Sakurai, and M. Yokoo. Incentives in ridesharing with deficit control. In *Proceedings of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*, pages 1021–1028, 2014.

Cooperative Scheme - An Alternative Approach to Equitable and Pareto-Improving System Optimum

Sayeeda Bint Ayaz* Song Gao† Hyoshin Park‡

1 Introduction

User Equilibrium (UE) is based on the assumption that travelers behave selfishly in a non-cooperative manner to minimize their own travel cost. System Optimum (SO) is a traffic state where the total cost of the system is minimized. However, SO is not stable as drivers on slow routes will likely shift to the fast routes, and cause the system to revert back to UE. Therefore, drivers need to be penalized through charges or compensated through rewards for the system to move towards SO. The study of congestion pricing is traced back to the early twentieth century when Pigou (1920) recommended a tax to be levied on any market activity that generates negative externalities. Vickrey (1969) proposed a time-varying toll that could completely eliminate queuing delay, and thereby maximize system efficiency. A plethora of studies have been conducted in this area since then.

Despite its theoretical appeal, congestion pricing continues to be a hard sell to people. Major proposals have been remonstrated by public or political opposition. For example, cordon tolling schemes for Edinburgh and Manchester in the UK were rejected by public referenda (2005 and 2008). An online petition to the UK government (2007) attracted more than 1.8 million signatures against road pricing, and effectively put an end to plans for a national scheme in the UK for the time being. A cordon toll plan for New York City was stopped by the New York state legislature (2008) when it declined to vote on the proposal. These setbacks illustrate the difficulties of designing congestion pricing schemes that are both efficient and publicly acceptable (de Palma and Lindsey, 2011).

There are a wide range of factors for the setbacks, and equitability is one of the most cited. Congestion pricing sometimes is characterized as a "regressive tax" (Small, 2007) in that high income travelers who usually have a high value of time (VOT) could benefit at the cost of low income travelers' loss. Innovative solutions to the equitability problem have focused on the so-called Pareto-improving schemes, where no traveler is worse off compared to the no-toll case. Examples include a hybrid scheme between rationing and pricing (Daganzo, 1995), alternating charging a given fraction of the drivers (Daganzo and Garcia, 2000) credit-based scheme (e.g., Kockelman and Kalmanje, 2005), tradable credit scheme (e.g., Yang and Wang, 2011; Nie, 2015), and toll-and-subsidy scheme (e.g., Guo and Yang, 2010; Liu et al., 2009; Nie and Liu, 2010). Lawphongpanich and Yin (2010) solve the Pareto-optimum pricing problem in a general network.

*University of Massachusetts, Amherst. sbayaz@umass.edu
†University of Massachusetts, Amherst. sgao@umass.edu
‡University of Massachusetts, Amherst. hyoshinpark@umass.edu

2 Research Summary

In this study, an alternative approach is adopted to achieve equitable and Pareto-improving system optimum based on cooperation among travelers assisted by defector penalty. As reviewed in Klein and Ben-Elia (2016), cooperation has been studied extensively in behavioral economics and game theory to resolve social dilemmas such as Prison's Dilemma (Dawes, 1980). Evolutionary game theory provides a competent theoretical framework for addressing the subtleties of cooperation in such situations (e.g., Murnighan and Roth, 1983; Fehr and Gächter, 2002; Boyd et al., 2003; Wu and Wang, 2007). Helbing et al. (2005) conducted experiments on humans playing two-person route choice games in a computer laboratory to study decision behavior in repeated games. Results show that a taking-turn strategy that achieves SO emerges after the two players have enough experience to perceive the value of cooperation. However, computer simulations and additional experiments indicate that oscillatory cooperation in route choice games with four players emerge only after a long time period (rarely within 300 iterations).

Almost all traffic equilibrium studies make the assumption that travelers are non-cooperative, for a good reason. With the large number of travelers, the time it takes for cooperation to emerge is too long for the assumption to be practically valid. Penalty to defectors (people who do not cooperate) has been suggested (Helbing et al., 2005; Klein and Ben-Elia, 2016) to promote cooperation. This study operationalizes the idea by performing theoretical analyses in a general network with heterogenous VOT among travelers. In relation to the literature, a hybrid scheme between pricing and rationing in Daganzo (1995) for a single bottleneck with flexible demand bears the most resemblance to the current work. The major distinction is that the cooperative scheme in this study is applied to route choice, both deterministic and stochastic, in a general network.

In the preliminary analysis, it is shown that under mild conditions, a cooperative scheme always exists for a single-origin-destination (OD) two-route network to ensure equitable and Pareto-improving SO, regardless of VOT distribution, even when the Pareto-improving is in terms of travel time only, exclusive of potential rewards received by cooperators. When the VOT is bounded from above, a cooperative scheme without financial transactions always exists, in which case the defector penalty is high enough so that all travelers cooperate. This last feature is especially appealing in that the public tends to view congestion pricing as tax and this perception almost dooms any such pricing schemes even if they are Pareto-improving. A more practical and potentially more appealing case is also discussed where a certain number of defectors exist. Intended contributions of planned work include: 1) extension of the analysis to a general network where a traveler might travel between multiple OD's and the decision to participate in the cooperative scheme is person-specific while not OD-specific, and 2) extension of the analysis to probabilistic route choice so that stochastic UE (SUE) emerges without interventions.

It is the authors' belief that traveler cooperation will bring about transformative changes to how the transportation system is managed, and its implementation can be accelerated by technologies including connected and autonomous vehicles. Specifically, with fully autonomous vehicles, the barrier to participation in the cooperative scheme due to cognitive constraint (e.g., inertia against regular switching to potentially unfamiliar routes) and disruption to the execution due to human errors (e.g., failing to follow specified route) can be significantly reduced, and even eliminated. Future research will be focused on the behavioral and institutional aspects to provide a comprehensive assessment of the feasibility and efficacy of the cooperative scheme.

3 Preliminary Analysis in a Simple Network

Consider a single-OD network with fixed demand d, connected by two routes: Fast Route with a travel time t_F at SO and Slow Route with a travel time t_S at SO. The two routes have the same travel time at UE, t^{UE}, and $t_F < t^{UE} < t_S$. x_F and x_S are the SO flows (positive integers) on Fast Route and Slow Route respectively, and $x_F + x_S = d$. Travel time is a strictly increasing function of flow.

A cooperative scheme is defined as travelers taking turns to use Fast Route. Participants of the cooperative scheme are called "cooperators". They use Fast Route on some days and Slow Route on other days following the guidance of a central controller to maintain an SO flow pattern on each day, even though the composition of the flow varies from day to day due to turn taking.

When all travelers are cooperators, a naive turn-taking strategy is such that in each cycle of d days, each cooperator uses Fast Route for x_F days and Slow Route for x_S days. The average travel time for each cooperator over a cycle is the average SO travel time:

$$t^{SO} = \frac{x_F t_F + x_S t_S}{d} \tag{1}$$

It is evident that $t_F < t^{SO} < t^{UE} < t_S$. A shorter cycle is preferred as it demonstrates the value of turn-taking in shorter time and thus more appealing for getting public acceptance. With an improved strategy, cooperators take turns by blocks. Let g be the greatest common factor of x_F and x_S. Cooperators are grouped into d/g blocks. In each cycle of d/g days, each block of cooperator uses Fast Route for x_F/g days and Slow Route for x_S/g days. The resulting average travel time over a cycle is still t^{SO}. In practical applications, the cycle length might need to be controlled below a fairly small number, say, 5 working days, to have a realistic chance of acceptance. In such cases, indifference to small travel time differences (e.g., a 5-minute difference for a 1-hour trip) can be exploited such that a small number of approximately equal-sized blocks result in approximately equal average travel time for each cooperator with the differences under a certain threshold.

The cooperative scheme is not stable, as a "defector" who stays on Fast Route all the time has a lower travel time of t_F than a cooperator. As more travelers defect, the cooperative scheme is broken and the system reverts back to UE. One way to maintain the cooperative scheme is to impose a defector penalty τ to make defection more costly than or at least as costly as cooperation.

Homogeneous VOT

If a single VOT of $\bar{\beta}$ is assumed for all the travelers, $\tau \geq \bar{\beta}(t^{SO} - t_F) = \bar{\beta}(t_S - t_F)x_S/d$.

When $\tau > \bar{\beta}(t_S - t_F)x_S/d$, defection is more costly than cooperation, and thus no defectors exist and no financial transactions.

When the penalty is exactly $\bar{\beta}(t_S - t_F)x_S/d$ and the penalty collected from defectors is distributed evenly to cooperators, defection is as costly as cooperation. Multiple cooperative schemes exist with different number of defectors, n, ranging from 0 to x_F. When $n = 0$, the cooperative scheme is the same as described above. When $0 < n < x_F$, cooperators have to use Fast Route proportionally less often than when $n = 0$ to maintain the SO flow pattern, and their average travel time

$$t^{CS}(n) = \frac{(x_F - n)t_F + x_S t_S}{d - n}. \tag{2}$$

$t^{CS}(n)$ is an increasing function of n, and $t^{CS}(n) > t^{CS}(0) = t^{SO}, \forall n > 0$. In other words, given that the total system travel time remains at the SO value, the reduction in travel time for defectors ($t_F < t^{SO}$)

is at the cost of cooperators in terms of increased travel time. The increased travel time is compensated for by the re-distributed defector payment. It can be shown mathematically that defectors and cooperators have the same generalized cost (combining time and monetary costs) equal to t^{SO} (in time units). Intuitively, the payment is a transfer within the system and thus does not affect total system cost, which remains the SO total travel time. The generalized costs of a defector and cooperator are equal as there are positive numbers of both (non-corner solution), and as a result, they must be equal to the average SO travel time. When $n = x_F$, the cooperative scheme degenerates to traditional congestion pricing with toll re-distribution, where no turn-taking is happening as Fast Route is filled up by defectors. Note that a cooperative scheme that does not maintain an SO flow pattern (but still better than UE in total travel cost) is still possible when $n = x_F$.

Heterogeneous VOT

Realistically, VOT is heterogeneous among travelers. Let β be the random VOT distributed over travelers. If the support of β has an upper bound $\check{\beta}$, it is trivial to show that a cooperative scheme (maintaining SO flow pattern) always exists with the defector penalty $\tau \geq \check{\beta}(t_S - t_F)x_S/d$. This is a much milder condition compared to those for Pareto-improving congestion pricing with toll re-distribution (Nie and Liu, 2010). An added advantage, as mentioned previously, is that no financial transactions are needed and thus the dooming perception of "tax" is avoided.

Practical considerations might lead to an upper bound on the defector penalty, for example, to avoid the perception of forced cooperation with the government. When $\tau < \check{\beta}(t_S - t_F)x_S/d$, a traveler with the threshold VOT, $\ddot{\beta} = (1/\tau)(t_S - t_F)x_S/d$, is indifferent between cooperation and defection. Travelers with a VOT higher than $\ddot{\beta}$ will defect while those with a VOT lower than $\ddot{\beta}$ will cooperate. The existence condition of the scheme is thus a condition to ensure that the number of defectors is no larger than x_F. With the same re-distribution scheme, it can be shown that every traveler is better off compared to UE[1], although the generalized cost is not equalized among travelers due to heterogeneous VOT.

A potentially more appealing scheme is to set the penalty to a value high enough but not too high, so that a certain number of defectors exist while the travel time of cooperators is strictly better than that in UE. It appeals to high-VOT travelers by giving them an option to pay for better travel time; it appeals to low-VOT travelers by reducing their travel time and on top of that, providing monetary rewards (re-distributed defector penalty). The number of defectors to ensure a strictly improving travel time for cooperators is such that

$$n < \frac{d(t^{UE} - t_S) + x_F(t_S - t_F)}{t^{UE} - t_F}. \tag{3}$$

Let $F_\beta(\cdot)$ be the cumulative distribution function of β. The penalty corresponding to a given number of defectors n is $F_\beta^{-1}(d - n)(t_S - t_F)x_S/d$.

[1]The traveler with threshold VOT, $\ddot{\beta}$, has a generalized cost of t^{SO} that is strictly better than t^{UE}. S/he can serve as a reference for a regular defectors or cooperators whose VOT is different from $\ddot{\beta}$. A regular defector has a higher VOT and the same travel time as the threshold defector. It follows that the travel time saving $t^{UE} - t_F$ is more valuable than for the threshold defector in terms of offsetting the penalty. The threshold defector is better off and thus the regular defector is also better off. A regular cooperator has a lower VOT and the same travel time as the threshold cooperator. When the regular cooperator's travel time is lower than that in UE, s/he is better off since the re-distributed penalty can only further reduce the generalized cost. When the regular cooperator's travel time is higher than that in UE, the increase in travel time is less detrimental than for the threshold cooperator in terms of offsetting the re-distributed penalty. The threshold cooperator is better off and thus the regular cooperator is also better off.

References

Boyd, R., Gintis, H., Bowles, S. and Richerson, P. J. (2003). The evolution of altruistic punishment, *Proceedings of the National Academy of Sciences* **100**(6): 3531–3535.

Daganzo, C. F. (1995). A pareto optimum congestion reduction scheme, *Transportation Research Part B: Methodological* **29**(2): 139–154.

Daganzo, C. F. and Garcia, R. C. (2000). A pareto improving strategy for the time-dependent morning commute problem, *Transportation Science* **34**(3): 303–311.

Dawes, R. M. (1980). Social dilemmas, *Annual Review of Psychology* **31**: 169–193.

de Palma, A. and Lindsey, R. (2011). Traffic congestion pricing methodologies and technologies, *Transportation Research Part C: Emerging Technologies* **19**(6): 1377–1399.

Fehr, E. and Gächter, S. (2002). Altruistic punishment in humans, *Nature* **415**(6868): 137–140.

Guo, X. and Yang, H. (2010). Pareto-improving congestion pricing and revenue refunding with multiple user classes, *Transportation Research Part B: Methodological* **44**(8): 972–982.

Helbing, D., Schönhof, M., Stark, H.-U. and Hołyst, J. A. (2005). How individuals learn to take turns: Emergence of alternating cooperation in a congestion game and the prisoner's dilemma, *Advances in Complex Systems* **8**(01): 87–116.

Klein, I. and Ben-Elia, E. (2016). Emergence of cooperation in congested road networks using ict and future and emergingtechnologies : A game-based review, *Transportation Research Part C: Emerging Technologies* **72**: 10–28.

Kockelman, K. M. and Kalmanje, S. (2005). Credit-based congestion pricing: a policy proposal and the public's response, *Transportation Research Part A: Policy and Practice* **39**(7): 671–690.

Lawphongpanich, S. and Yin, Y. (2010). Solving the pareto-improving toll problem via manifold suboptimization, *Transportation Research Part C: Emerging Technologies* **18**(2): 234–246.

Liu, Y., Guo, X. and Yang, H. (2009). Pareto-improving and revenue-neutral congestion pricing schemes in two-mode traffic networks, *NETNOMICS: Economic Research and Electronic Networking* **10**(1): 123–140.

Murnighan, J. K. and Roth, A. E. (1983). Expecting continued play in prisoner's dilemma games a test ofseveral models, *Journal of conflict resolution* **27**(2): 279–300.

Nie, Y. M. (2015). A new tradable credit scheme for the morning commuteproblem, *Networks and Spatial Economics* **15**(3): 719–741.

Nie, Y. M. and Liu, Y. (2010). Existence of self-financing and pareto-improving congestion pricing: impact of value of time distribution, *Transportation Research Part A: Policy and Practice* **44**(1): 39–51.

Pigou, A. C. (1920). *The economics of welfare*, Palgrave Macmillan.

Small, Kenneth A. an Verhoef, E. T. (2007). *The Economics of Urban Transportation*, 2nd edn, Routledge.

Vickrey, W. S. (1969). Congestion theory and transport investment, *The American Economic Review* pp. 251–260.

Wu, Z.-X. and Wang, Y.-H. (2007). Cooperationenhanced by the difference between interaction and learning neighborhoods for evolutionary spatial prisoner's dilemma games, *Physical Review E* **75**(4): 041114.

Yang, H. and Wang, X. (2011). Managing network mobility with tradable credits, *Transportation Research Part B: Methodological* **45**(3): 580–594.

Traffic & Mobility

TC2: Pricing Shared Mobility

Thursday 2:45 – 4:15 PM

Session Chair: Tunay Tunca

2:45 **Split the Bill by Sharing the Ride-Sharing Services: A Study on Optimal Ride-Sharing Pricing**
Jagan Jacob, Ricky Roet-Green*
University of Rochester

3:15 **Inventory Rebalancing and Minimum Stop-Over Routes for One-Way Electric Vehicle Sharing Systems**
[1]Yinglei Li, [2]Sung Hoon Chung*
[1]State University of New York at Binghamton, [2]Binghamton University

3:45 **An Empirical Analysis of Price Formation, Utilization, and Value Generation in Ride Sharing Services**
[1]Liu Ming, [2]Tunay Tunca, [1]Yi Xu, [2]Weiming Zhu*
[1]University of Maryland, [2]IESE Business School

Split the bill by sharing the ride-sharing services: A study on optimal ride-sharing pricing

Jagan Jacob & Ricky Roet-Green
Simon Business School
University of Rochester

Abstract

Real-time ride sharing is a one-time car sharing service made between matching users often on short notice (Amey et.al. 2011, Kowshik 1996). It is a two-sided market model where the firm offers a platform for peer-to-peer interaction between potential drivers and passengers. For instance, a big player in this relatively new transportation model is Uber. Uber is an online transportation company that offers its customers the service of transportation from their current location. Similar "Ride Sharing Service Provider" (RSP) companies include Lyft, Sidecar, Didi etc. Uber has shown tremendous growth since its inception in 2009 (Hall & Krueger 2015). Studies show that ride-sharing services are going to continue to grow (Hall & Krueger 2015, Chen et.al. 2015, Watanabe et.al. 2016). Unlike traditional taxi services, an RSP doesn't own a single vehicle in its "fleet". Instead, third party contractors, called "driver-partners" get transportation service requests from nearby, potential passengers via a smartphone (iPhone, Android etc.). The requests are sent by the passenger after she browses through all the drivers nearby and the expected waiting time for arrival. Using the GPS navigation system on the phone, the driver picks up the customer from her location. After the ride, the passenger pays for the service through the app. The RSP keeps a certain percentage of the net fare and the rest goes to the driver.

RSP's offers mainly two classes of services: (1) Shared/Pooled ride and (2) Solo ride. Shared/Pooled rides matches passengers with other passengers travelling in the same direction and charge them a lower fare than what they would have had to pay if they chose to travel alone. Services like UberPool and Lyft Line are examples of such modes. Thus, if a successful match is found, the passenger's travel expense is lower, and the lower price attracts more passengers thereby increasing the revenue of drivers and the RSP. Also, if more passengers choose to pool, there will be less cars on the road, thus reducing traffic congestion and environmental pollution. On the flip side however, there is a great deal of uncertainty whether a match will be found within a reasonable time. Passengers also incur a cost of sharing for riding with a stranger who may or may not be pleasant. From the drivers' side, they prefer two separate solo rides to one pooled ride since the former pays more and it is often too much of a hassle to locate, pick-up and drop off two different passengers in a single ride. In this study, we model the pros and cons of offering shared rides by the RSP and find the optimal strategies that the RSP should take depending on passenger parameters; when to offer pooled rides besides solo rides and how to price it optimally so as to maximize the revenue.

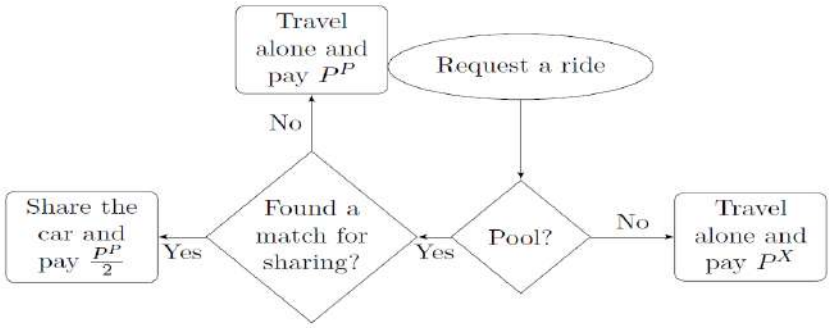


Fig 1: Travel Process

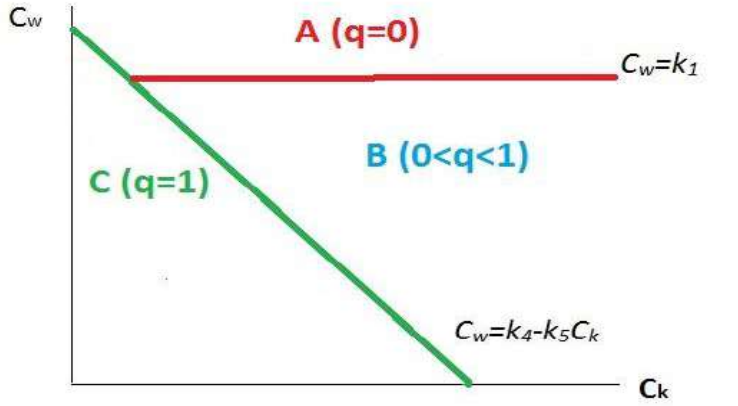
Whenever a passenger has to choose her options, it is reasonable to assume that the top two criteria she looks into are her total cost and expected arrival time at destination. It is also reasonable to assume that there are different types of passengers whose utility for cost and time are different such that they prefer different classes of vehicles. This opens up the potential for market differentiation. Since perfect price discrimination is not feasible in real life, the RSP can extract more consumer surplus using screening mechanism. In our model, we have included the cost of sharing besides the cost of waiting. The cost of sharing represents the 'costs' or disutility incurred by sharing the ride, that is not captured by the fare charged or the value of time for the passenger. . It encompasses the cost of privacy, the risk associated with travelling with a stranger etc. For example, if the passenger wants to make a phone call or is with a friend such that sharing the ride with a third party would be uncomfortable due to space constraints or if the passenger simply doesn't like to travel with a stranger, her cost of sharing will be high.

Traditional research on service industry based on queueing theory focuses on observable metrics like price and time. Vast literature on optimal service policies and design of queueing systems based on minimizing the expected delay or waiting time is available (Carmon et.al. 1995, Larson 1987, Anderson 1973). But not accounting the sharing cost and ignoring the perception of the customers towards waiting while being serviced can lead to policies that may be optimal in terms of minimizing the average delay, but may not be in the best interest of the firm in terms of long term profit. Hence to develop a model that captures the decision making process of passengers closer to real life scenario, we incorporate cost of sharing into our model as well. In this regard, our work is an addition to the existing literature.

We assume that there are sufficiently high number of cars available in the area so that the only waiting a passenger needs to make is to find a potential match to share the ride (if she chooses to share). If the request was for a solo ride, the passenger starts her journey and the journey time is distributed according to an exponential distribution. At the end of the journey, the passenger pays a specific fare. In contrast, if the request was for a pooled ride, the RSP searches for a potential match and if a match is not found within a maximum specified time, the passengers makes the journey solo, with journey time distributed according to an exponential distribution with same rate as a solo ride but pays a lower fare. However, if a match is found, she shares the ride and has to pay only half of the above rate, but her journey time is now distributed according to an exponential distribution, with mean travel time greater than that of a solo ride.

We begin our analysis by assuming that all passengers are identical with respect to their cost of waiting time and cost of sharing the ride. Later we will consider different types of passengers with respect to the costs. Our goal is to find the equilibrium strategy ‘q’, which is the probability of a passenger choosing to share the ride. Hence q=0 implies nobody wants to share the ride, q=1 implies

everyone opts to share and 0<q<1 implies passengers are indifferent between sharing and riding solo, which can be viewed as if the passengers decide the mode of travel by flipping a biased coin. We find explicit expressions for the probability of successfully finding a match if a passenger choose to share. Based on the utility functions for solo and shared ride, the probability of getting a match and all other given parameters, we find that a unique symmetric Nash equilibrium strategy exists. We provide explicit conditions over the cost of waiting and the cost of sharing to determine whether the unique equilibrium strategy is pure (to ride solo or to pool) or mixed (pool with certain probability). We find that omitting sharing costs from the model, in at least some cases give different equilibrium strategies, which indicates that RSPs should include it in addition to the cost of waiting time for the passengers.

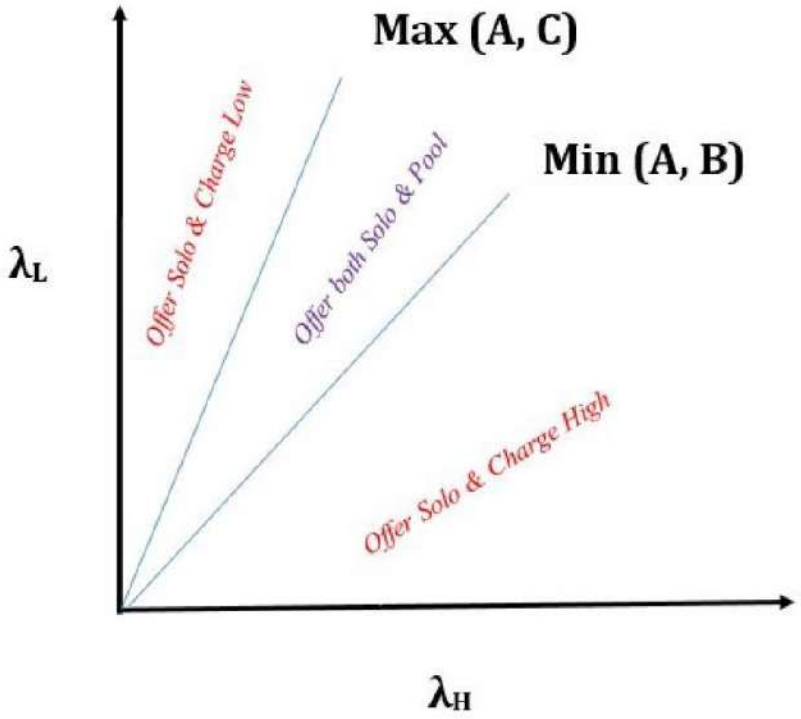


Equilibrium strategy of passengers based on their cost of sharing (Ck) and cost of waiting (Cw)

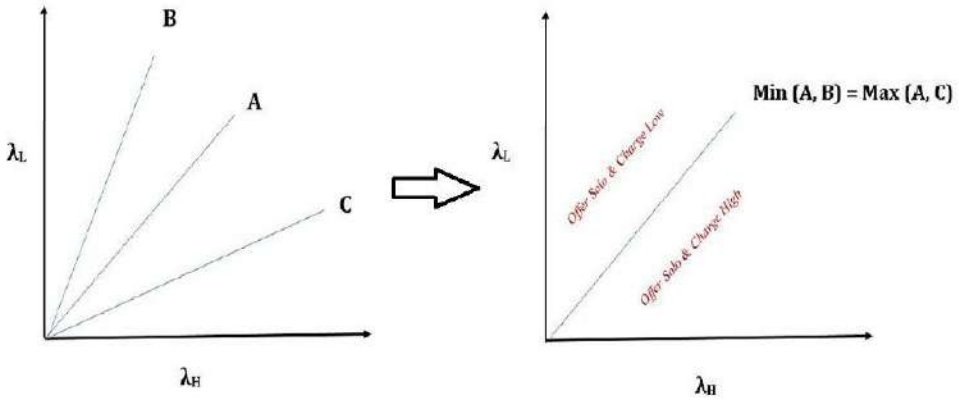
Once we find the equilibrium strategies for the passengers, we find the optimal prices to charge the two modes of travel that maximizes the RSPs revenue. Both homogeneous and heterogeneous passenger base are considered. For a homogenous set of passengers, the optimal strategy of the RSP is found to offer only solo rides and set the price that extracts the entire surplus from the passengers. In case of heterogeneous passenger base, we consider two types of passengers with different journey valuation and cost of waiting. The type of a passenger is private information, i.e., only the passenger knows their type with certainty and the RSP cannot distinguish between the two. Every other parameters are assumed to be common knowledge.

Because of asymmetrical information, perfect price discrimination is not possible by the RSP. Hence it implements second degree price discrimination based on service quality differentiation (Fudenberg & Tirole 1991, Varian 1989, Maskin 1984). We employ self-selection mechanism design and find explicit conditions required for a strategy to be optimal. We also find conditions under which it is never optimal to offer pooled rides.

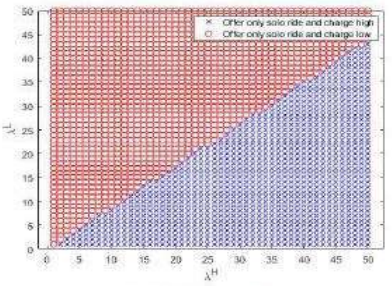
So far we have assumed that the number of drivers was unlimited. However, the prices that the RSPs like Uber charge the customers has a direct effect on the willingness of drivers to provide the service as the drivers get paid a specific proportion of the revenue. So the availability of drivers and cars in the market is a finite value, dependent on the prices charged by the RSP. Our research aims to analyze the impact of limiting the number of available cars on the design of the optimal price menu. We also try to analyze how the pricing strategies would affect the traffic density in the steady state by finding the proportion of pooled rides vs. solo rides.



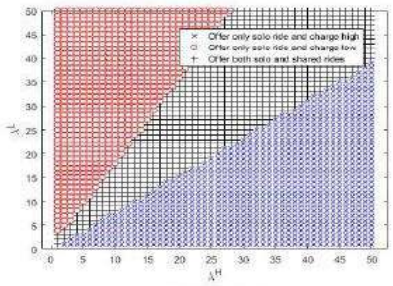
(a) Optimal RSP strategy depending on λ^H, λ^L. Note A $= \frac{GS_x^H}{GS_x^L} - 1$, B $= \frac{GS_p^H}{GS_p^L} - k$, C $= \frac{EC_p^H - (1-k)GS_x^L}{EC_p^L} - k$



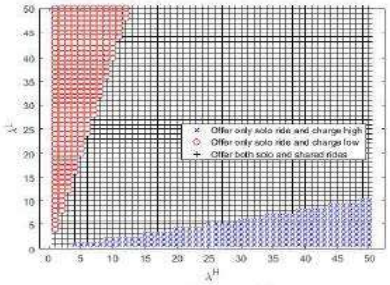
(a) Let A $= \frac{GS_x^H}{GS_x^L} - 1$, B $= \frac{GS_p^H}{GS_p^L} - k$, C $= \frac{EC_p^H - (1-k)GS_x^L}{EC_p^L} - k$. If C<A<B, then it is never optimal for the RSP to offer pooling



(a) $C_k^L >> C_k^H$



(b) $C_k^L = C_k^H$



(c) $C_k^L << C_k^H$

As the cost of sharing for the high type increase more than that of the low type, the region where providing both pooled and solo rides expanded. The reason being, the high type are willing to pay even more to avoid sharing their ride which increases the total revenue under that strategy. The opposite happens when the cost of sharing for the low type increase more than the high type, since the higher cost of sharing eats into the gross surplus, leaving only less surplus for the RSP to extract.

References

- Amey, A., Attanucci, J., Mishalani, R.A., (2011) "Real-time ridesharing: Opportunities and challenges in using mobile phone technology to improve rideshare services." *Transportation Research Record: Journal of the Transportation Research Board*, 2217:103-110.
- Anderson, Rolph E., (1973) "Consumer dissatisfaction: The effect of disconfirmed expectancy on perceived product performance." *Journal of Marketing Research*, 10(1):38-44.
- Carmon, Z., Shanthikumar, J.G., Carmon, T.F., (1995) "A psychological perspective on service segmentation models: The significance of accounting for consumers' perceptions of waiting and service." *Management Science*, 41:1806-1815.
- Hall, J., Krueger, A.B., (2015) "An analysis of the labor market for Uber's driver-partners in the United States." Working Papers (*Princeton University. Industrial Relations Section*).
- Kowshik, R.R., (1996) "Evaluation of the Sacramento-area real-time rideshare matching field operational test final report." *Institute of Transportation Studies, University of California, Davis: Research Report* UCD-ITS-RR-96-02.
- Larson, Richard C., (1987) Perspectives on queues: "Social justice and the psychology of queueing." *Operations Research*, 35:895-905.
- Mislove A. Wilson C. Chen, L., (2015) "Peeking beneath the hood of Uber". *Proceedings of the 2015 ACM Conference on Internet Measurement Conference*
- Watanabe, C., Naveed, K., Neittaanmaki, P, (2016) "Co-evolution of three mega-trends nurtures un-captured GDP - Uber's ride-sharing revolution." *Technology in Society*, 46: 164-185.
- Riley J. Maskin, E. "Monopoly with incomplete information". *RAND journal of economics*, 15(2):171-197, 1984.
- Tirole J. Fudenberg, D. "Game theory" *The MIT Press*, 1991.
- H.R. Varian. "Price discrimination". *Handbook of industrial organization*, pages 598-654, 1989.

Inventory Rebalancing and Minimum Stop-Over Routes for One-Way Electric Vehicle Sharing Systems

Yinglei Li Sung Hoon Chung

Abstract

Electric vehicles are generally considered as an environmentally friendly alternative to conventional vehicles that use fossil fuels. Car sharing systems have gained an increasing attention due to their positive aspects such as increasing mobility, reducing pollution, and mitigating traffic congestion. Electric vehicle sharing systems then have a great potential to further promote environmental sustainability. In this paper, we tackle the problem of rebalancing the inventory of one-way electric vehicle sharing systems by proposing a novel dynamic pricing approach. We propose a model that helps establish a self-balancing inventory system through dynamic pricing for which a system operator proposes to incoming users a better price if they choose an alternative combination of pick-up and drop-off stations that will improve the system inventory balance. The inventory matrix is updated dynamically every time a reservation is made and a vehicle return occurs, and the requirement that electric vehicles need to be recharged after a return to be available for the next user is taken into account. In addition, we design a stop-over station recommendation system for trips longer than the range of electric vehicles, which we call minimum stop-over path cover problem. Furthermore, we present computationally tractable algorithms to implement the proposed model, which is a requisite for implementing dynamic pricing and stop-over station recommendation system. We provide examples in which the proposed dynamic pricing approach and a staff-based rebalancing method are compared.

Keywords: Electric vehicle sharing, One-way car sharing, Dynamic pricing, Minimum stop-over path cover

Electric vehicles (EVs) are generally considered as an environmentally friendly alternative to conventional vehicles, as EVs can reduce local air pollution, green house gas (GHG) emissions, and traffic noise. Car manufacturers have commercialized several EVs, especially targeting the urban

market. There are, however, barriers that prevent the wide spread adoption of EVs, such as the lack of charging infrastructure and higher price than conventional cars.

In urban areas, owning a car could be challenging due to expensive parking fees, high maintenance cost, and traffic congestion especially during rush hours. Public transportation is widely used for a daily commute and for some other activities but urban residents may still need a car from time to time for, e.g., shopping and weekend excursions. A car sharing system has reduced the need for car ownership as users can borrow a car when it is needed (on-demand) in a convenient (payment using a smart phone application) and cost effective way (by-the-minute payment, no worries for parking fee). Car sharing systems also have a great potential to contribute to the environment and sustainability thanks to its positive aspects such as reducing pollution and mitigating traffic congestion. Recently, a one-way car sharing system has gained a surge of interest as it provides enhanced mobility and flexibility by allowing users to have flexible return times and stations. There are, however, management issues arising with the successful launch of one-way car sharing systems, one of which is the inventory imbalance problem caused by spatially dispersed, asymmetric return demands that are different than pick-up locations. The question is then how to effectively manage the system inventory balance, such that users can pick up a vehicle from a location of preference at a desired time and return it to a drop-off location close to a final destination when they want to return.

In light of the aforementioned discussions, of our interest is to consider an electric vehicle for car sharing systems as a reasonable means to foster the introduction of electric vehicles and to promote car sharing systems at the same time. We claim that the combination of the two, i.e., electric vehicle sharing systems have a great potential to further promote environmental sustainability, even though this may pose new challenges such as limited driving range and battery recharging issues in the context of vehicle sharing. Indeed, several electric vehicle sharing systems and services have been developed and introduced in some countries, and it is expected that electric vehicle sharing systems will continue to grow all around the world.

In this paper, we tackle the problem of rebalancing the inventory of one-way electric vehicle sharing systems by proposing a novel dynamic pricing approach. We propose a model that helps establish a self-balancing inventory system through dynamic pricing for which a system operator proposes to incoming users a better price if they choose an alternative combination of pick-up and

drop-off stations that will improve the system inventory balance. A sophisticated pricing mechanism is presented and the inventory matrix that allows flexible return times and stations is developed for the implementation of the proposed model. The inventory matrix is updated dynamically every time a reservation is made and a vehicle return occurs. Note that unlike the conventional car sharing systems, for electric vehicles there is a gap between the time of vehicle return and the time at which the vehicle becomes available for the next user. This requirement is explicitly taken into account in the inventory matrix. We compare the proposed dynamic pricing approach with a staff-based rebalancing method with a focus on the total cost for rebalancing. In addition, we design a stop-over station recommendation system that will allow trips longer than the range of electric vehicles, which we call minimum stop-over path cover (MSPC) problem. Furthermore, we present computationally tractable algorithms to implement the proposed model, which is a requisite for implementing dynamic pricing and stop-over station recommendation system.

We consider two approaches for rebalancing the system inventory. The first approach is to use a dynamic pricing method in an attempt to minimize the number of imbalance stations over the course of the planning horizon. We assume that a system operator obtains information on pick-up and drop-off stations and rental duration when a reservation is made or when an on-demand pick-up occurs. That is, while the electric vehicle sharing system allows on-demand pick-ups, a trip information is required before a user can actually pick-up a vehicle and start a trip. Based on this assumption, we propose a model that helps establish a self-balancing inventory system. If the drop-off station (for the case of on-demand pick-ups) or the pair of pick-up and drop-off stations (for the case of reservations made in advance) make the system inventory more imbalanced, then the operator can suggest nearby stations at a better price, which will improve the system inventory balance. We also assume that users always seek the most economical ways, such that they will choose the alternative stations if the total cost including the traveling cost from original stations to nearby stations is lower than the one between original stations.

Even though the proposed dynamic pricing approach is designed to establish a self-balancing inventory system, there is still a chance that imbalanced stations will remain at the end of the planning horizon. The second approach is to use staff members to relocate vehicles at the end of the planning horizon, thereby providing all balanced stations before the beginning of the next planning horizon, while providing vehicle rental prices proportional to the rental duration. We compare these

two approaches in terms of the number of imbalanced stations at the end of the planning horizon and the total cost to make the system inventory balanced for the next planning horizon through several examples.

We provide 3 examples to show how our proposed models can be used in electric vehicle sharing problems. In particular, example 1 is used to show the relatively small area with 50 randomly generated EV stations. Example 2 is used to show the relatively large area with 100 randomly generated EV stations. Example 3 is used to simulate the distribution of zipcar stations in New York City. We consider three cases in example 1. We evaluate each example with six performance measures: (1) number of imbalanced stations at the end of planning horizon, (2) percentage of imbalanced stations, (3) total staff travel distance, (4) total number of unserved customers, and (5) total cost. From the results, we conclude that the dynamic pricing method has a significant potential to improve the service level by reducing the potential number of unserved customers and by reducing the total cost. There are several extensions possible in the future work, examples of which include customers' willingness to use alternative stations depending on the distance to neighborhood stations and monetary compensation.

An Empirical Analysis of Price Formation, Utilization, and Value Generation in Ride Sharing Services

Liu Ming*[1], Tunay I. Tunca†[1], Yi Xu‡[1], and Weiming Zhu§[2]

[1]Robert H. Smith School of Business University of Maryland

[2]IESE Business School, University of Navarra

January 5, 2017

Extended Abstract

Motivation

The advancement of mobile Internet technologies have led to the birth of ride-sharing platforms such as Uber and Lyft and their Chinese counterpart Didi. These ride-sharing platforms link consumers and drivers in an efficient and real-time way through mobile applications. The dynamics of both demand (consumers) and supply (drivers) on ride-sharing platforms are distinctively different from the ones in traditional taxi industry. Through ride-sharing mobile applications, consumers can choose a range of car services and can instantly review the number and locations of nearby active cars in real time, and make decisions based on up-to-date information about availability and pick up times. On the supply side, the capacity or the number of active drivers is not centrally controlled by the ride-sharing platforms. Instead, the number of drivers that are working at a specific time is indirectly determined by the independent choices of individual drivers.

Given these unique market dynamics, ride-sharing platforms often set prices dynamically to balance the demand and supply for their services. Fare setting rules and proprietary algorithms used by these services for managing their operations play an important role in value generation for their customers as

*Robert H. Smith School of Business, University of Maryland, College Park, MD 20742. E-mail:liu.ming@rhsmith.umd.edu
†Robert H. Smith School of Business, University of Maryland, College Park, MD 20742. E-mail:ttunca@rhsmith.umd.edu
‡Robert H. Smith School of Business, University of Maryland, College Park, MD 20742. E-mail:yxu@rhsmith.umd.edu
§IESE Business School, 21 Avenida Pearson, Barcelona 08034, E-mail:wzhu@iese.edu

well as their drivers, and important questions center around their pricing and regulation. In this paper, we develop an empirical model, and analyze price formation and surplus generation of these services using a comprehensive dataset obtained from Didi, the largest ride-sharing platform in China (with a market share of more than 90%), who provides taxi as well as ride-sharing services (known as Kuaiche) on its platform while dynamically setting the prices to better match the supply with the demand. The particular research questions we study are

1- What are the factors that determine the consumer demand and driver willingness to participate in ride-sharing platforms?

2- What are the welfare effects of certain proposed government regulations on pricing and capacity of ride-sharing services?

Empirical Model and Estimation

We first develop an empirical model to capture the demand and supply dynamics on a ride-sharing platform like Didi. We apply a discrete choice model to describe both customers' choices among different services and drivers' choices on whether to drive at a given time. On the demand side, when a consumer is looking for a service on Didi's platform, the consumer needs to select a service/option $j \in \{K, T, O\}$ where K denotes *Kuaiche*, T the *Taxi* and O the outside option such as taking public transportation. Let A denote the consumer (demand) side and B denote the driver (supply) side of Didi's platform. Following a multinomial logit choice model framework, at any time t, consumer i's utility function of choosing Kuaiche can be written as

$$u_{it}^{KA} = \mu^{KA} + \alpha^{KA} N_{it}^{KB} + \beta^{KA} p_{it}^{KA} + \gamma^{K} w_{it}^{K} + \delta^{KA} D_t + \xi_t^{KA} + \epsilon_{it}^{KA}, \tag{1}$$

where μ^{KA} denotes the fixed effect of Kuaiche on Didi platform on the consumer side; N_{it}^{KB} be the number of available Kuaiche drivers consumer i observes at time t; p_{it}^{KA} is the total price for consumer i's trip by Kuaiche at time t. w_{it}^{K} is the waiting time for consumer i's order to be taken by a Kuaiche driver at time t. ξ_t^{KA} and ϵ_{it}^{KA} are random effects of the Kuaiche service faced by consumers where ϵ_{it}^{KA} follows an $i.i.d.$ Gumbel distribution for each i at time t. D_t is a vector of control variables including day dummies, highest and lowest temperatures, weather and air quality. Consumer i's utility function of choosing Taxi is given in a similar form. Consumer i's utility of choosing the outside option is given as

$$u_{it}^{OA} = \mu^{OA} + \epsilon_{it}^{OA}, \tag{2}$$

where μ^{OA} is the fixed effect of the outside option, and ϵ_{it}^{OA} is the random effect of the outside option faced by consumer i, and have $i.i.d.$ Gumbel distribution at each time t. Without loss of generality, we normalize

μ^{OA} to 0. A consumer will select the option that will generate the highest utility. So, consumer i will choose service/option $j \in \{K, T, O\}$ at time t, if $u_{it}^{jA} > u_{it}^{kA}$, for all $k \neq j$, and $k \in \{K, T, O\}$. Given the utility functions, the market share of service $j \in \{K, T\}$ offered on the platform at time t can be derived based on the multinomial logit choice probability calculations, which specify the demand dynamics on the platform.

On the supply side, Didi offers two services: Kuaiche and Taxi. For Kuaiche driver i, we assume that the utility of driving, i.e., taking Kuaiche orders from the Didi platform, at time t is given as

$$u_{it}^{KB} = \mu^{KB} + \alpha^{KB} N_{it}^{KA} + \beta^{KB} p_t^{KB} + \gamma^K v_{it}^K + \delta^{KB} D_t + \xi_t^{KB} + \epsilon_{it}^{KB}, \tag{3}$$

where μ^{KB} denotes the fixed effect of Kuaiche service on the platform in the driver's side; N_{it}^{KA} be the number of Kuaiche orders submitted on the Didi platform around driver i's location at time t; p_t^{KB} is the expected payoff Kuaiche drivers receive on a trip at time t; v_{it}^K is the cumulative money Kuaiche drivers has earned up to time t in a particular day. D_t is the vector of control variables. ξ_t^{KB} and ϵ_{it}^{KB} are random effects of driving Kuaiche, and ϵ_{it}^{KB} have *i.i.d.* Gumbel distribution. If a Kuaiche driver decides not to drive at time t, the driver can take the outside option. We use superscript O_K to denote the not driving or outside option for Kuaiche drivers. For Kuaiche driver i, we assume that the utility of not driving at time t is

$$u_{it}^{O_K B} = \mu^{O_K B} + \epsilon_{it}^{O_K B}, \tag{4}$$

where $\mu^{O_K B}$ is the fixed effect of not driving for a Kuaiche driver, and $\epsilon_{it}^{O_K B}$ is the random effect of not driving for Kuaiche driver i which follows an *i.i.d.* Gumbel distribution at each time t. Without loss of generality, we normalize $\mu^{O_K B}$ to 0. Based on the utitlity functions, the proportion of registered Kuaiche drivers that decide to drive at time t can be derived, again through multinomial logit choice probability calculations, specifying the supply dynamics of Kuaiche drivers on the platform. Taxi drivers' participation decisions are modeled in a similar way.

Based on the empirical model of demand and supply on the platform, we then estimate the model parameters using three datasets, two from Didi and one from public sources, to conduct the empirical analysis. The first dataset we obtained from Didi contains transaction level data in Beijing including observations on the locations at which each customer is picked up and dropped off, the time the customer submits a request through the mobile app and the time the driver accepts the request, the drop-off time, and the total fare of the ride. The second dataset from Didi contains the driver's information. Finally, we also employ an independent public data to control for other factors such as the weather variation during the timespan of the Didi dataset.

Performing the estimation, we find that price has different effects on consumers' choices of Kuaiche in

different time periods in a day. It has a significantly negative effect on consumer's choice of Kuaiche during the night time, i.e., 11 pm-6 am, while its effect is not significant during morning rush hour from 7 to 9 am. The after-dinner period, from 8-11 pm, also demonstrate lack of consumer sensitivity to the price. This can be explained from two aspects, namely reduced availability of transportation choices in the evening, and consumers' higher valuation of convenience and comfort.

On the supply side, Kuaiche drivers' choices of whether to drive or not depend on the observable parameters like the number of open orders around, current estimated price per order as well as how much money the driver has made so far in a day. The number of consumers is always a positive incentive for Kuaiche drivers in all time periods. The price effect, however, is negative and significant in 11 pm-6 am, 6-7 am and 9-10 am and positive in other time periods. This can be explained by drivers' target-driven behavior. Drivers care less about the price during the morning rush hour since the number of orders is the dominating consideration given the higher congestion level. Taxi market share is affected by the number of available taxi drivers positively for all time periods. Price per order for taxi service is negative or insignificant with its market share for all time periods except the night time. Limited public transportation forces consumers to take taxi during night time even if the night price is higher.

Counterfactual Analysis on Government Regulation and Welfare

We then utilize our estimation results on the demand and supply dynamics to investigate the impact of potential government regulations on pricing and capacity through counterfactual analysis. Governments around the globe are often concerned about the lack of regulations in ride-sharing industry. Based on concerns that existing taxi companies may get severely affected by the low base price and the large number of Kuaiche drivers of Didi, and that customers may suffer from the surge price, in China there are proposed regulations commonly focusing on enforcing constraints on pricing and limiting the number of Kuaiche drivers in a city. To evaluate the impact of government regulations on ride-sharing platforms, we estimate both consumer and driver welfare through log-sum method, and evaluate the difference between the current levels and scenarios that assume the implementation of price caps and capacity limitations on Kuaiche service.

We first examine the impact of price caps imposed by a regulator on Kuaiche prices based on taxi prices. Under the proposed price caps, Kuaiche prices cannot be lower than taxi prices at a specific time. With other model parameters remaining the same, we solve the nonlinear equation system estimated to study the dynamics of Kuaiche and taxi drivers' responses to the pricing regulation on Kuaiche in different time periods. We find that when the price for Kuaiche is regulated to be no lower than that of taxi, the customer welfare will decrease in both weekdays and weekends with an average reduction of 61.31%. The decrease

comes from the interaction of various connected factors. First the the increase of price reduces consumer's welfare directly. The drop of consumer orders taking Kuaiche afterwards further decreases consumer welfare because Kuaiche drivers' willingness to drive is also negatively affected. As a result, although a group of consumers may benefit from shorter waiting times due to larger decrease of consumers requesting Kuaiche than the number of available Kuaiche drivers, price regulation dramatically decreases consumer welfare as a whole. What is more, the increased price leads to a 5.31% decrease in total Kuaiche driver welfare.

We find that taxi drivers that participate in the platform do not benefit from the price caps on Kuaiche either, with an average welfare loss of 20.12%. There are two reasons for this unexpected observation. First, although the Kuaiche prices are regulated to be no less than taxi prices, many Kuaiche customers would likely respond by taking choose an outside transportation option. This, creates an imbalance in favor of supply for the Kuaiche service, which results in relatively less waiting times for Kuaiche, and attracts some taxi consumers to switch to Kuaiche. Therefore, a demand reduction occurs in the taxi market, which reduces the surplus of the taxi drivers that offer services on Didi's platform.

We then analyze the effects of potential capacity regulation on Kuaiche by considering the proposed scenario of limiting the number of Kuaiche drivers from the current 95,000 to the current number of taxi cabs, 66,000 in Beijing. Performing the analysis, we can observe a significant decrease of 50.06% in total consumer welfare and a decrease of 56.35% in Kuaiche driver welfare under this scenario. We find that the shortage of Kuaiche capacity is likely decrease consumers' willingness to choose the Didi platform and further jeopardize Kuaiche drivers' incentives to participate because of the decreased number of orders. However, our analysis shows that taxi drivers would also experience a loss of surplus in this case, and the joint effect of the changes of consumers and drivers due to capacity regulation causes various degrees of losses to all participants on the platform.

Our results suggest that common government regulations on pricing and capacity of ride-sharing platforms could perform poorly in enhancing the social welfare and balancing demand and supply in the market. Hence, more careful design of regulating mechanisms is needed to improve the efficiency and social benefits of ride-sharing platforms.

Traffic & Mobility

TD2: Optimization in Shared mobility

Thursday 4:30 – 6:00 PM

Session Chair: Carolina Osorio

4:30 **Optimizing the Profitability and Quality of Service in Carshare Systems Under Demand Uncertainty**
[1]*Mengshi Lu**, [2]*Siqian Shen*, [2]*Zhihao Chen*
[1]*Krannert School of Management-Purdue University*, [2]*Industrial and Operations Engineering-University of Michigan*

5:00 **Ridesharing in a Mobility-On-Demand System**
[1]*Samitha Samaranayake**, [2]*Harshith Guntha*, [3]*Kevin Spieser*, [4]*Emilio Frazzoli*
[1]*Cornell University*, [2]*IIT Chennai*, [3]*Uber*, [4]*Massachusetts Institute of Technology*

5:30 **A Discrete Simulation-Based Optimization Algorithm for Two-Way Car-Sharing Network Design**
[1]*Tianli Zhou**, [2]*Carolina Osorio*, [3]*Evan Fields*
[1]*MIT-CEE*, [2]*MIT-CEE/MIT-ORC*, [3]*MIT-ORC*

Optimizing the Profitability and Quality of Service in Carshare Systems under Demand Uncertainty

Mengshi Lu
Email: mengshilu@purdue.edu
Krannert School of Management, Purdue University, West Lafayette, IN, USA

Siqian Shen, Zhihao Chen
E-mails: siqian, czhihao@umich.edu
Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI, USA

Abstract

We consider allocating a carshare fleet to service zones under uncertain one-way and round-trip demand. We use a two-stage stochastic integer program, where in the first stage, we allocate shared vehicle fleet and purchase parking lots or permits; in the second stage, we generate finite samples to represent uncertain location-time-based carshare demand and construct a spatial-temporal network for each sample to model vehicle movement. We minimize the expected total costs minus profit, and develop branch-and-cut algorithms with mixed-integer rounding (MIR)-enhanced Benders cuts. We also implement the two-stage model in a rolling-horizon fashion for real-time vehicle relocation. We test instances generated based on Zipcar data in Boston to demonstrate the efficacy of our approaches and draw insights on carshare management.

1 Introduction

Increasing energy prices and parking fees in large cities have led to increased vehicle ownership costs, resulting in many individuals turning to other means of transportation. In recent years, carsharing has become a popular means of alternative transportation, serving as a middle ground between public transport and private ownership. Carshare systems can be broadly categorized into two types: reservation-based, in which customers must reserve vehicles prior to using them (e.g., Zipcar), and free-float, in which customers can pick up any available car for immediate use (e.g., Car2Go). Carshare rentals can be categorized into one-way and round-trip rentals, with the former allowing customers to rent and return vehicles in different locations, and the latter only allows returning cars to the same location.

From the customers' point of view, the availability of one-way rentals provides two main benefits. Firstly, there can be more flexibility in vehicle use. Secondly, customers can potentially save on rental fees by splitting a round-trip rental into two one-way rentals. From the point of view of a carshare company, providing such a service has not been a priority because of added management complexities, among which, the most significant one is planning for imbalances in demand, e.g., via vehicle relocation, which can be costly.

1.1 Problem Description and Solution Approaches

In this paper, we consider allocating a carshare fleet in a region serviced by a carshare company to satisfy uncertain travel demand. The region is discretized into smaller zones, with parking costs different from zone to zone. To regulate carshares, city governments issue parking lot contracts and free-float parking permits for shared cars. We are tasked to allocate a homogeneous fleet of vehicles to zones, by determining the number of contracted parking lots (for reservation-based systems) or the number of parking permits (for free-float systems) to purchase in each zone.

The above car fleet allocation decisions are made "here-and-now" before one-way and round-trip demand is realized at discrete time periods over a finite horizon. In each period, should the

demand in a zone exceeds the number of vehicles available for use at that zone, any excess demand is immediately lost (i.e., we assume that any excess demand is not carried over to the next period). During the finite horizon, we may relocate vehicles as recourse actions, and customers will be unable to use vehicles that are being relocated. Given each demand realization, we find the optimal vehicle movement to maximize the net profit minus the penalty of undesirable QoS results associated with unserved demand.

1.2 Contribution and Main Results

This paper focuses on a strategic planning problem of purchasing parking lots/permits and allocating initial vehicle fleet in service zones, to satisfy uncertain one-way and round-trip carshare demand. We optimize both profitability (given by carshare revenue minus the relocation cost) and QoS (determined by unserved demand, total vehicle idleness, and denied demand trips). We formulate a two-stage stochastic programming model with integer variables in the first stage, while using spatial-temporal networks to model vehicle movement corresponding to different demand realizations in the second stage.

We conduct a numerical study using augmented real-world data, to demonstrate the effectiveness of stochastic optimization approaches for carshare management under uncertainty. Our results show that the proportion of one-way rentals can significantly impact the profitability and QoS of carshare systems. However, the specific impact can be completely different depending on whether one-way demand is exogenously or endogenously generated. If the proportion of one-way demand is exogenously given, we show that higher one-way proportion can increase a carshare system's profitability. On the other hand, if the proportion of one-way demand is endogenously determined by pricing and strategic customer behavior, we show that higher one-way proportion could decrease a carshare system's profitability. When one-way demand is exogenous, increasing the proportion of one-way demand will first cause QoS measures (e.g., unserved demand and denied trips) to worsen. However, as the proportion further increases, the QoS measures will improve. When one-way demand is endogenous, increasing the proportion of one-way demand always leads to improvements of the QoS. Furthermore, the effective use of vehicle relocation plays an important role in carshare systems with one-way demand. The number of vehicle relocations will increase as one-way demand increases. The reservation-based system with limited parking spaces has higher cost of vehicle relocation but lower cost of parking compared to the free-float system that uses parking permits.

2 Problem Formulation

Consider the problem where a carshare company needs to allocate a given budget of S vehicles in a set of zones, denoted I, to maximize its profit and QoS over T time periods, using contracted parking lots (for a reservation-based system, denoted by $m = 1$) and purchased parking permits (for a free-float system, denoted by $m = 2$). Decisions include the number of parking lots to purchase in zone i, denoted w_i, the number of vehicles deployed in zone i that require contracted parking spaces, denoted x_i^1, and the number of vehicles allocated in zone i with purchased parking permits, denoted x_i^2, for all $i \in I$. Let c_i^{lot} be the cost of acquiring one parking space, c_i^{loc} be the cost of allocating a vehicle in zone i, for all $i \in I$, and c^{ffp} be the cost of one free-float parking permit. We denote the demand for one-way rentals from zone i to zone j starting at period t and ending at period s by d_{ijts}^{one}, and denote the demand for round-trip rentals from zone i starting at period t and ending at period s by d_{its}^{two}. The time taken to travel from zone i to zone j is denoted ℓ_{ij}, and the data satisfies $s - t \geq \ell_{ij}$ for any $d_{ijts}^{\text{one}} > 0$. We denote the revenue per period per vehicle from one-way rentals by $r^{\text{one}} \geq 0$ and that from round-trip rentals by $r^{\text{two}} \geq 0$; the relocation cost per

period per vehicle is $c^{\text{rel}} \geq 0$. When a vehicle is in use, whether during one-way rentals, round-trip rentals, or relocation, it incurs a maintenance cost of $c^{\text{mnt}} \geq 0$ per period; when it is not in use, it incurs an idle cost of $c^{\text{idle}} \geq 0$ per period.

To model zone-to-zone vehicle movement over T periods, we construct a spatial-temporal network $G(N, A)$, with each node $n_{it} \in N$ representing a zone $i \in I$ at period $t \in \{0, 1, \ldots, T\}$. The set A is the union of four types of arcs, namely, one-way, round-trip, relocation, and idle arcs, i.e., $A = A^{\text{one}} \cup A^{\text{two}} \cup A^{\text{rel}} \cup A^{\text{idle}}$. We denote the unit cost of flow and the capacity of arc a by f_a and u_a, respectively, while $\delta^+(n_{it})$ and $\delta^-(n_{it})$ denote the sets of arcs for which n_{it} is their origin and destination node, respectively. The unit flow costs and capacities of each type of arcs are summarized in Table 1.

Table 1: Unit flow costs and capacities for each arc type

Type of arc	Cost per unit flow f_a	Capacity u_a
One-way arc (n_{it}, n_{js})	$-(r^{\text{one}} - c^{\text{mnt}})(s-t)$	d^{one}_{ijts}
Round-trip arc (n_{it}, n_{is})	$-(r^{\text{two}} - c^{\text{mnt}})(s-t)$	d^{two}_{its}
Relocation arc $(n_{it}, n_{j,t+\ell_{ij}})$	$(c^{\text{rel}} + c^{\text{mnt}})\ell_{ij}$	$+\infty$
Idle arc $(n_{it}, n_{i,t+1})$	c^{idle}	w_i $(m=1)$; $+\infty$ $(m=2)$

2.1 A Two-Stage Stochastic Integer Programming Formulation

We employ a two-stage stochastic program, where $\mathbf{w} \in \mathbb{Z}_+^{|I|}$ denotes the vector of w_i's, and $\mathbf{x}^m \in \mathbb{Z}_+^{|I|}$ denotes the vector of x_i^m's with $m = 1, 2$. Moreover, let $c_i^1 = c_i^{\text{loc}}$, $c_i^2 = c^{\text{ffp}} + c_i^{\text{loc}}$, for each zone $i \in I$, and $M = \{1, 2\}$. We formulate the carshare fleet allocation problem as

$$\min_{\mathbf{w},\mathbf{x}^1,\mathbf{x}^2} \quad \sum_{i\in I}\left(c_i^{\text{lot}} w_i + \sum_{m\in M} c_i^m x_i^m\right) + Q(\mathbf{w}, \mathbf{x}^1, \mathbf{x}^2) \tag{1}$$

$$\text{s.t.} \quad (\mathbf{w}, \mathbf{x}^1, \mathbf{x}^2) \in X = \left\{\mathbf{w}, \mathbf{x}^1, \mathbf{x}^2 \in \mathbb{Z}_+^{|I|} : \sum_{i\in I}\sum_{m\in M} x_i^m \leq S,\ x_i^1 \leq w_i,\ \forall i \in I\right\}. \tag{2}$$

The function $Q(\mathbf{w}, \mathbf{x}^1, \mathbf{x}^2)$ returns the optimal cost in the second stage, which optimizes flows in the spatial-temporal network given that the supply level at each node n_{i0} is x_i and the capacity on each arc a is u_a determined by the random demand and the number of parking spaces w_i, for all $i \in I$. We define recourse decisions $y_a^m,\ a \in A,\ m \in M$, to represent reservation-based $(m = 1)$ and free-float $(m = 2)$ vehicle movements in the spatial-temporal network. The feasible region of shared vehicle movement is given by $Y(\mathbf{w}, \mathbf{x}^1, \mathbf{x}^2, \mathbf{u})$, consisting of constraints in the corresponding multi-commodity flow problem, with two commodities $m = 1, 2$.

In this paper, we employ Monte Carlo sampling to generate a finite number of demand scenarios from a joint distribution of one-way and round-trip rentals. We index the scenarios by $k \in K$, denote the vector of capacities of the arcs in scenario k by $\mathbf{u}^k = [u_a^k,\ a \in A^{\text{one}} \cup A^{\text{two}}]^{\mathsf{T}}$, and the probability of occurrence of scenario k by p^k. For the objective, we minimize the expected cost of scenario-based vehicle movements $\mathbf{y}^k = [y_a^{km},\ m \in M,\ a \in A]^{\mathsf{T}},\ k \in K$, plus some random penalty incurred from unserved demands. Thus,

$$Q(\mathbf{w}, \mathbf{x}^1, \mathbf{x}^2) = \min_{\mathbf{y}^1,\ldots,\mathbf{y}^{|K|}} \quad \sum_{k\in K} p^k \sum_{a\in A} f_a \sum_{m\in M} y_a^{km} + \sum_{k\in K} p^k \sum_{a\in A^{\text{one}}\cup A^{\text{two}}} G_a(u_a^k - \sum_{m\in M} y_a^{km}) \tag{3}$$

$$\text{s.t.} \quad \mathbf{y}^k \in Y(\mathbf{w}, \mathbf{x}^1, \mathbf{x}^2, \mathbf{u}^k) \qquad \forall k \in K. \tag{4}$$

The first term in the objective function (3) denotes the expected cost of vehicle traveling, idleness, and relocation, while the second denotes the penalty incurred in all the scenarios.

2.2 Rolling-Horizon Model for Vehicle Relocation

For ease of exposition, we consider the reservation-based system, i.e., $M = \{1\}$. Let $s \in \{1, 2, \cdots, T-1\}$ denote the current time period. A spatial-temporal network, denoted $G(V_s, A_s)$, is constructed in the same way as before, except that it only includes periods $s+1, \ldots, T$ and that $\mathbf{w}$ and $\mathbf{x}^1$ are now fixed (obtained in period 0). Let v_{is} be the number of currently idle vehicles in zone i, and v_{it} the number of vehicles that are currently rented out, but will be returned to zone i at time $t > s$. We define decision variables z_{ij} as the number of vehicles that start being relocated from zone i to zone j in the current period. Recall that ℓ_{ij} is the time for relocating a vehicle from i to j. Let c_{ij}^{rel} denote the cost of relocating a vehicle from i to j. For ease of exposition, let z_{ii} denote the number of vehicles that remain idle in zone i, with $\ell_{ii} = 1$ and $c_{ii}^{\text{rel}} = c^{\text{idle}}$. Let $J(i,t) = \{j \in I : \ell_{ji} = t\}$, i.e., the set of zones from which it takes t periods to relocate a vehicle to zone i. The rest of the notation follows those previously defined. We present the vehicle relocation problem in period s as

$$\min_{\mathbf{z}} \quad \sum_{i\in I}\sum_{j\in J} c_{ij}^{\text{rel}} z_{ij} + \sum_{k\in K} p_k Q_k(\mathbf{z}) \tag{5}$$

$$\text{s.t.} \quad \sum_{j\in J} z_{ij} = v_{is},\ z_{ii} \le w_i, \quad \forall i \in I \tag{6}$$

$$z_{ij} \in \mathbb{Z}_+, \quad \forall i,\ j \in I \tag{7}$$

Let $I_s(t) = 1$ if $t = s$ and $I_s(t) = 0$ if $t \neq s$. For each $k \in K$, the second-stage cost $Q_k(\mathbf{z})$ in (5) can be specified as

$$Q_k(\mathbf{z}) = \min_{\mathbf{y}} \quad \sum_{a\in A_s} f_a y_a + \sum_{a\in A_s^{\text{one}}\cup A_s^{\text{two}}} G_a(u_a^k - y_a) \tag{8}$$

$$\text{s.t.} \quad \sum_{a\in\delta^+(n_{it})} y_a - \sum_{a\in\delta^-(n_{it})} y_a = v_{it} + \sum_{j\in J(i,t-s)} z_{ji} - I_T(t) x_i^1, \quad \forall (i,t) \in V_s \tag{9}$$

$$0 \le y_a \le u_a^k, \quad \forall a \in A_s. \tag{10}$$

The one-time relocation model can be applied as follows. First, we solve for a daily initial vehicle assignment plan. Then, every period, e.g., an hour, we fulfill demand as much as possible. At the end of the period, we use the one-time relocation model to optimize vehicle relocation. This process continues until period T when the system restores initial vehicle allocation for the next day.

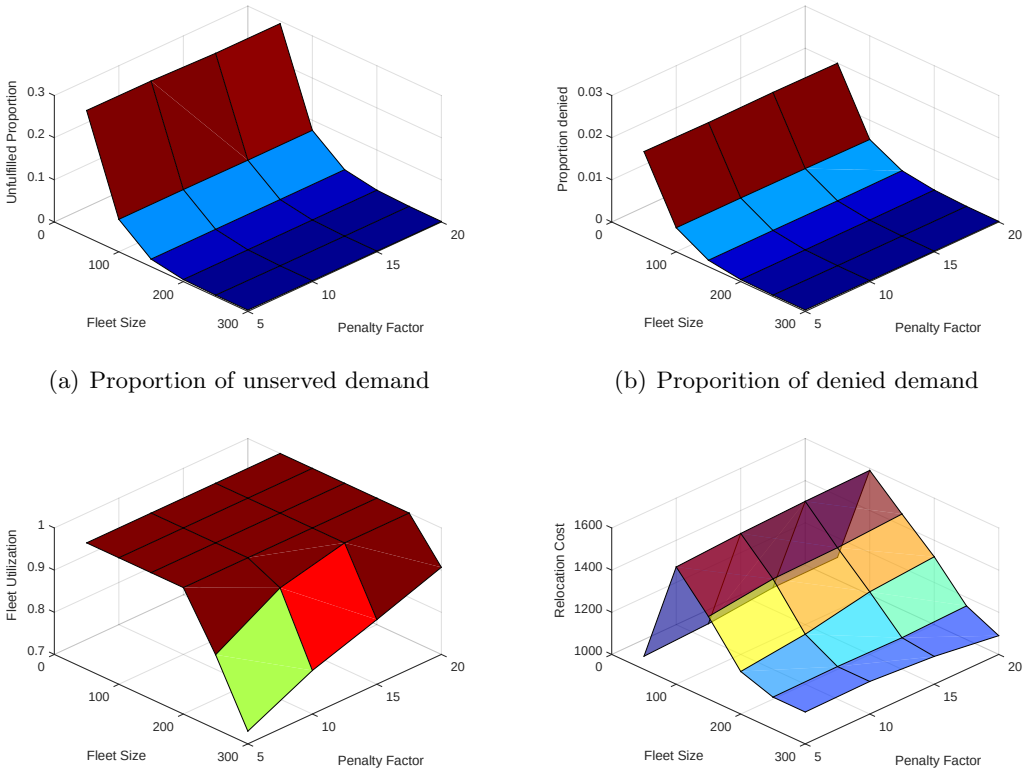
2.3 Mixed-Integer Rounding (MIR)-Enhanced Benders Decomposition

We develop a branch-and-cut algorithm with cuts enhanced by mixed-integer rounding (MIR) [1]. The basic procedure branches on the integer variables and solves individual nodes via Benders decomposition [2]; the Benders cuts at each node are added to the master problems of subsequent nodes, which, however, could be weak due to the relaxed integer constraints in the first-stage. Thus, we apply MIR to pairs of previously generated Benders cuts to obtain stronger valid cuts.

3 Numerical Results

We test instances generated based on Zipcar's rental data collected from the Boston-Cambridge area and conduct computational studies to demonstrate the results of our approach. The data set is from Oct 1 to Dec 1, 2014, containing the information of the starting-ending time, and the zipcodes of the origin and destination zones of each rental. In our sampling-based method, we generate one-way and round-trip rentals for $T = 24$ peirods using Gamma distributions (validated by Kolmogorov-Smirnov test) with means and variances equal to the empirical means and variances of the Zipcar data. The cost/revenue parameters are set according to real practice.

We first study the impact of other important factors such as the size of available fleet, and the penalty cost of lost demand, shown in Figure 1. Then, we analyze how one-way and round-trip de-



(a) Proportion of unserved demand

(b) Proportion of denied demand

(c) Utilization of available fleet

(d) Relocation cost

Figure 1: The impact of fleet size and penalty factor on some system performance measures

mand proportions affect solution profitability and QoS, given fixed prices used in practice, and also examine the profitability and QoS results corresponding to different one-way proportions, endogenously generated based on varying one-way rental prices. In these two cases, a higher proportion of one-way demand may have opposite effects. Thus, understanding the market dynamics of one-way rentals is of great importance for carshare companies as well as local governments that try to solve their urban transportation problems by promoting carsharing. Finally, we illustrate how to implement our two-stage model for dynamic vehicle relocation, and demonstrate the efficacy of MIR-enhanced Benders cuts implemented in parallel, which outperforms the traditional Benders.

References

[1] Bodur, M. and Luedtke, J. R. (2016). Mixed-integer rounding enhanced Benders decomposition for multiclass service system staffing and scheduling with arrival rate uncertainty. Forthcoming in *Management Science*.

[2] Van Slyke, R. M. and Wets, R. (1969). L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17(4):638–663.

Ridesharing in a Mobility-on-Demand System

Samitha Samaranayake
School of Civil and Environmental Engineering,
Cornell University, Ithaca, New York, U.S.A. Email: {samitha}@mit.edu

Kevin Spieser
Uber Technologies, San Francisco CA, U.S.A
Email: kspieser@gmail.com

Harshith Guntha
Department of Electrical and Computer Engineering,
Indian Institute of Technology, Chennai, India
Email: hgunta@gmail.com

Emilio Frazzoli
Laboratory for Information and Decision Systems,
Massachusetts Institute of Technology, Cambridge MA, U.S.A.
Email: {frazzoli}@mit.edu

1 Introduction

Mobility-on-demand (MoD) services have the potential for transporting passengers between locations efficiently and reliably, providing a service comparable to a private car, while removing many of the more burdensome aspects of ownership, e.g., maintenance and parking. Moreover, by collectively time-sharing use of the fleet, significant fixed costs are distributed over a large user base, drastically reducing the cost to access mobility. Although vehicle sharing services offer a number of benefits, they are plagued by the inefficiency of needing to rebalance empty vehicles to ensure the supply of vehicles remains aligned with the demand for transport. Rebalancing has the unfortunate side-effect of increasing the total vehicle mileage driven throughout the system, which raises concerns about worsening congestion on city streets. Recognizing that many vehicles can carry two or more passengers, one way to curb this effect is to incorporate ridesharing and permit vehicles to transport multiple passengers simultaneously. Services such as UberPool, Lyftline, and Via speak to, relatively recent, commercial interest in building a carsharing platform capable of leveraging the efficiency gains of ridesharing.

Despite progress in resolving the core issues of carsharing and ridesharing indepen-

dently, the associated results rarely extend to services that fuse carsharing and ridesharing functionality. For example, much of the ridesharing literature (see Agatz et al. (2012) for a detailed taxonomy) does not apply to carsharing systems because it assumes a private vehicle ownership model or ignores the issue of how to rebalance empty vehicles, Santi et al. (2014). Similarly, much of the carsharing work, while accounting for rebalancing Spieser et al. (2014, 2016), assumes unit capacity vehicles[1]. As an alternative, software solutions that exploit ridesharing opportunities with real-time requests are discussed in Ma et al. (2013). There, a heuristic quality of service rule is used to match and route up to thousands of vehicles that are viable candidates sharing a ride. Throughout, we tacitly assume a door-to-door MoD service does not require riders to switch between vehicles.

This work makes an initial contribution to the space by considering a MoD system, in which a shared fleet of vehicles, each capable of carrying two passengers at a time[2], is used to transport passengers. Inherent to the formulation are two important attributes: (i) the need to rebalance empty vehicles and (ii) the ability to identify lucrative ridesharing corridors by means of trip chaining. Note that although the later functionality is essential to capture ridesharing in its most general form, it is absent from the majority of existing works that, for a variety of reasons, limit the extent to which rides may be shared. We present a mixed-integer linear programming (MILP) formulation of the problem and show how an heuristic (feasible) solution to the problem can be obtained in polynomial-time by independently solving the ride-matching and rebalancing problems. This approximate solution can be used as a intial guess when solving the coupled problem via a MILP solver. Numerical results based on the NYC Yellow Taxi dataset show a performance improvement over the work presented in Santi et al. (2014).

2 A Steady-State Model of Ridesharing in MoD Systems

Although the ridesharing problem is inherently dynamic in nature, we focus on the case of steady-state demands for clarity and conciseness. The results are readily extensible to the dynamic ridesharing problem with an appropriate time expansion of the variables.

The MoD system is described by a network $G = (V, E)$ of fully connected stations. The time required to travel from station i to j is $T_{ij} > 0$ and represents the shortest physical path between the stations. Passengers enter and exit the system at stations, and are transported between stations by a fleet of m self-driving vehicles. Each vehicle is able to transport up to two passengers at a time. Passengers that enter a vehicle may depart the vehicle only upon reaching their destination, preventing transfers which are a known

[1]There is a large literature on these topics that we have not references to in this abstract due to space limitations.

[2]The model can be extended to the case of carrying more passengers, with the number of variables in the optimization problem increasing by a factor of N (the number of stations) per increase in occupancy.

deterrent to the adoption of transit systems (Guo and Wilson, 2011)[3]. Vehicles traveling with zero, one, or two passengers are referred to as empty or rebalancing, single-occupancy, and dual-occupancy vehicles, respectively. Single occupancy vehicles may be diverted to pickup a second passenger. We wish to find the optimal vehicle and passenger flows along each $ij \in E$ for different optimality criteria.

We begin by defining the following variables. x_i^k is the flow rate of empty vehicles traveling from node i to node $k \neq i$. y_i^k is the rate of single-occupancy vehicles transporting a passenger from node i to the passenger's destination node $k \neq i$. $\tilde{y}_i^{jk}$ is the rate of single occupancy vehicles en-route from i to j with a passenger destined for k. $z_i^{k,k'}(t)$ is the rate of dual-occupancy vehicles traveling from node i to node k, with one passenger destined for node $k \neq i$, and the other passenger destined for node $k' \neq i$[4]. λ_i^k is the rate at which passengers arrive at node i destined for node k. The set $\{\lambda_i^k : i,\ k \neq i \in V\}$ is referred to as the travel demand at time t.

Steady state equations. In steady state, neither passengers nor vehicles accumulate within the system. Consequently, for each node i, the rate at which vehicles enter i must equal the rate at which vehicles leave i,

$$\sum_{j \in V \setminus i} \left(x_j^i + y_j^i + \sum_{k \in V \setminus j} (z_j^{ik} + \tilde{y}_j^{ik})\right) = \sum_{j \in V \setminus i} \left(x_i^j + y_i^j + \sum_{k \in V \setminus i} (z_i^{jk} + \tilde{y}_i^{jk})\right) \quad \forall i \in V. \tag{1}$$

Similarly, for each node i, the passengers entering the system, exiting the system, and just passing through must be conserved,

$$\lambda_i^k + \sum_{j \in V \setminus i} (z_j^{ik} + \tilde{y}_j^{ik}) = y_i^k + \sum_{j \in V \setminus \{i,k\}} \tilde{y}_i^{jk} + \sum_{k' \in V \setminus \{i,k\}} \left(z_i^{kk'} + z_i^{k'k}\right) + 2z_i^{kk} \quad \forall i \in V,\ k \in V \setminus i. \tag{2}$$

Finally, as per our earlier restriction, we do not want passengers to be transferred while en-route. With $\kappa = \sum_{j \in V \setminus i} (z_j^{ik} + \tilde{y}_j^{ik})$, this is ensured via the following constraint

$$z_i^{kk} \leq \begin{cases} \kappa + \frac{1}{2}(\lambda_i^k - \kappa) & \text{, if } \lambda_i^k \geq \kappa \\ \lambda_i^k & \text{, otherwise} \end{cases}, \quad \forall i \in V,\ k \in V \setminus i. \tag{3}$$

If $\lambda_i^k \geq \kappa$, then z_i^{kk} is maximized when as many of the λ_i^k demands get paired with the incoming passengers heading to k and the remaining $\lambda_i^k - \kappa$ share rides in pairs. If $\lambda_i^k < \kappa$, then at most λ_i^k vehicles will have two passengers heading to k. Conveniently, (3) may be written as the following pair of inequalities:

$$z_i^{kk} \leq \lambda_i^k, \quad z_i^{kk} \leq \frac{\lambda_i^k}{2} + \frac{1}{2}\kappa. \tag{4}$$

[3]This restriction is added to make the system more practically applicable and can be easily removed.

[4]Note that $k = k'$ is allowed, corresponding to both passengers traveling to the same destination.

Because the number of flows originating from i, i.e., x_i^k, y_i^k, $\tilde{y}_i^{jk}$, and z_i^{jk}, scales as $O(N+N+N^2)$, where $|V|=N$, the total number of variables scales as $O(N^3)$. However, in practice, it is likely that many z_i^{jk} are zero, and that therefore we achieve more favorable scaling[5].

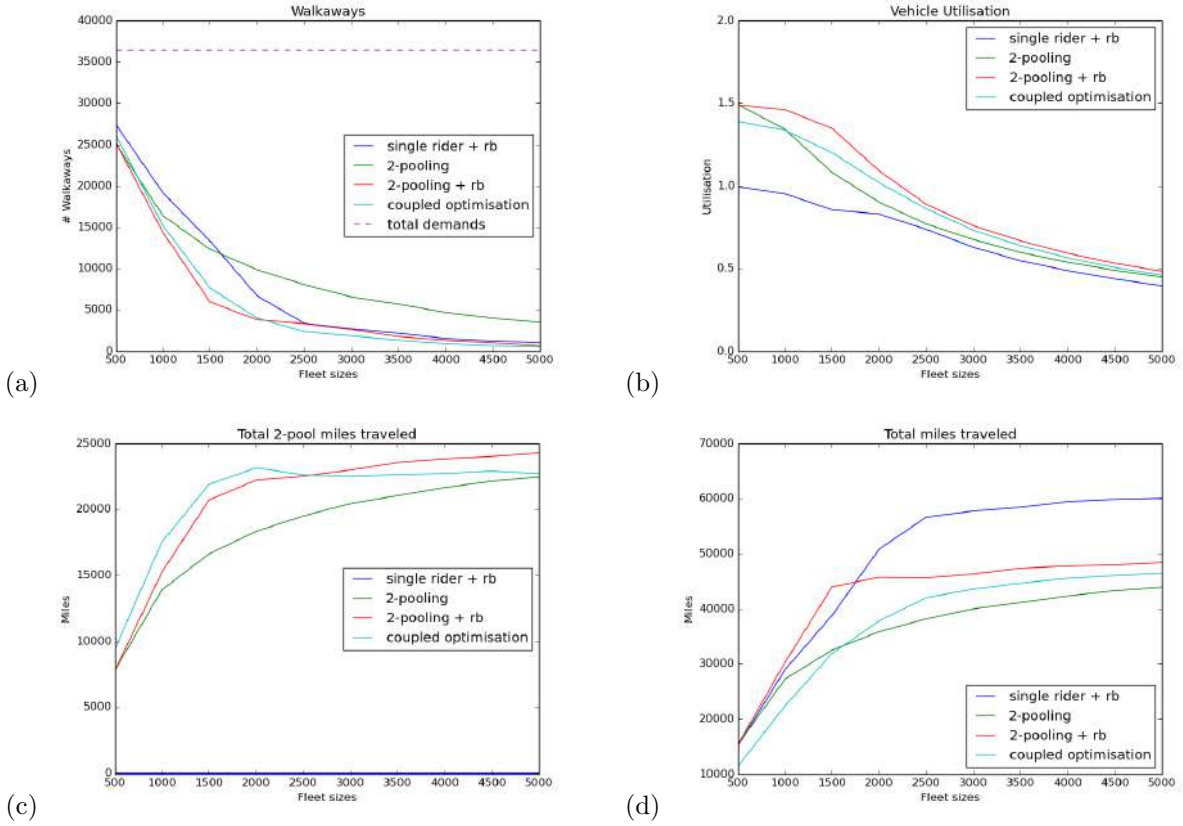


Figure 1: Numerical results for different on-demand taxi models (rb = rebalancing): - (a) The number of customer walkaways as a function of fleet size. A walkaway occurs when a customer is not services within 6 minutes, (b) Vehicle utilization (average number of passengers) as a function of fleet size, (c) Total shared (2-pooled) miles traveled, and (d) Total miles traveled. The coupled optimization minimized the total number of miles traveled and shows that this is achieved without maximizing the number of shared trips.

3 Optimal ridesharing for stationary demands

In the steady-state setting, efficient ridesharing amounts to finding vehicles flows that satisfy (1)–(3) and optimize a sensible performance metric.

A major advantage of ride-sharing is the ability to reduce the fleet size, m, required to service a certain volume of users (assuming users are willing to share). However, there

[5]For example, $z_i^{jk}=0$ when i is in Midtown Manhattan, j is in the Bronx, and k is in Lower Manhattan.

remains a fundamental tradeoff between m and the average time a customer is in transit. An overly aggressive approach to ride-sharing may cause passengers to spend exorbitantly long en-route to their destination. Conversely, customer travel-times can be minimized by assigning each passenger a vehicle. The system operator is responsible for finding the routing policy and fleet size that provides a high quality at a reasonable operational cost.

This problem results in a MILP when using a linear performance metric. Given recent advances in commercial solvers medium scale problems can be directly solved using the MILP formulation. However, for larger problems, this approach may not be feasible. For such cases, we propose solving the matching and rebalancing problems separately as described in (Santi et al., 2014; Pavone et al., 2012) and using the resulting approximate solution as a feasible starting point for a MILP solver. The solver can then be run for a fixed time duration to improve the quality of the polynomial time approximation. Initial numerical evaluations based on the NYC Yellow Taxi dataset provide promising early results (see Figure 1), which we are currently expanding on.

References

N. A. H. Agatz, A. L. Erera, M. W. P. Savelsbergh, and X. Wing. Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, 2012.

Z. Guo and N. H. M. Wilson. Assessing the cost of transfer inconvenience in public transport systems: A case study of the london underground. *Transportation Research Part A: Policy and Practice*, 45(2):91–104, 2011.

S. Ma, Y. Zheng, and O. Wolfson. T-share: A large scale dynamic taxi ridesharing service. *Proceedings of IEEE ICDE*, 2013.

M. Pavone, S.L. Smith, E. Frazzoli, and D. Rus. Robotic load balancing for mobility-on-demand systems. *International Journal of Robotics Research*, 2012.

P. Santi, G. Resta, M. Szell, S. Sobolvesky, S.H. Strogatz, and C. Ratti. Quantifying the benefits of vehicle pooling with shareability networks. *PNAS*, 2014.

K. Spieser, K. Treleaven, R. Zhang, E. Frazzoli, D. Morton, and M. Pavone. *Toward a systematic approach to the design and evaluation of automated mobility-on-demand systems: a case study in Singapore.* Road Vehicle Automation, 2014.

K. Spieser, S. Samaranayake, W. Gruel, and E. Frazzoli. Shared-vehicle mobility-on-demand systems: A fleet operator's guide to rebalancing empty vehicles. *Proceedings of the 95th Annual Meeting of the Transportation Research Board*, 2016.

A Discrete Simulation-Based Optimization Algorithm for Two-Way Car-Sharing Network Design

Tianli Zhou* [1], Carolina Osorio†[1,2], and Evan Fields‡[2]

[1]Department of Civil and Environmental Engineering, Massachusetts Institute of Technology
[2]Operations Research Center, Massachusetts Institute of Technology

January 4, 2017

Car-sharing has become a popular transportation mode in urban areas in the past decades. It has been deployed in about 1,100 cities in 26 nations, with at least 31,665 vehicles and 1,251,504 members in 2010 worldwide (Shaheen and Cohen, 2013). Certain studies indicate that car-sharing can reduce greenhouse gas emissions and the vehicle miles traveled, thereby benefiting both the environment and traffic congestion (Martin et al., 2010; Martin and Shaheen, 2011). Car-sharing can be viewed as a supplement to public transportation service in urban areas. It can be used to solve the first-mile or the last-mile problem.

Two-way car-sharing service, which requires customers to return the car to the same place where they picked it up, is one of the most common forms in today's car-sharing market. In different places in a city, a car-sharing company rents several parking spaces and assigns its fleet to these spaces. It is important for the car-sharing company to decide where to rent the spaces and how many spaces to rent in each area of the city for a certain period of time. From the company's perspective, we are looking to design a system to increase its profit; from the public perspective, we would like to design a system such that residents can have close access (from their home or workplace) to either a car-sharing station, a bike-sharing station, or a public transit stop. The goal of this work is to design a car-sharing network to complement the other existing mobility services (bike-sharing and transit).

In this work, we propose a simulation-based optimization (SO) metamodel algorithm to assign a two-way car-sharing fleet. We use a simulator that samples from real data from a car-sharing company. We evaluate the performance of our proposed algorithm with a Boston case study. We compare its performance to analytical optimization algorithms, and to traditional discrete simulation-based optimization methods. In the following section, we first discuss our motivations. Then we present our methodology, including the simulator and the metamodel SO algorithm. At the end of this abstract, we will briefly describe the experiments carried out to validate the algorithm.

1 Motivation

To use a two-way car-sharing service, a user should first find a suitable car-sharing station and reserve an available car in advance. For the reservation, the user needs to specify the pick-up time and how many hours he wants to use this car. At the reservation start time, the user should go to the location, use the car and return the car to the same place on time. If the user finds that the cars at the desired location are not available for the desired using period, he can choose another start location, change start time, change service length, or choose not to make a reservation. This procedure can be described as "first-come-first-reserve".

*tzhou90@mit.edu
†osorioc@mit.edu
‡efields@mit.edu

Car-sharing companies may keep a record of each reservation, including the reservation booking time, car pick-up location, reservation start time, booked length, and actual return time. Sometimes they also record each change made by the user to a reservation.

To find good fleet assigning strategy, the most common method is using mixed integer programming (MIP) models to approximate the operation of a car-sharing system. Chen (2016) proposed a two-stage stochastic programming model to decide how to allocate a two-way car-sharing fleet and how many free-float fleet parking permits to purchase. That work featured a case study that divided the study area into 9 smaller regions, each of which corresponds to two first-stage decision variables. The second-stage problem of this study contains 100 demand scenarios. Boyacı et al. (2015) proposed a multi-objective MIP model to plan a one-way electric vehicle-sharing system. The case study in that work contains 210 regions. The demand was assumed to be fixed. By considering a vehicle-sharing system as a supplement of public transportation, Nair and Miller-Hooks (2014) used an MIP model to determine the location and capacity of vehicle-sharing stations so it could be better connected to an existing transit network. This study also assumed fixed demand. OMahony (2015) used a continuous-time Markov process to model a bike-sharing system. The author developed a very fast greedy algorithm to solve the problem to optimality. Detailed review about vehicle-sharing systems can be found in Jorge and Correia (2013) and Brandstätter et al. (2016).

In our case, we consider the uncertainty of reservation's pick-up location, start time and length. The analytical MIP formulation of our considered problem leads to a high-dimensional problem. Another approach is to model the interactions between demand and supply with a stochastic simulator. There is currently a lack of efficient SO algorithms for high-dimensional discrete problems. Industrial Strength COMPASS (ISC) becomes slow when problem dimension becomes larger than 10 (Xu et al., 2010). Adaptive Hyperbox Algorithm (AHA) has higher efficiency than COMPASS, and ISC-AHA (AHA combined with ISC) can handle problems with 100 discrete decision variables (Xu et al., 2013). Wang et al. (2013) proposed R-SPLINE, a method using the idea of "phantom gradients". It converges to a local optimal solution faster than AHA in some examples with less than 20 decision variables. According to the authors, R-SPLINE is better at finding good solutions locally than globally, while ISC and ISC-AHA have a global phase before searching for local optimal solution. A more detailed recent review of discrete SO algorithms can be found in Hong et al. (2015). This paper proposes a discrete SO algorithm suitable to address the high-dimensional car-sharing fleet assignment problem.

2 Problem Description

Assume that a city can be divided into smaller pieces, each piece is called a *cell*. Also, assume that each cell will generate customer demand for the two-way car-sharing service. Some cells are candidate locations for car-sharing stations.

We have developed a data-driven microscopic simulator for a two-way car-sharing system based on real data from a car-sharing company. Given the total demand (reservation-hours wanted) of each cell and the number of cars assigned to each cell, the simulator samples empirical reservations from the data and assigns a reservation to a selected station. If no car is available at the desired cell and time, the simulator will assign this reservation to a nearby cell, give it a different start time, or reject this reservation. This simulator has been validated using the data from the car-sharing company.

The simulator can be used to evaluate various objective functions. In this work, we want to maximize the profit of the company.

Important notations to describe the fleet assignment problem are as follows:

$\mathcal{I}$	set of all customer generating cell;
$\mathcal{I}^*$	set of cells that have other modes available, and $\mathcal{I}^* \subseteq \mathcal{I}$;
$\mathcal{I}_i$	set of cells that is considered "near" the cell $i \in \mathcal{I}$;
N^i	capacity of cell $i \in \mathcal{I}$;
X	total number of cars the company can assign;
x_i	decision variable, number of cars we put at the station in cell $i \in \mathcal{I}$.

The feasible region of the fleet assignment problem is defined as follows:

$$\sum_{i \in \mathcal{I}} x_i \leq X, \tag{1}$$

$$x_i \leq N^i, \qquad \forall i \in \mathcal{I}, \tag{2}$$

$$\sum_{j \in \mathcal{I}_i} x_j \geq 1, \qquad \forall i \in \mathcal{I} \setminus \mathcal{I}^*, \tag{3}$$

$$x_i \in \mathbb{Z}_+, \qquad \forall i \in \mathcal{I}. \tag{4}$$

Equation (1) means that the total number of cars that can be assigned should be smaller than the upper bound. Equation (2) means that the number of cars at each station should be smaller than the capacity of that station. Equation (3) says that for each cell, there must be at least one car nearby, if there is no transit (train or key bus route) stops or bike sharing stations nearby. This requirement is aimed at ensuring that residents of the city have access to at least one of the following services: car sharing, bike sharing, train and key bus route. Equations (1) to (4) specify the feasible region of this problem. For the rest of this abstract, we use $\mathcal{F}$ to denote the feasible region.

3 Methodology

We propose to use a metamodel SO method to solve this problem. Computationally efficient metamodel formulations for continuous SO problems have been proposed in Osorio and Nanduri (2015a), Osorio and Nanduri (2015b), Osorio and Chong (2015) and Osorio and Selvam (Forthcoming), which stems from the general metamodel idea of Osorio and Bierlaire (2013). Here we formulate metamodel ideas for discrete SO problems. This is the first algorithm that is designed to be efficient for discrete SO transportation problems (i.e., coming up with good solutions with few simulation evaluations). To achieve this, this paper formulates an analytical approximation of the simulation-based objective function. This approximation is then used as part of a metamodel (as in Osorio and Bierlaire (2013)), and is embedded within a general purpose discrete SO algorithm.

The metamodel is an analytical approximation of the objective function - profit. Important notations for the metamodel are

$\mathcal{L}$	set of all possible reservation lengths;
$\mathcal{T}$	set of possible service start times, $\mathcal{T} = \{1, 2, \ldots, t_{\max}\}$;
r_l	revenue for the company brought by a reservation with length l;
c_i	cost of having a space in cell i over the planing period, including space renting fee and other costs;
d^i_{tl}	number of customers who are from cell $i \in \mathcal{I}$ and want to start service with length $l \in \mathcal{L}$ at time $t \in \mathcal{T}$, estimated by using simulator.

We have two types of auxiliary variables:

z^i_{tl}	number of customers whose start time is $t \in \mathcal{T}$, service length is $l \in \mathcal{L}$ and will actually be served by the station in cell $i \in \mathcal{I}$;
z^{ji}_{tl}	number of customers from cell $j \in \mathcal{I}$, with start time $t \in \mathcal{T}$ and service length $l \in \mathcal{L}$ and will actually be served by a nearby station $i \in \mathcal{I}_j$

The objective of the problem is to maximize the profit of the company. For the metamodel, the analytical approximation of the simulation-based objective function has the following form:

$$T(x) = \sum_{i\in\mathcal{I}} \sum_{t\in\mathcal{T}} \sum_{l\in\mathcal{L}} r_l z^i_{tl} - \sum_{i\in\mathcal{I}} c_i x_i, \tag{5}$$

which equals the total revenue minus the total cost.

We use the following constraints to describe the system:

$$\sum_{j\in\mathcal{I}_i} z^{ji}_{tl} = z^i_{tl}, \qquad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \forall l \in \mathcal{L}, \tag{6}$$

$$\sum_{j\in\mathcal{I}_i} z^{ij}_{tl} \leq d^i_{tl}, \qquad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \forall l \in \mathcal{L}, \tag{7}$$

$$\sum_{l\in\mathcal{L}} z^i_{tl} + \sum_{l\in\mathcal{L}} \sum_{t'\in\mathcal{T}_1(t,l)} z^i_{t'l} \leq x_i, \qquad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \tag{8}$$

$$x \in \mathcal{F}, \tag{9}$$

$$z^i_{tl} \in \mathbb{R}_+, \qquad \forall i \in \mathcal{I}, \forall t \in \mathcal{T}, \forall l \in \mathcal{L}, \tag{10}$$

$$z^{ij}_{tl} \in \mathbb{R}_+, \qquad \forall i \in \mathcal{I}, \forall j \in \mathcal{I}_i, \forall t \in \mathcal{T}, \forall l \in \mathcal{L}, \tag{11}$$

where $\mathcal{T}_1(t,l) = \{t' \in \mathcal{T} : t'+1 \leq t \leq t'+l-1\}$ is the set of service start time points for services that start before t and have not finished at t for a given service length value l. Equations (6) and (7) are customer flow conservation constraints. Equation (8) ensures that at each station, the sum of the number of cars that are being used at this time point and the number of cars that will start to be used at this time point is smaller than the number of cars that belong to the station. Equation (9) states that the fleet assignment should be feasible. This analytical model is a simplification of the simulator. For instance, it differs from the simulator in that it does not enforce the first-come-first-reserve rule. It also does not allow customers to change the start time or the duration of their reservation. These simplifications are made such as to have a tractable analytical model, which will be then embedded within the SO algorithm as part of a metamodel.

We embed the proposed metamodel formulation within an existing discrete SO algorithm. We use the ISC-AHA of Xu et al. (2013).

We evaluate the performance of our proposed algorithm with both a toy case study and a Boston case study. We compare its performance to analytical optimization algorithms, and to traditional discrete simulation-based optimization. We compare the algorithms both in terms of the quality of their proposed solutions and their computational efficiency. This is part of ongoing research. The most recent results will be presented at the conference.

Acknowledgment

This abstract is part of the project "Car-sharing services: optimal network expansions to integrate automotive with mass-transit systems and to mitigate congestion at the metropolitan scale" of the Ford-MIT Alliance. The authors thank the project members, including Perry MacNeille, Paul Beer and Arvind Kannan.

References

Burak Boyacı, Konstantinos G Zografos, and Nikolas Geroliminis. An optimization framework for the development of efficient one-way car-sharing systems. *European Journal of Operational Research*, 240(3): 718–733, 2015.

Georg Brandstätter, Claudio Gambella, Markus Leitner, Enrico Malaguti, Filippo Masini, Jakob Puchinger, Mario Ruthmair, and Daniele Vigo. Overview of optimization problems in electric car-sharing system de-

sign and management. In *Dynamic Perspectives on Managerial Decision Making*, pages 441–471. Springer, 2016.

Zhihao Chen. *Strategic Network Planning Under Uncertainty with Two-Stage Stochastic Integer Programming*. PhD thesis, University of Michigan, 2016.

L Jeff Hong, Barry L Nelson, and Jie Xu. Discrete optimization via simulation. In *Handbook of Simulation Optimization*, pages 9–44. Springer, 2015.

Diana Jorge and Gonçalo Correia. Carsharing systems demand estimation and defined operations: a literature review. *European Journal of Transport and Infrastructure Research*, 13(3):201–220, 2013.

Elliot Martin, Susan A Shaheen, and Jeffrey Lidicker. Carsharing's impact on household vehicle holdings: Results from a north american shared-use vehicle survey. institute of transportation studies, university of california. Technical report, Davis. Research Report UCD-ITSRR-10-05, 2010.

Elliot W Martin and Susan A Shaheen. Greenhouse gas emission impacts of carsharing in north america. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1074–1086, 2011.

Rahul Nair and Elise Miller-Hooks. Equilibrium network design of shared-vehicle systems. *European Journal of Operational Research*, 235(1):47–61, 2014.

Carolina Osorio and Michel Bierlaire. A simulation-based optimization framework for urban transportation problems. *Operations Research*, 61(6):1333–1345, 2013.

Carolina Osorio and Linsen Chong. A computationally efficient simulation-based optimization algorithm for large-scale urban transportation problems. *Transportation Science*, 49(3):623–636, 2015.

Carolina Osorio and Kanchana Nanduri. Energy-efficient urban traffic management: a microscopic simulation-based approach. *Transportation Science*, 49(3):637–651, 2015a.

Carolina Osorio and Kanchana Nanduri. Urban transportation emissions mitigation: Coupling high-resolution vehicular emissions and traffic models for traffic signal optimization. *Transportation Research Part B: Methodological*, 81:520–538, 2015b.

Carolina Osorio and Krishna K Selvam. Simulation-based optimization: achieving computational efficiency through the use of multiple simulators. *Transportation Science*, Forthcoming.

Eoin Daniel OMahony. *Smarter Tools For (Citi) Bike Sharing*. PhD thesis, Cornell University, 2015.

Susan A Shaheen and Adam P Cohen. Carsharing and personal vehicle services: worldwide market developments and emerging trends. *International Journal of Sustainable Transportation*, 7(1):5–34, 2013.

Honggang Wang, Raghu Pasupathy, and Bruce W Schmeiser. Integer-ordered simulation optimization using r-spline: Retrospective search with piecewise-linear interpolation and neighborhood enumeration. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 23(3):17, 2013.

Jie Xu, Barry L Nelson, and L Jeff Hong. Industrial strength compass: A comprehensive algorithm and software for optimization via simulation. *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 20(1):3, 2010.

Jie Xu, Barry L Nelson, and L Jeff Hong. An adaptive hyperbox algorithm for high-dimensional discrete optimization via simulation problems. *INFORMS Journal on Computing*, 25(1):133–146, 2013.

Traffic & Mobility

FA2: Traffic Management

Friday 8:30 – 10:30 AM

Session Chair: Lili Du

8:30 **Morning Commute Problem with Queue-Length-Dependent Bottleneck Capacity**
[1]Jin-Yong Chen, [2]Rui Jiang, [2]Xin-Gang Li, [1]Mao-Bin Hu, [2]Bin Jia*
[1]University of Science and Technology of China, [2]Beijing Jiaotong University

9:00 **Modelling User Behavior at a Stochastic Bottleneck**
Daphne van Leeuwen, Peter van de Ven*
CWI

9:30 **Recasting Intersection Automation as a Connected-and-Automated-Vehicle (CAV) Scheduling Problem within Heterogeneous Traffic Environment**
[1]Pengfei (Taylor) Li, [2]Xuesong Zhou*
[1]Mississippi State University, [2]Arizona State University

10:00 **Distributed Computation Based Constrained Model Predictive Control for a Mixed Flow Platoon**
Siyuan Gong, Lili Du*
Illinois Institute of Technology

Morning commute problem with queue-length-dependent bottleneck capacity

Jin-Yong Chen[a], Rui Jiang[b], Xin-Gang Li[b], Mao-Bin Hu[a], Bin Jia[b]
[a] School of Engineering Science, University of Science and Technology of China, Hefei 230026, China
[b] MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University, Beijing 100044, China

Vickrey's bottleneck model is a classical model used to study morning commute problem. Since the model can be investigated analytically, it has been used to study implications of transportation policies, such as congestion pricing, tradable mobility credits, flexible work schedule, parking management, and variable speed limit.

In the Vickrey model, a fixed number (denoted by N) of commuters go from home to work through a road with a potential bottleneck. Once the bottleneck is activated, a queue will emerge and, thus, commuters incur a queuing delay. Commuters also incur an early/late cost if they arrive at work earlier/later than desired arrival times. The cost of a commuter is calculated by

$$c(t) = \alpha T(t) + \beta \max\{0, t^* - t - T(t)\} + \gamma \max\{0, t + T(t) - t^*\}. \quad (1)$$

Here T(t) is the travel time for a departure time t, t* is preferred arrival time, α is the value of unit travel time, and β and γ are the schedule penalties for a unit time of early arrival and late arrival respectively. As usual, we assume that γ>α>β>0. Each individual chooses departure time to minimize his/her commute cost that combines queuing delay and early/late cost, and a unique user equilibrium (UE) state can be achieved.

In the tolled highways, usually more tollgates will be activated to alleviate traffic congestion in rush hours. Via considering a queue-length-dependent bottleneck capacity, this paper extends Vickrey model to depict the situation. It is shown that diverse situations can be identified. In particular, multiple user equilibrium states and oscillation of queue length have been observed.

In our model, bottleneck capacity is assumed to increase from s_1 to s_2 as the queue length exceeds a threshold D_1, and decreases back to s_1 as the queue length reduces and becomes smaller than another threshold $D_2 < D_1$.

Other symbols used are listed as follows

Symbols	definition
k_D	$k_D = D_1/D_2 > 1$
k_s	$k_s = s_2/s_1 > 1$
δ	$\delta = \beta\gamma/(\beta+\gamma)$
t_q	The earliest departure time
$t_{q'}$	The latest departure time
t_1	The time that the queue length increases to D_1
t_2	The time that the queue length decreases to D_2 ($t_2 > t_1$)
$\tilde{t}_1$	The departure time corresponding to arrival time t_1
$\tilde{t}_2$	The departure time corresponding to arrival time t_2
$\tilde{t}$	The departure time corresponding to arrival time t*

Firstly, we consider the situation that capacity does not change or changes (defined as increasing to s_2 and backing to s_1) only once. Under this situation, three scenarios are classified.

Scenario (I): $k_D \geq \frac{\alpha+\gamma}{\alpha} k_s$

In this scenario, three UE states can be observed, as shown in Fig.1. In case 1, bottleneck capacity does not change. The model thus reduces to a standard Vickrey model. The cost function is

$$p_1(N) = \delta \frac{N}{s_1}, \quad (2)$$

and this case can be observed if $N \leq \frac{\alpha}{\delta} D_1$ .In case 2, bottleneck capacity changes. The cost function

is

$$p_2(N) = \alpha \frac{D_1}{s_2} + \frac{\delta}{s_2}\left(N - \tilde{N}_{\min}\right), \tag{3}$$

and this case can be observed if $N \geq \tilde{N}_{\min}$. Here

$$\tilde{N}_{\min} = \frac{\alpha}{\delta} D_1 - \frac{k_s - 1}{k_s}\left(\frac{\alpha - \beta}{\beta} k_D + \frac{\alpha + \gamma}{\gamma} k_s\right) D_2 . \tag{4}$$

Case 3 can be observed if $k_s > \frac{\alpha + \gamma}{\alpha}$ and $\tilde{N}_{\min} \leq N < \tilde{N}_{\max}$. Here

$$\tilde{N}_{\max} = \frac{\alpha}{\delta} D_1 + \frac{\alpha + \gamma}{\gamma} \frac{k_s - 1}{k_s}\left(k_D - k_s\right) D_2 . \tag{5}$$

The cost function is

$$p_3(N) = \alpha \frac{D_1}{s_2} + \frac{\delta}{s_1}\left(N - \tilde{N}_{\min}\right). \tag{6}$$

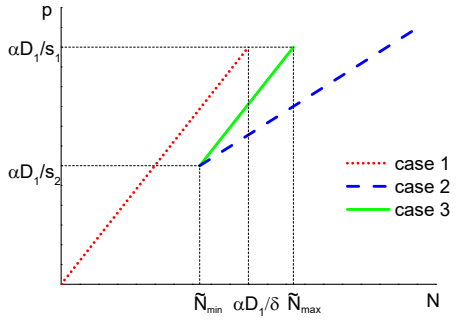


Fig. 1. Cost function in the three cases in scenario (I)

Scenario (II): $\frac{\alpha - \beta}{\alpha} k_s < k_D < \frac{\alpha + \gamma}{\alpha} k_s$

In this scenario, case 1 can also be observed as in Scenario (I). However, when $N \geq YD_2$, a different case 4 can be observed, and the cost function is

$$p_4(N) = \delta\left(\frac{\alpha}{\beta}\frac{D_1}{s_2} + \frac{\alpha + \gamma}{\gamma}\frac{D_2}{s_1}\right) + \frac{\delta}{s_2}\left(N - YD_2\right), \tag{7}$$

Here

$$Y = \frac{\alpha - \beta}{\beta}\frac{k_D}{k_s} + k_D + \frac{\alpha + \gamma}{\gamma}, \tag{8}$$

If $k_s \neq \frac{\alpha - \beta}{\alpha - 2\beta}$ and N is between $\tilde{N}$ and YD_2, another different case 5 can be observed, and the cost function is

$$p_5(N) = \delta\left(\frac{\alpha}{\beta}\frac{D_1}{s_2} + \frac{\alpha + \gamma}{\gamma}\frac{D_2}{s_1}\right) + \frac{\delta}{s_1}\frac{k_s - \frac{\alpha}{\alpha - \beta}}{\frac{\alpha - 2\beta}{\alpha - \beta} k_s - 1}\left(N - YD_2\right). \tag{9}$$

Here

$$\tilde{N}=\left[Y+\frac{\alpha-\beta}{\beta}\left(\frac{\alpha-2\beta}{\alpha-\beta}-\frac{1}{k_s}\right)\frac{k_s-\frac{\alpha}{\alpha+\gamma}k_D}{k_s-\frac{\alpha}{\alpha+\gamma}}\right]D_2 . \tag{10}$$

Note that when $\frac{1}{k_s}>\frac{\alpha-2\beta}{\alpha-\beta}$, $\tilde{N}<YD_2$ (Fig.2a-c); when $k_s>\frac{\alpha-\beta}{\alpha-2\beta}>1$, $\tilde{N}>YD_2$ (Fig.2d). Also note that the slope of the cost function in case 5 could be positive (Fig.2a and d), negative (Fig.2c) or equal zero (Fig.2b). The negative and zero slope is counterintuitive, in which increasing the total number of commuters does not change or even decreases the average cost of each commuter.

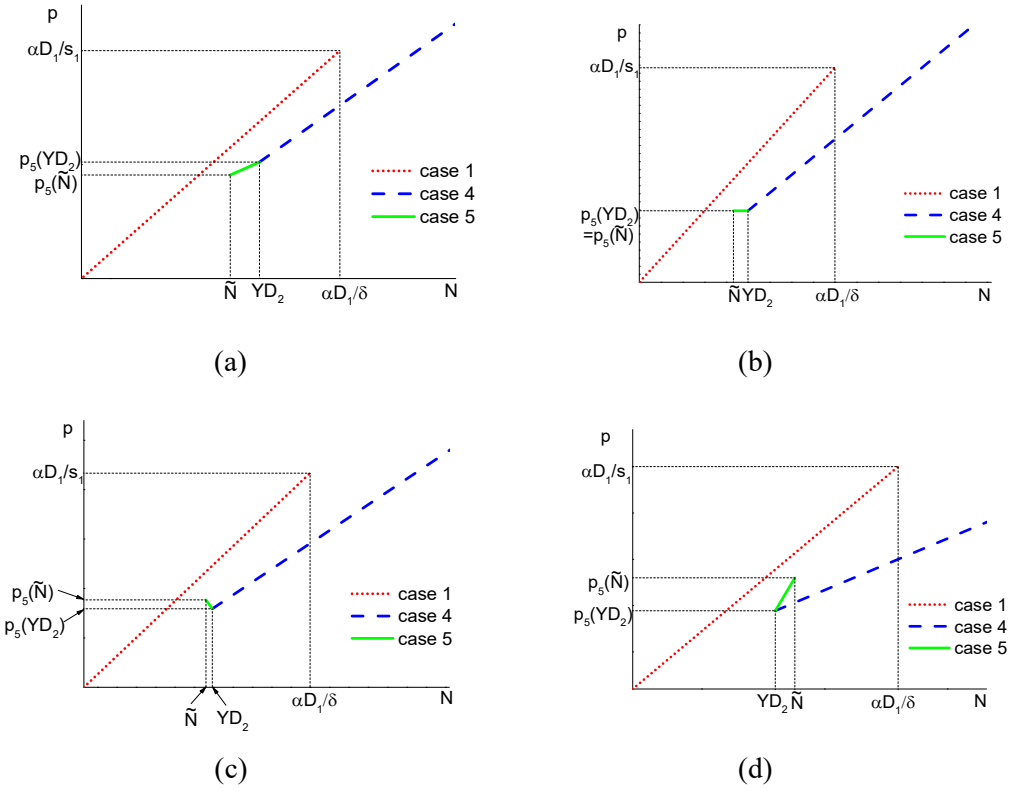


Fig.2. Cost function in the three cases 1, 4, 5 in scenario (II). (a) $k_s<\frac{\alpha}{\alpha-\beta}$, (b) $k_s=\frac{\alpha}{\alpha-\beta}$, (c) $k_s>\frac{\alpha}{\alpha-\beta}$ and $\frac{1}{k_s}>\frac{\alpha-2\beta}{\alpha-\beta}$, (d) $\frac{1}{k_s}<\frac{\alpha-2\beta}{\alpha-\beta}$

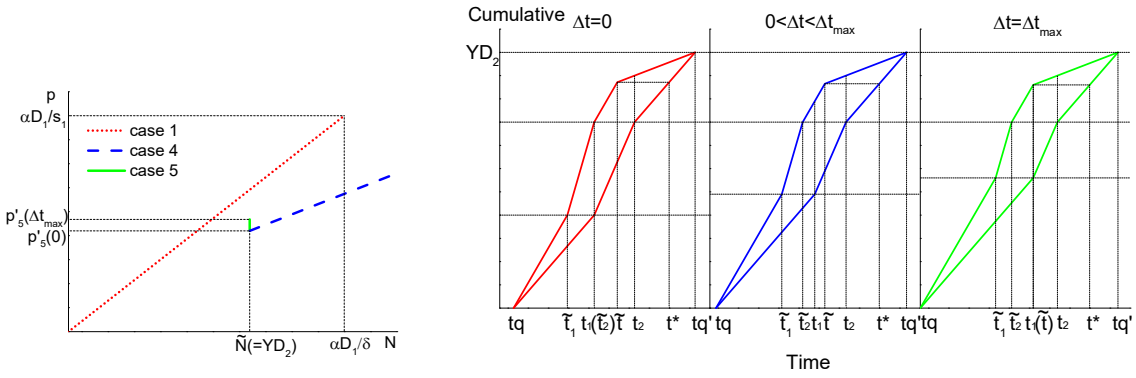
In the special case $k_s=\frac{\alpha-\beta}{\alpha-2\beta}$, $\tilde{N}=YD_2$. Under the circumstance, if $N=YD_2$, case 5' can be observed in which there exist infinite UE solutions. Let us introduce $\Delta t=t_1-\tilde{t}_2$, then the cost function is

$$p_5'(\Delta t)=\delta\left(\frac{\alpha}{\beta}\frac{D_1}{s_2}+\frac{\alpha+\gamma}{\gamma}\frac{D_2}{s_1}\right)+\delta\frac{\alpha}{\beta}\left(1-\frac{\alpha}{\alpha-\beta}\frac{1}{k_s}\right)\Delta t , \tag{11}$$

where $0\le\Delta t\le\Delta t_{\max}$, and

$$\Delta t_{\max} = \frac{\alpha-\beta}{\alpha} \frac{k_s - \dfrac{\alpha}{\alpha+\gamma} k_D}{k_s - \dfrac{\alpha}{\alpha+\gamma}} \frac{D_2}{s_1}, \tag{12}$$

see Fig.3(a). Fig.3(b) shows three different UE solutions. One can see that the departure curve is the same for $t > \tilde{t}_2$ and the arrival curve is the same for $t > t_2$.



(a) (b)

Fig.3. (a) Cost function in the three cases 1, 4, 5' in scenario (II); (b) Three different UE solutions in case 5'.

Scenario (III): $k_D \le \dfrac{\alpha-\beta}{\alpha} k_s$

In this scenario, apart from case 1, there are five different cases. The existence condition and the cost function of Case 6-9 are listed below in Table 1, where

$$Y_1 = Y + \frac{\alpha}{\beta}\left(\frac{\alpha-\beta}{\alpha} k_s - k_D\right), \tag{13}$$

$$Y_2 = Y + \frac{\alpha-\beta}{\beta}\left(\frac{\alpha-2\beta}{\alpha-\beta} - \frac{1}{k_s}\right) \frac{k_s - \dfrac{\alpha}{\alpha-\beta} k_D}{k_s - \dfrac{\alpha}{\alpha-\beta}}, \tag{14}$$

and $Y_1 > Y_2$ always holds. Fig.4 shows the cost function of case 1 and case 6-9.

Table 1 The existence condition and the cost function of Case 6-9

Case	existence condition	cost function
6	$N \ge Y_1 D_2$	$p_6(N) = \alpha \dfrac{D_2}{s_1} + \dfrac{\delta}{s_2}(N - Y_1 D_2)$
7	$k_s \ne \dfrac{\alpha-\beta}{\alpha-2\beta}$ and N is between $\tilde{N}$ and $Y_2 D_2$	$p_7(N) = \alpha \dfrac{D_2}{s_1} + \dfrac{\delta}{s_1} \dfrac{k_s - \dfrac{\alpha}{\alpha-\beta}}{\dfrac{\alpha-2\beta}{\alpha-\beta} k_s - 1}(N - Y_2 D_2)$
8	$Y_1 D_2 \le N < Y_1 D_2 + \dfrac{\alpha}{\delta}(D_1 - D_2)$	$p_8(N) = \alpha \dfrac{D_2}{s_1} + \dfrac{\delta}{s_1}(N - Y_1 D_2)$
9	$Y_2 D_2 \le N < Y_2 D_2 + \dfrac{\alpha}{\delta}(D_1 - D_2)$	$p_9(N) = \alpha \dfrac{D_2}{s_1} + \dfrac{\delta}{s_1}(N - Y_2 D_2)$

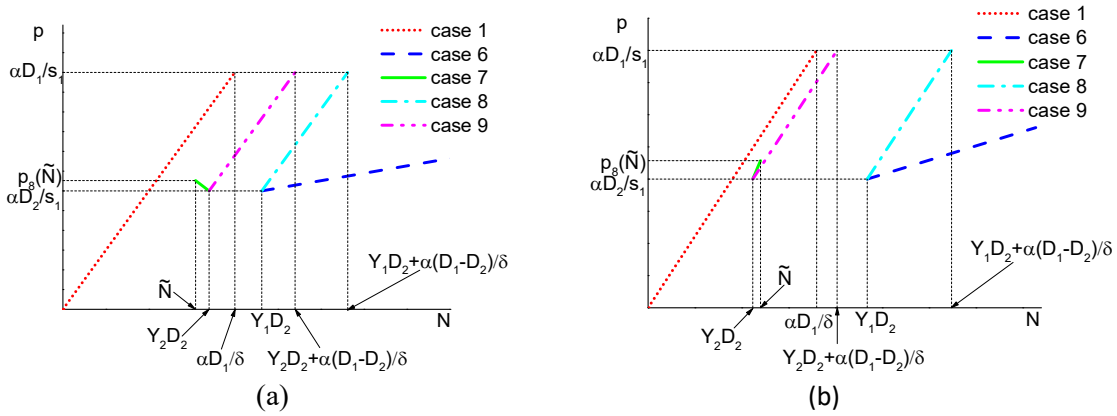


Fig.4. Cost function in case 1 and case 6-9 in scenario (III). (a) $\frac{1}{k_s} > \frac{\alpha-2\beta}{\alpha-\beta}$; (b) $\frac{1}{k_s} < \frac{\alpha-2\beta}{\alpha-\beta}$

Similar to scenario (II), in the special case $k_s = \frac{\alpha-\beta}{\alpha-2\beta}$, $\tilde{N} = Y_2D_2$. Under the circumstance, if $N=Y_2D_2$, case 7' can be observed in which there exist infinite UE solutions.

Capacity changes more than once

Fully discussing the situation that capacity changes more than one time is very complex. As an example, Fig.5 shows an equilibrium solution in which $\frac{2}{\frac{\alpha-\beta}{\alpha}k_s-1}+\frac{\alpha}{\beta} \le k_D < \frac{3}{\frac{\alpha-\beta}{\alpha}k_s-1}+\frac{2\alpha}{\beta}$, and $N = M_2D_2$. Here $M_2 = \frac{\alpha}{\delta}+2\frac{k_s-1}{k_s-\frac{\alpha}{\alpha-\beta}}\left(k_D-1\right)$. In the example, capacity changes twice. Oscillation of queue length has been observed.

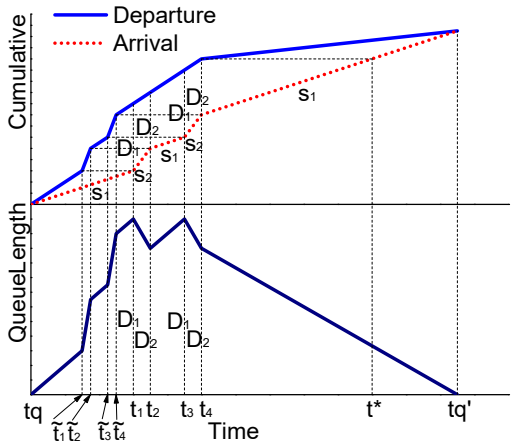


Fig.5. An example that capacity changes twice.

Modelling user behaviour at a stochastic bottleneck

D. van Leeuwen, CWI, The Netherlands, D.vanleeuwen@cwi.nl
P.M. van de Ven, CWI, The Netherlands, ven@cwi.nl

December 9, 2016

1 Introduction

Bottlenecks are common in road traffic networks: Due to high fluctuations in travel demand over time and space, capacities at parts of the network are temporarily exceeded, resulting in congestion. Predictions on daily arrival patterns are used to forecast the delays in the near-future. Often, these methods are deterministic and only consider average delay. However, traffic congestion is a non-stationary process which is predictable to a certain degree. Variations in capacity and demand often lead to unpredictable delays. A better understanding of these bottlenecks and user behaviour is the first step towards designing better road networks.

A well-known model capturing travel delay and user behaviour at a bottleneck is due to [1], later adjusted by [2]. In this deterministic setting the Wardrop equilibrium can be computed, yielding an arrival curve into this queue such that all travellers encounter the same costs and no traveller can decrease his/her costs by altering their departure time. This bottleneck model and its variants have been studied extensively in econometric and transportation literature, for examples see [2][3]. Although these fluid queues are mathematically tractable, they may not provide an accurate representation for the behaviour at a bottleneck, in particular for local roads where a bottleneck would be comprised of relatively few travellers.

To study the effects of variability, we extend this model using the formulation of [2]. In this extension we model the arrivals and departures in the system by a stochastic process. In this model we compute the arrival and departure rates with the same input parameters and approximate the equilibrium for the stochastic model. We also determine the difference in costs if we would use the original deterministic bottleneck models output values and show the effects for our extended model including stochastics in arrival and departure. Finally, we

quantify the impact in time-dependent arrival rate, given that we quantify the costs of uncertainty.

It is worth pointing out that this work is part of a larger trend towards modelling uncertainty by using the bottleneck model. Some researchers argue that variability is of great importance to understand user behaviour. For example Noland (1996) points out: "The inability to accurately predict travel times may be more costly to travellers than the extra travel time due to congested traffic" [4]. However, most variants of the fluid bottleneck model that include stochasticity effects make use of exogenous variability. Such as in [5] and [6], where a probability of extra travel time due to an incident creates variability. In our study we investigate the impact of endogenous effects, where we include the stochasticity effects of arrival behaviour.

2 Model

We consider a queueing model with discrete customers for which we determine the arrival rate given Poisson distributed arrival instant and exponentially distributed service durations, to more accurately capture the arrival pattern at a bottleneck. In particular, we model this bottleneck by a transient analysis of the $M_t/M/1$ queue, and aim to quantify and correct for the error made by the deterministic bottleneck model. More specifically, we show that the impact is large for small number of travellers N, while for $N \to \infty$ the stochastic model tends to the solution of the original model.

The original model in [2] considers a single bottleneck queue where travellers are modelled as a fluid. The total amount of fluid is denoted by N and the capacity of the bottleneck by s, the duration of the bottleneck period $[t_a, t_b]$ is fixed and is given by N/s. All travellers want to depart from the bottleneck at the same time denoted by t^*. By means of a cost function, the discomfort of travellers departing from this queue is captured, in the simplest form, costs are equal for all travellers. This cost function of the fluid model is defined by to the following function

$$c(t) = \alpha w(t) + \beta \max\{0, t^* - t - w(t)\} + \gamma \max\{0, t + w(t) - t^*\} \qquad (1)$$

where $w(t)$ is the waiting time when a traveller arrives at time t at the bottleneck, α gives the costs for waiting, β the costs for arriving earlier and γ for departing later than the desired time. Costs are given per unit of time.

The cost function for the stochastic model is identical to that of the deterministic model, with the delay $w(t)$ replaced by its stochastic counterpart $W(t)$. To compute the impact of stochasticity over time for an arbitrary day, we compute the expected costs

$$\mathbb{E}C(t) = \alpha \mathbb{E}W(t) + \beta \mathbb{E}[(t^* - t - W(t))] + \gamma \mathbb{E}[(t + W(t) - t^*)] \qquad (2)$$

In the stochastic model travellers divide themselves over time such that on average they encounter the same costs. To determine the difference in impact for traveller with large discomfort compared to small discomfort due to uncertainty, we can include a cost parameter for uncertainty. We define this by

$$\mathbb{E}C(t) = \alpha\mathbb{E}W(t) + \beta\mathbb{E}[(t^* - t - W(t))] + \gamma\mathbb{E}[(t + W(t) - t^*)] + \delta Var[W(t)], \tag{3}$$

where δ gives the costs per unit of time due to uncertainty.

3 Analysis

We numerically compute the waiting time distribution by discretising t into small time steps of length Δ, where Δ is small, to approximate this cost function. This gives

$$\begin{aligned}\mathbb{E}C(t) \approx \tilde{C}(t) =& \alpha \sum_{\tau=0}^{\infty} \tau * f(\tau;t) + \beta \sum_{\tau=0}^{(t^*-t)/\Delta} (t^* - t - \tau) * f(\tau;t) \\ &+ \gamma \sum_{\tau=t^*-t}^{\infty} (t + \tau - t^*) * f(\tau;t)\end{aligned} \tag{4}$$

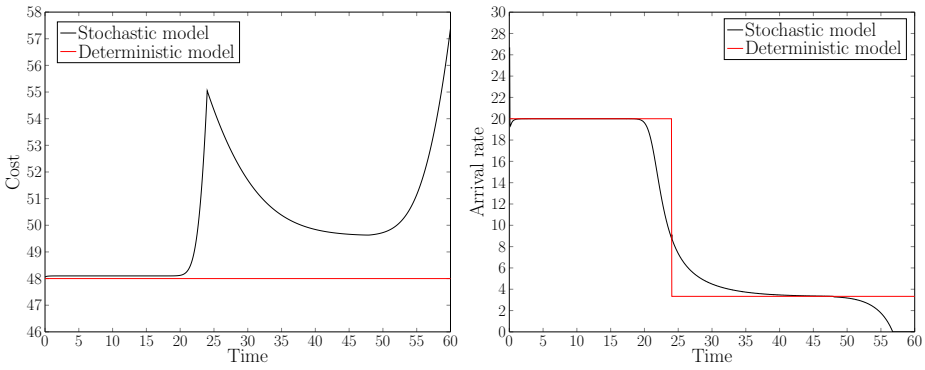
where $f(\tau;t) = \sum_{n=0}^{N} \pi_n(t) * g_n(\tau)$ represents the density function $g_n(\tau)$ of the sojourn time τ conditioned on the queue length $\pi_n(t)$ at time t for n travellers at the bottleneck.

4 Results

Input fluid model						Output fluid model				
N	s	t^*	α	β	γ	r_1	r_2	t_n	Eq. Cost	Timespan $[t_a, t_b]$
600	10	480	2	1	4	20	3.33	240	480	[0,600]

Table 1: Input/ output parameters for the deterministic bottleneck.

To analyse the impact of stochasticity in this bottleneck model we first compute the Wardrop equilibrium of the fluid queue as explained in **??**. We use a numerical example with parameter values of Table 1. The resulting equilibrium arrival rate, cost, time span and peak moment t_n are shown as well. We use the



(a) Cost function arrival rate r_1 and r_2. (b) Arrival rate function with cost $c_f(t)$.

Figure 1: Cost and arrival rate function for the deterministic model and the approximated stochastic model with parameter set 1.

resulting arrival curve to evaluate the costs in the equivalent $M_t/M/1$ queue. The arrival curve is a piecewise linear function given by

$$\lambda_t = \begin{cases} 0 & t \in (-\infty, t_a) \\ r_1 & t \in [t_a, t_n) \\ r_2 & t \in [t_n, t_b) \\ 0 & t \in [t_b, \infty) \end{cases} . \tag{5}$$

where $[t_a, t_b]$ gives the duration of the bottleneck period.

Figure 1a demonstrates that the error this (expected) costs incur vary notably between travellers, in particular putting travellers arriving towards the end of the busy period in a disadvantageous position. The large increase in costs just before the peak moment is due to the costs of late arrival which are not encountered the deterministic model. At the end of the bottleneck period there is a fair chance that the queue did not disappear at time t_b. Using this as a starting point, we then attempt to numerically repair this cost imbalance by adapting the arrival rate function. This means that we set the equilibrium cost function $c_f(t) \equiv C$ obtained from the fluid model and use this to compute the corresponding arrival rate of our stochastic model by

$$c_f(t) \approx \mathbb{E}C(t) = \alpha \mathbb{E}w(t) + \beta \mathbb{E}[(t^* - t - w(t))] + \gamma \mathbb{E}[(t + w(t) - t^*)]. \tag{6}$$

The resulting time-dependent arrival rate can be seen in Figure 1b together with the rate of the original fluid model for comparison. This shows that the clear cut that is obtained for the deterministic model behaves more fluently over time. The smaller the total amount of travellers N, the smoother this function will become.

Having obtained an approximation of the equilibrium in the stochastic model allows us to investigate the uncertainty by plotting the mean against the stan-

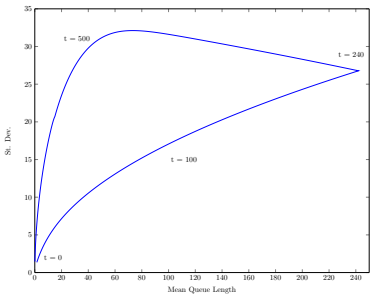


Figure 2: Mean queue length and St.dev. over time.

dard deviation over time shown in Figure 2. We can see that the standard deviation increases when the congestion increases at the bottleneck, however, when the congestion decreases, the standard deviation decreases more slowly. This indicates that uncertainty at the end of the bottleneck period has a larger impact than at the beginning. Similar results where shown by Fosgerau in [7] in which he refers to this as a looping phenomenon which has also been observed in empirical data. In this study we continue our analysis by adding an extra cost parameter to the stochastic model. This research is still in progress.

References

[1] W.S. Vickrey. *Congestion Theory and Transport Investment.* The American Economic Review, Vol. 59, No. 2, pp. 251-260, 1969.

[2] R. Arnott, A. De Palma, R. Lindsey. *Economics of a bottleneck.* Journal of urban economics Vol. 27 No. 1, pp 111-130, 1990.

[3] K.A. Small, *The Bottleneck Model: An Assessment and Interpretation.* Forthcoming, Economics of Transportation, 2015.

[4] R. Noland, *Commuter responses to travel time uncertainty under congested conditions: expected costs and the provision of information.* Journal of urban economics 41.3, 1997.

[5] R. Noland, K. Small, *Travel-time uncertainty, departure time choice, and the cost of the morning commute.* Institute of Transportation Studies, University of California, Irvine, 1995.

[6] A. De Palma, M. Fosgerau. *Random queues and risk averse users.* European Journal of Operational Research 230.2, 2013.

[7] M. Fosgerau, *On the relation between the mean and variance of delay in dynamic queues with random capacity and demand.* Journal of Economic Dynamics and Control 34.4: 598-603, 2010.

Recasting intersection automation as a connected-and-automated-vehicle (CAV) scheduling problem: A sequential branch-and-bound search approach in Phase-Time-Traffic hypernetwork

Pengfei (Taylor), Li [a1] , Xuesong Zhou [b]

[a] Department of Civil and Environmental Engineering, Mississippi State University, MS, 39762, USA
[b] School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ, 85287, USA

1. Introduction

With the rapid development of wireless communication, sensing and computing techniques, it has become a common vision that connected and automated vehicles (CAV) will soon go beyond testbeds to appear on roads with traditional (i.e., human-driven) vehicles at large scale. In the meantime, it is also envisioned that we will live in an era with both CAVs and traditional vehicles for a long time before CAVs completely replace traditional vehicles. CAVs will undoubtedly bring new potential of improving the automation level at intersections however the co-existence of CAVs and traditional vehicles in the real world makes it challenging to describe the dynamic interaction between heterogeneous vehicle movements and signal operations and then further develop systematic intersection automation policies (IAP) benefiting both CAVs and traditional vehicles. Within the heterogeneous traffic environment, CAVs are viewed as a player with dual roles at intersections: (a) a part of traffic flow and interact with traditional vehicles and (b) identities sending green requests to intersection(s). The CAV's first role (as vehicles) is explicitly separated from the second (as green request senders). Furthermore, we view the urban traffic dynamics as a *sequential decision process* or SDP. Traffic vehicle locations or states at time t are the direct results of traffic signal operations from time 0 to time t. On the other hand, traffic signal operations from t to the future also needs to consider both current and future traffic states according to their inherent traffic vehicular dynamics. As such the target problem modeling and solution algorithms developed in this paper aim to fully capture the nature of sequential decision process through coupling heterogeneous traffic and traffic signal operations.

Extended from the notations due to Karp and Held (1967), traffic propagation over time can be represented by a quintuple $\mathbb{N} = (A, R, r_0, r_T, \Gamma)$, where $\mathbb{N}$ is the subject traffic system; A is the set of time-dependent inputs such as traffic lights; R is the set of all feasible traffic states including vehicles' locations, traffic signal status and the set of served green requests; r_0 represents the initial state; r_T represents the final traffic state and Γ is the state transition function. In addition, it is also necessary to define ρ, an input data set for SDP including the link travel times, saturation flow rates, etc. and two conceptual functions:

1. State transition function h: $R \times A \times P \to h$. For a given data set $\rho \in P, a \in A$, quantity of $h(r, a)$ gives the cost of transiting from the initial state r to an updated traffic state $r' = \Gamma(r, a)$ due to a, where Γ is the traffic state transition function. Function h is critical in traffic signal modeling and optimization because it is necessary to evaluate the traffic state evolution as well as incurring costs under a possible traffic operation during the optimization. As examples, in addition to signal phasing transition rules, state transition function h is most commonly defined based on cumulative vehicle counting curves in which traffic signal operations change the departure curve (the "D curve") and control delays (Daganzo, 1997). h is also defined as changes to arterial "green band" under traffic signal timings at intersections in a mixed-integer linear programming (MILP) form in the literature due to Little et al.(1981) and Gartner et al. (1991); defined as the changes to queue length and delay changes at intersections under a traffic signal operation by Mirchandani and Head (2001); defined as traffic dynamics changes under given traffic signal operations in the cell transmission model (CTM) developed by Daganzo (1994) in the literature due to Lin and Wang (2004) and Lo (1999); as the propagation of vehicles stored on links by Aboudalas et al. (2009); as the propagation of traffic shockwaves by Wu and Liu (2011) and as changes to vehicle routing plan in the literature due to Li et al. (2015).
2. Cost function k: $P \times R \to k$, for a given data $\rho \in P$, quantity of $k(r)$ gives the cost of traffic state $r \in R$.

Based on the concept of SDP, the intersection automation policy optimization can be defined as follows: construct an algorithm which schedules the traffic signal operations during the time horizon T, given the (prior known) data set ρ so as to make $\mathbb{N}$ reach a feasible final state q_T at the minimum cost. In particular, within the heterogeneous traffic environment, a traffic state r contains two groups of objects: CAV requests for green and traffic dynamics comprised of all vehicles. The CAV requests can be represented in a form like "*Request green on movement X at t*" while all vehicle locations on roads are governed by traffic dynamics defined in R and Γ.

2 Construct Phase-Time-Traffic (PTR) hypernetwork to model heterogeneous traffic propagation under various intersection automation policies

[1] Corresponding author. Email: pengfei.li@cee.msstate.edu ;

With the SDP modeling framework, the challenge is that elements including traffic signal operations, heterogeneous traffic propagation as well as CAV schedules all involve additional hard constraints. The complex interactions between those elements makes it nearly impossible to solve the problem efficiently with mathematical-programming-based solvers. To address these issues, we develop a new phase-time-traffic hypernetwork model, referred to as PTR hypernetwork, to "pre-build" most hard constraints and complex interactions among heterogeneous traffic, traffic signal operations and CAV requests into the PTR hypernetwork itself to ensure the traffic dynamics feasibility and signal transition feasibility during the sequential traffic decision progress.

Traffic signal operation representation: three items are necessary to define a complete traffic signal control strategy at intersections: signal phases, phasing sequence and respective phase durations. The first two items can be well represented in the classic state-transition diagram but it cannot reflect the phase durations. To address this issue, Li et al. (2015) developed the phase-time network model or PT network model to reflect all three items in an acyclic, discrete and forward network. Each arc in the PT network is interpreted as one feasible traffic signal operation. Without loss of generality, we assume the minimum time step is 1s.

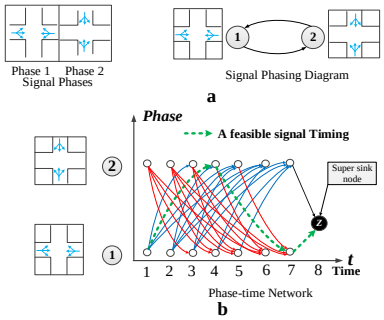


Figure 1 Demonstration of traffic signal operations in phase-time network
(min green=2 s; max green=3s; adapted from Li et al. (2015))

Seeking the optimal traffic signal timing in the phase-time network is in essence to seek the shortest path from the start PT node (current signal phase at current time) to the end PT node (the virtual sink node after the time horizon). All hard constraints for traffic signal control systems, including the minimum and maximum green requirements on each phase, mutual exclusiveness of incompatible phases and cyclic or flexible phasing sequences, can be "prebuilt" into the PT network.

CAV request for green with time window representation: One feature of CAVs is that they move along the scheduled routes and their arriving times at downstream intersections can be predicted. Therefore, it is possible for CAVs to send green requests in terms of time windows to several intersections in advance. All PT-arcs able to serve a CAV request can be divided into two types: admissible PT arcs to a CAV request turning the corresponding phase to green during the time window; "early-green" PT arcs of this CAV request to start the green early and hold until the CAV request's time window starts. Mathematically they are defined as below:

Denote a CAV request on phase p with a time window as $(p,t), t_1 \le t \le t_2, t_1, t_2$ are boundaries of this request's time window. Then a PT arc l_1 is an "admissible PT arc" to this CAV request if $l_1 = (p', t', p,\ t)$ where $p' \neq p$; a PT arc l_2 is an "early-green" PT arc of this CAV request if $l_2 = (p, t', p, t'')$ where $t' < t_1$ and $t'' > t_1$

Two types of PT arcs are illustrated in Fig. 2 and both of them are superimposed with additional large negative cost $-\lambda$ as the incentive for search algorithms to select those arcs while searching a least-cost path in the phase-time network. The second restriction is introduced to ensure one and only one PT arc, which is either the "admissible" PT arc to or the "early-green" PT arc of a given CAV request, is selected.

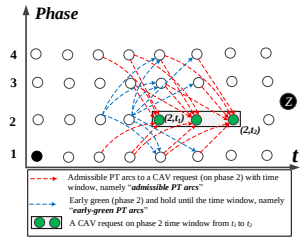


Figure 2 illustration of modified phase-time network model to accommodate future green requests

Please note that if multiple CAV requests on the same phase with overlapping time windows and some PT arcs happen to be able to serve multiple CAV requests, those PT arcs will be superimposed with multiple $-\lambda$ values to reflect a preference

of using one green phase to service multiple green requests. Fig. 3 shows this concept. Accordingly, the search algorithm will favor those overlapping PT arcs with multiple negative λs more in seeking least-cost paths.

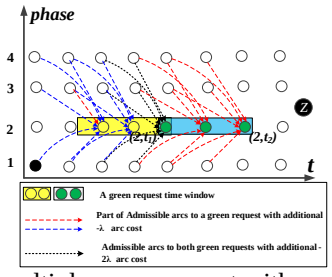


Figure 3 PT arc costs for multiple green request with overlapping time windows

Heterogeneous traffic propagation under traffic signal operations in phase-time-traffic (PTR) hypernetwork: Fig. 4 illustrates the concept of the new PTR hypernetwork in which all vehicles, including CAVs and traditional vehicles, are released into the space-time plane and propagate. Interactions between vehicles and traffic signal operations can significantly affect vehicle trajectories. Constructing a PTR hypernetwork starts from a two-dimension PT network model. Each PT arc can be expanded to a series of three-dimension PTR edges along the traffic dimension. As an illustration, assuming there are only two traffic conditions whose indices are r_1, r_2 in Fig. 4. A two-dimension PT arc (dash arrow line) can be expanded to a series of three-dimension PTR edges (solid arrow line) according to feasible traffic state transitions. Since a traffic state index uniquely maps a traffic state, we use "traffic state" to refer to the traffic dimension in the rest of this paper unless confusion is caused.

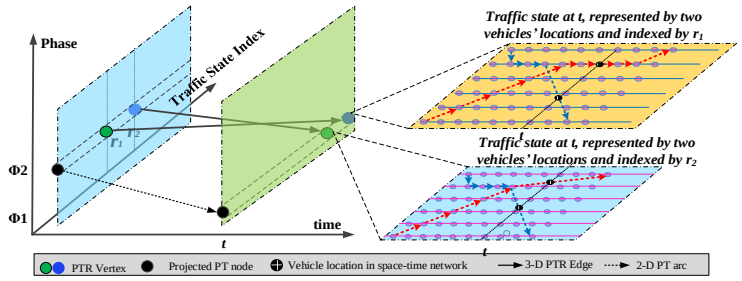


Figure 4 illustration of the PTR hypernetwork concept

3. Mixed-integer linear programming (MILP) formulation for IAP optimization in the PTR hypernetwork

The problem is described in a multi-commodity formulation:

$$\text{Minimize } P = \sum_{(p,t,r,p\prime,t',r')\in E}\left(c_{(p,t,r,p',t',r')} - N_{(p,t,r,p',t',r')} \times \lambda\right) y_{(p,t,r,p\prime,t',r')} \quad (1)$$

Where p, p' are phase numbers, t, t' are time indices and r, r' are traffic state indices

Subject to:

Flow conservation constraints at origin vertex

$$\sum_{(o,0,r_0,p',t',r')\in E} y_{(o,0,r_0,p',t',r')} = 1 \qquad \forall (p',t',r') \in V \quad (2)$$

Flow conservation constraints at destination vertex

$$\sum_{(p,t,r,z,T,r')\in E} y_{(p,t,r,z,T,r')} = 1 \qquad \forall (p,t,r) \in (z,T,r_T) \quad (3)$$

Flow conservation constraints at intermediate vertices

$$\sum_{(p,t,r,p',t',r')\in E} y_{(p,t,r,p',t',r')} - \sum_{(p,t,r,p',t',r')\in E} y_{(p,t,r,p',t',r')} = 0$$

$$\forall (p',t',r') \in V\backslash\{(o,0,r_0),(z,T,r'),p \in R\} \quad (4)$$

Multiple services for one CAV request prevention

$$\sum_{(p,t,r,p',t',r')\in E_k} y_{(p,t,r,p',t',r')} = 1;\ \forall r,r' \in R, \forall k \in K, (p',r') \in E_k \quad (5)$$

Edge costs $c_{(p,t,r,p',t',r')}$ in (1) are not constant but dynamically determined. A traffic state can be simply defined using cumulative vehicle counting curves $A_m(t), V_m(t), D_m(t)$ and the area between $V_m(t), D_m(t)$ or $S\left(V_m(t), D_m(t)\right)$ on all approaches. Accordingly, $c_{(p,t,r,p',t',r')}$ can be defined as:

$$c_{(p,t,r,p',t',r')} = \sum_{m\in M}\left(S\left(V_m(t'), D_m(t')\right) - S\left(V_m(t), D_m(t)\right)\right) \quad (6)$$

4. Sequential branch-and-Bound Search Algorithm in PTR hypernetwork

A sequential branch-and-bound search algorithm is developed to search the optimal IAP and special pruning (dominance) rule is designed for this context.

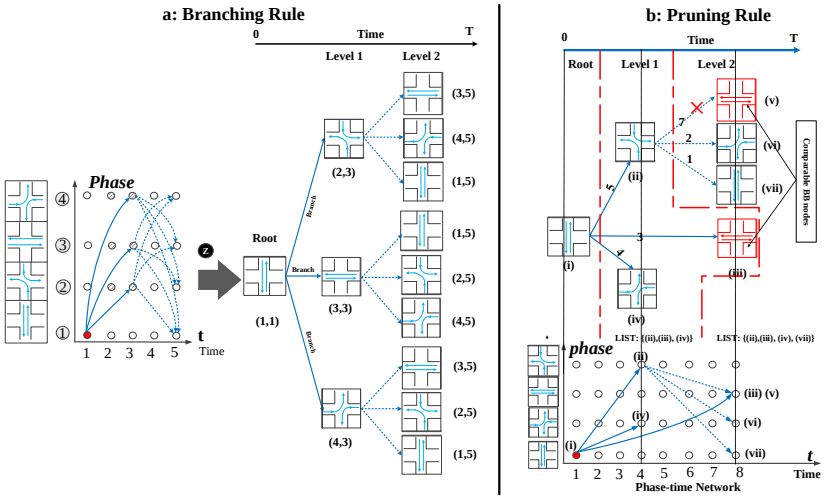


Figure 5 illustration of branching and pruning rules for the BB search algorithm

5. Experiments and findings

The subject intersection is Intersection 5 and the baseline Traffic condition is formed after 200 MSA iterations.

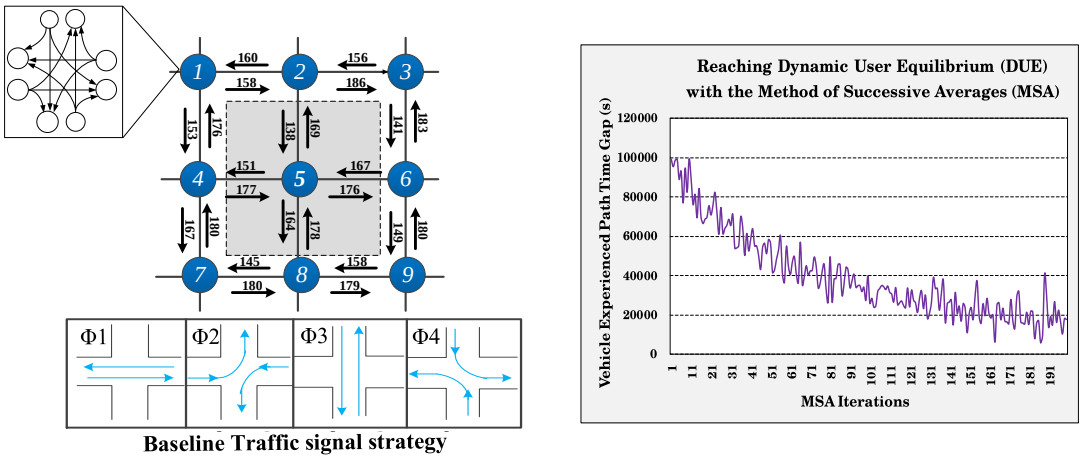
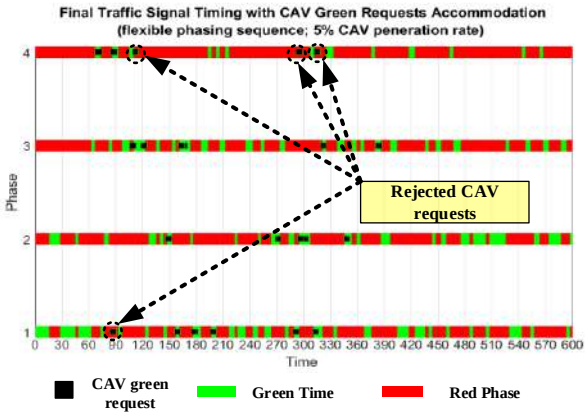


Figure 6 Configurations of experiment network

Operational Performance Analysis under the proposed intersection control strategy. Most CAV requests are served and the total delay is also lower than the baseline.



	North Bound			South Bound		
	Right	Through	Left	Right	Through	Left
New control policy	11.8	8.6	9.6	3.9	4.4	4.2
Baseline	12.8	9.6	9.8	3.9	4.7	4.9
	West Bound			East Bound		
	Right	Through	Left	Right	Through	Left
New control policy	14.4	13.9	13.5	13.4	12.7	11.6
Baseline	13.5	14.5	16.4	13.2	12.9	11.2

Figure 7 CAV green request service rate under the proposed intersection automation policy

Concluding remarks

The optimal IAP will serve the CAV requests at its maximum potential and maintain an acceptable traffic mobility. Nowadays many special vehicles like fire trucks or buses today already have most CAV features and therefore the proposed IAP optimization framework can also be promising to solve special-vehicle-related problems at intersections.

The IAP optimization framework also has potential for network-wide CAV scheduling as well in the sense that (a) the definition of "phases" at intersections can be generalized as a group of "openable simultaneously" movements across intersections; (b) all the CAV requests are proactively placed according to their scheduled paths and the proposed IAP optimization framework at congested intersections can also suggest certain CAVs to drop their requests and re-schedule their paths to avoid congestions. In the future, we plan to combine the proposed IAP with other CAV path scheduling policies to ensure the CAV requests not to exceed intersections' maximal potential. In doing so, it is expected that balanced CAV requests can be served at each intersection and so travel time reliability will be greatly improved. We also plan to further improve the efficiency of parallel computing in the branch-and-bound search algorithm for large-scale applications.

Reference

Aboudolas, K., Papageorgiou, M., Kosmatopoulos, E., 2009. Store-and-forward based methods for the signal control problem in large-scale congested urban road networks. Transportation Research Part C: Emerging Technologies 17(2), 163-174.

Daganzo, C.F., 1994. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. Transportation Research Part B: Methodological 28(4), 269-287.

Gartner, N.H., Assman, S.F., Lasaga, F., Hou, D.L., 1991. A multi-band approach to arterial traffic signal optimization. Transportation Research Part B: Methodological 25(1), 55-74.

Gartner, N.H., Little, J.D., Gabbay, H., 1975. Optimization of traffic signal settings by mixed-integer linear programming: Part I: The network coordination problem. Transportation Science 9(4), 321-343.

Kamal, M.A.S., Imura, J.-i., Hayakawa, T., Ohata, A., Aihara, K., 2015. A vehicle-intersection coordination scheme for smooth flows of traffic without using traffic lights. IEEE Transactions on Intelligent Transportation Systems 16(3), 1136-1147.

Karp, R.M., Held, M., 1967. Finite-state processes and dynamic programming. SIAM Journal on Applied Mathematics 15(3), 693-718.

Li, P., Mirchandani, P., Zhou, X., 2015. Solving simultaneous route guidance and traffic signal optimization problem using space-phase-time hypernetwork. Transportation Research Part B: Methodological 81, 103-130.

Little, J.D., Kelson, M.D., Gartner, N.M., 1981. MAXBAND: A PROGRAM FOR SETTING SIGNALS ON ARTERIES AND TRIANGULAR NETWORKS. Transportation Research Record 795.

Lo, H.K., 1999. A novel traffic signal control formulation. Transportation Research Part A: Policy and Practice 33(6), 433-448.

Mahmoudi, M., Zhou, X., 2016. Finding optimal solutions for vehicle routing problem with pickup and delivery services with time windows: A dynamic programming approach based on state–space–time network representations. Transportation Research Part B: Methodological 89, 19-42.

Mirchandani, P., Head, L., 2001. A real-time traffic signal control system: architecture, algorithms, and analysis. Transportation Research Part C: Emerging Technologies 9(6), 415-432.

Wu, X., Liu, H.X., 2011. A shockwave profile model for traffic flow on congested urban arterials. Transportation Research Part B: Methodological 45(10), 1768-1786.

Distributed Computation based Constrained Model Predictive Control for a Mixed Flow Platoon

Siyuan Gong, Lili Du*[1]

Extended Abstract

1. Background and motivation

It has been shown that improper car-following behavior, e.g., overreacting or timid car-following, is one of the key factors to trigger traffic congestion, oscillations (stop-and-go or slow-and-fast traffic), and accidents. The connected and autonomous vehicle (CAV) technologies provide tremendous opportunities to develop cooperative adaptive cruise control (CACC) technologies, which target cooperative driving among vehicles to ensure traffic safety and efficiency. Motivated by this view, extensive interests have been attracted from both industry and academia sides and produce numerous control strategies/laws such as car-following model based CACC and machine learning based CACC. However, most of existing CACC technologies either focus on individual CAVs or a pure CAV platoon. Few studies work on mixed flow platoon control which includes CAVs and human-drive vehicles (HDVs). HDVs do not communicate with neighborhood vehicles or follow control law, thus they will very likely worsen traffic flow and impair the advantages of CACC technologies. Motivated by this view, this research aims to develop a cooperative platoon control using model predictive control and distributed computation for a mixed flow platoon, in order to improve traffic smoothness, stability, and efficiency of the entire platoon (i.e. system performance).

2. Preliminaries

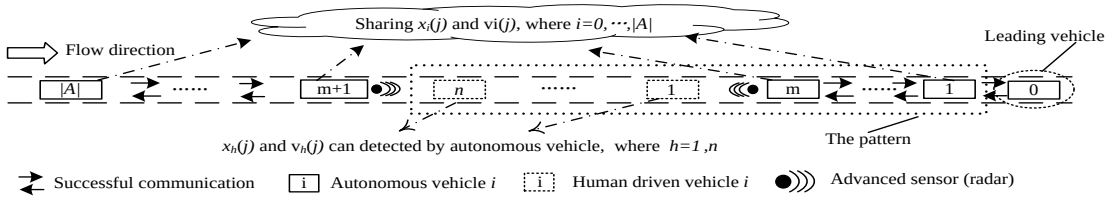


Figure 1. The simple and standard arrangement of a platoon in mixed traffic flow

To develop rigorous study, this research makes the following agreements/assumptions for a sample platoon shown in Figure 1. Specifically, this sample platoon contains a leading vehicle numbered as vehicle 0 with unpredictable movement. Then from the downstream to upstream, there are three platoons following this leading vehicle. They are respectively a CAV sub-platoon A_1 including m number of CAVs, a HDV sub-platoon H including n number of HDVs, and another CAV sub-platoon A_2 including the CAVs from $m+1$ to $|A|$. We define CAV set $A = A_1 \cup A_2$ and use $|A_1|$, $|A_2|$, and $|A|$ to denote the total number of CAVs in the sets of A_1, A_2 and A, respectively. It is easy to show that repeating this sample platoon will produce more general mixed flow platoons. Moreover, we consider a semi-global information structure for the development of the control algorithm. Namely, the movement of all CAVs and their adjacent HDVs can be detected by on-board sensors and then shared among CAVs in this mixed platoon. In other words, we consider that all CAVs are well connected, including CAVs m and $m+1$, even though they are spatially separated by some HDVs.

Built upon the above assumptions, this study seeks to develop a control scheme for this sample mixed flow platoon, explicitly putting platoon smoothness and stability under control while still ensuring individual vehicles' mobility. *Mainly, the following research challenges will be addressed*. First, HDVs introduce uncertain responses to traffic variation and unknown platoon length. These issues raise the challenges to forecast the movement of HDVs and thus affect the control algorithm design. Second, a vehicle platoon is a self-organized system. It may consist of a large number of vehicles and is subject to frequent topology variation. To implement such platoon control, centralized computation usually is inefficient or infeasible due to the lack of roadside computing facilities except around intersections. This challenge calls for the need of distributed computation, which takes advantages of CAV's local computing capability but may introduce extra communication load and challenges of ensuring the convergence. The presented research below addresses these challenges with our innovative approaches.

[1]Corresponding author. ldu3@iit.edu. Illinois Institute of Technology. 3201 South Dearborn Street Chicago, Illinois 60616, U.S.A.

3. Methodology

This section develops the cooperative platoon control using traffic flow theory, model predictive control (MPC), machine learning, and optimization techniques. Briefly, we apply the one-step MPC developed in [3] for the CAVs in A_1 since the movement of its leading vehicle 0 is unpredictable. Next, we consider the first CAV in the platoon A_2 is able to learn and predict the movement of its leading vehicle (HDV n), which is affected by the platoon A_1. Accordingly, a P-step MPC combined with a machine learning algorithm is developed to control the car-following movement of the CAVs in the platoon A_2. The cooperative platoon control aims to ensure both transient traffic smoothness and asymptotic traffic stability of the collective movement of vehicles in the platoon. The following presents the key technical details of our approaches.

3.1. Vehicle dynamics

This research studies the vehicle movement at discrete sample times (indexed by $j = 1, \dots, J$) with time interval $\Delta t > 0$. The movements of CAVs and HDVs are modeled by different dynamics models. Accordingly, the discrete-time linear model for the longitudinal dynamics of a CAV is given below:

$$x_i(j+1) = x_i(j) + \Delta t v_i(j) + \frac{\Delta t^2}{2} u_i(j), \qquad i \in A = A_1 \cup A_2 \tag{1}$$

$$v_i(j+1) = v_i(j) + \Delta t u_i(j), \qquad i \in A = A_1 \cup A_2 \tag{2}$$

where $x_i(j)$ and $v_i(j)$ represent the location and speed of CAV i at the time step j and $u_i(j)$ represents the control variable, which is constant in each interval $[j\Delta t, (j+1)\Delta t]$.

For HDVs, this study only considers the HDVs adjacent to the CAV platoons (such as HDVs 1 and n in Figure 1). By applying the Newell's car-following model, the movement dynamics of the two HDVs are given by the formulations below.

$$x_1(j\Delta t + \tau_1) = x_m(j) - d_1 \tag{3}$$

$$x_n(j\Delta t + \sum_{h=2}^{n} \tau_h) = x_1(j) - \sum_{h=2}^{n} d_h \tag{4}$$

where τ_h and d_h are the reaction time and the minimum stop distance of human driver h.

Equations (3) and (4) indicate that the trajectory variation at HDV 1 will propagate to HDV n with an accumulated delay $\sum_{h=2}^{n} \tau_h$. However, the values of τ_h and d_h for $h = 1, \dots, n$ are unknown and vary among HDVs. Moreover, it is hard to know the number of the HDVs in-between CAVs m and $m+1$ (i.e., the value of n is unknown). These unknown factors cause the difficulty to estimate the HDV dynamics. To address this challenge, this study applies a curve matching algorithm [1], combining the online collected trajectory data of CAV m, HDVs 1 and n, and the features of the Newell's car-following model. Below briefly introduces this approach.

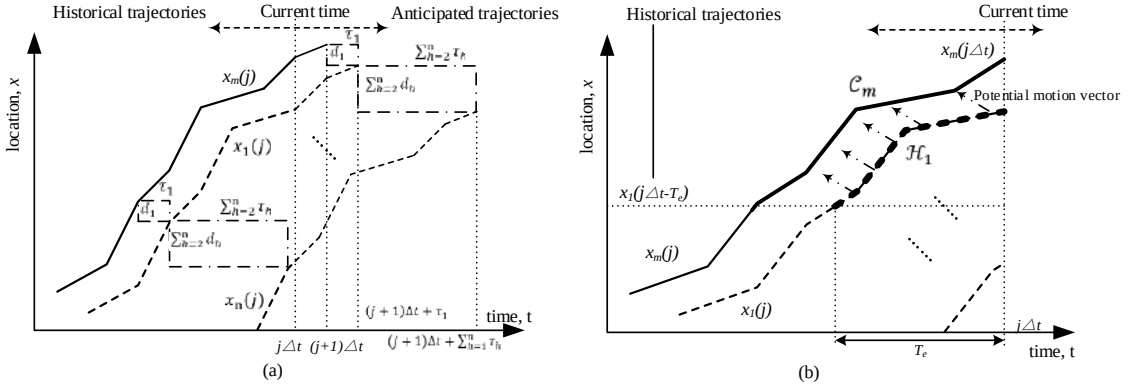


Figure 2. (a) Newell's car-following model. (b) Basic idea of curving matching.

According to the Newell's car-following model, it is recognized that the trajectories of CAV m, and HDVs 1 and n follow the same shape but with spatial-temporal delay (see Figure 2(a)). Thus, the curve matching algorithm [1] is used to estimate the spatial-temporal delay for the HDVs in real time (i.e. the parameters $\tau_1, d_1, \sum_{h=2}^{n} \tau_h$, and $\sum_{h=2}^{n} d_h$ in Equations (3) and (4)). We use $\mathcal{H}_1$ and $\mathcal{C}_m$ to denote the trajectories of HDV 1, and CAV m, respectively (see Figure 2(b)). The main idea of the curve matching algorithm is to iteratively move a trajectory (such as $\mathcal{H}_1$) by a motion vector $(-\tau_1, d_1)$ until coinciding with the other trajectory (such as $\mathcal{C}_m$). Specifically, the algorithm iteratively matches a point on $\mathcal{H}_1$ to the closest point on $\mathcal{C}_m$. A least-squares method is used to select the best motion vector $(-\tau_1^*, d_1^*)$, which minimizes the average distance of all paired points between the two trajectories. This procedure runs repeatedly until one trajectory can match the other with acceptable error. The same algorithm is also used

to estimate the moving vector $(-\sum_{h=2}^{n}\tau_h, \sum_{h=2}^{n} d_h)$ by matching the trajectories of the HDV n to the trajectory of HDV 1, i.e., $\mathcal{H}_n$ and $\mathcal{H}_1$. Once the $\tau_1, d_1, \sum_{h=2}^{n}\tau_h$, and $\sum_{h=2}^{n} d_h$ are well estimated for the HDVs, we can use Equations (3) and (4) to anticipate the trajectories of HDVs 1 and n during time intervals $[j\Delta t, (j+1)\Delta t + \tau_1)$ and $[j\Delta t, (j+1)\Delta t + \sum_{h=1}^{n}\tau_h)$, respectively.

3.2. Constrained model predictive control for the mixed platoon

Recall that this study considers that the mixed flow platoon in Figure 1 includes three sub-platoons: the CAV platoon A_1 following the lead vehicle 0, the CAV platoon A_2 and the HDV platoon H staying in-between the two CAV platoons. Accordingly, this research develops two interactive MPCs [2] to regulate the movements of CAVs in A_1 and A_2 so that we can improve traffic stability, smoothness, and efficiency of the entire mixed flow platoon, while still ensure individual vehicles' mobility. Below present our approaches in details.

Clearly, the movements of the CAVs in A_1 (CAVs from 1 to m in Figure 1) is affected by the movement of the leading vehicle 0. Accordingly, we define the control dynamics as follows subject to Equations (1) and (2).

$$z_1(j+1) \coloneqq (x_0(j+1) - x_1(j+1) - s_d, \dots, x_{m-1}(j+1) - x_m(j+1) - s_d)^T \in \mathbb{R}^{|A_1|}$$

$$z_1'(j+1) \coloneqq \left(v_0(j+1) - v_1(j+1), \dots, v_{m-1}(j+1) - v_m(j+1)\right)^T \in \mathbb{R}^{|A_1|}$$

Considering the movement of the leading vehicle 0 is unpredictable, this study applies one-step MPC for the CAV platoon A_1 with the constrained optimizer given in Equations (5) to (8).

Optimizer-I

$$\textbf{Min}\ \Gamma_1(u(j)) \coloneqq \frac{1}{2}\left[z^T(j+1)Q_z z(j+1) + \left(z'(j+1)\right)^T Q_{z'} z'(j+1)\right] + \frac{\Delta t^2}{2}\|u_I(j)\|_2^2 \quad (5)$$

subject to:

$$a_{min,i} \le u_i(j) \le a_{max,i},\ \ i \in A_1 \quad (6)$$

$$v_{min} \le v_i(j+1) \le v_{max},\ \ i \in A_1 \quad (7)$$

$$x_{i-1}(j+1) - x_i(j+1) \ge L_i + v_i(j+1)\,\Delta t - \frac{[v_i(j+1) - v_{min}]^2}{2a_{min,i}}, i \in A_1 \quad (8)$$

where, the control variables are $u_I(j) \coloneqq (u_1(j), \dots, u_m(j))^T \in \mathbb{R}^n$. Let $\alpha \coloneqq (\alpha_1, \dots, \alpha_m)$ and $\beta \coloneqq (\beta_1, \dots, \beta_m)$, where $\alpha_i > 0$ and $\beta_i > 0$ are penalty weights for $\forall\, i \in A_1$, we can define two diagonal matrices $D_\alpha = \mathrm{diag}(\alpha_1, \dots, \alpha_m)$ and $D_\beta = \mathrm{diag}(\beta_1, \dots, \beta_m)$. Then, let $Q_z \coloneqq R^T D_\alpha R$ and $Q_{z'} \coloneqq R^T D_\beta R$ be symmetric and positive definite matrices, where R is an orthogonal matrix which characterizes the interaction of the CAVs in this platoon through the control. By the matrices Q_z and $Q_{z'}$, the weights α and β will directly affect the eigenvalues of the linear closed-loop system, which makes the stability analysis of Optimizer-I much simpler. The constraints contain the acceleration/deceleration limit $[a_{min,i}, a_{max,i}]\ \forall\, i \in A_1$, speed limit $[v_{min}, v_{max}]$, and safety spacing, where $a_{min,i} < 0 < a_{max,i}$ and $v_{min} < v_{max}$.

The objectives in Γ_1 are chosen to minimize traffic flow oscillations using a mild control in control horizon. Accordingly, Γ_1 includes the penalties on three terms: (i) the errors between desired vehicle spacing and actual spacing; (ii) the fluctuations of the spacing between adjacent vehicles, i.e., the relative speed between adjacent vehicles; and (iii) the variations of individual vehicles' speed, i.e., vehicles' acceleration/deceleration.

The movements of the CAVs in A_2 (CAVs $i = m+1, \dots, |A|$ in Figure 1) will be affected by the movement of HDV n (i.e., the leading vehicle of the CAV platoon A_2). It is noticed that the movement of the HDV n will be influenced by the CAV m. Recall this study assumes that the CAVs m and $m+1$ are connected by wireless communication even though they are separated by the HDV platoon. Then, the CAV $m+1$ is able to predict the response of the HDV n according to the movement of CAV m during the time interval $[j\Delta t, (j+1)\Delta t + \sum_{h=1}^{n}\tau_h)$ through its vehicle dynamics function in Equation (4) by the real-time estimation of $\tau_1, d_1, \sum_{h=2}^{n}\tau_h$, and $\sum_{h=2}^{n} d_h$. Accordingly, this study adopts the P-horizon MPC [2] to control the car-following behavior in the CAV platoon A_2, where $P = 1 + \lfloor(\sum_{h=1}^{n}\tau_h)/\Delta t\rfloor$.

Similarly, for a given movement state $\left(x_{i,j}(0), v_{i,j}(0)\right)_{i=m+1}^{|A|} = (x_i(j), v_i(j))_{i=m+1}^{|A|}$ (i.e. setting the initial movement state of P-horizon MPC as the current movement state at time tick j) and the anticipated movement of the HDV n $\left(x_{n,j}(p), v_{n,j}(p)\right)_{p=1}^{P}$ at time tick j, the vehicle dynamics of CAVs in A_2 and HDV n in future P steps at sample point j can be anticipated by Equations (9) to (11).

$$x_{i,j}(p) = x_{i,j}(p-1) + \Delta t v_{i,j}(p-1) + \frac{\Delta t^2}{2} u_{i,j}(p-1), \quad i \in A_2, \quad p \in P \tag{9}$$

$$v_{i,j}(p) = v_{i,j}(p-1) + \Delta t u_{i,j}(p-1), \quad i \in A_2, \quad p \in P \tag{10}$$

$$x_{n,j}(p) = x_{1,j}(p\Delta t - \sum_{h=2}^{n} \tau_h) - \sum_{h=2}^{n} d_h, \quad p \in P \tag{11}$$

Then we define the control dynamics as follows subject to Equations (9) to (11).

$$z_{2,j}(p) \coloneqq \left(x_{n,j}(p) - x_{m+1,j}(p) - s_d, \dots, x_{|A|-1,j}(p) - x_{|A|,j}(p) - s_d\right)^T \in \mathbb{R}^{|A_2|},$$

$$z_{2,j}'(p) \coloneqq \left(v_{n,j}(p) - v_{m+1,j}(p), \dots, v_{|A|-1,j}(p) - v_{|A|,j}(p)\right)^T \in \mathbb{R}^{|A_2|}$$

Next, the optimizer of the P-horizon MPC at each control sample point j is given by Equations (12) to (16), in which the control variables are $\boldsymbol{u_j} = \{u_{i,j}(p-1)\}_{i=m+1,p=1}^{|A|,P} \in \mathbb{R}^{P|A_2|}$. In order to keep the stability of the P-horizon MPC, we introduce adjusting parameters $\varphi \coloneqq (\varphi_1, \dots, \varphi_{|A_2|})$, where $\varphi_i > 0$ for $i = 1, \dots, |A_2|$. We can define a diagonal matrix $D_\varphi = \mathrm{diag}(\varphi_1, \dots, \varphi_{|A_2|})$. Then, let $Q_\varphi \coloneqq R^T D_\varphi R$ be a symmetric and positive definite matrix, where R is an orthogonal matrix, which formats the Q_φ as a similar structure to matrices Q_z and $Q_{z'}$ and helps to simplify the stability analysis of this P-horizon MPC.

Optimizer II

$$\begin{aligned} \textbf{Min } \Gamma_2(\boldsymbol{u_j}) \coloneqq & \frac{1}{2}\left[{z_{2,j}}^T(P) Q_z z_{2,j}(p) + \left(z_{2,j}'(P)\right)^T Q_{z'} z_{2,j}'(P)\right] \\ & + \sum_{p=0}^{P-1}\left\{\frac{1}{2}\left[{z_{2,j}}^T(p)(Q_\varphi Q_z) z_{2,j}(p) + \left(z_{2,j}'(p)\right)^T (Q_\varphi Q_{z'}) z_{2,j}'(p)\right]\right\} + \frac{\Delta t^2}{2}\left\|\boldsymbol{u_j}\right\|_2^2 \end{aligned} \tag{12}$$

Subject to: for each $p = 1, \dots, P$

$$a_{min,i} \le u_{i,j}(p-1) \le a_{max,i}, \quad i \in A_2 \tag{13}$$

$$v_{min} \le v_{i,j}(p) \le v_{max}, \quad i \in A_2 \tag{14}$$

$$x_{i-1,j}(p) - x_{i,j}(p) \ge L_i + v_{i,j}(p)\,\Delta t - \frac{\left[v_{i,j}(p) - v_{min}\right]^2}{2a_{min,i}}, i = m+2, \dots, |A| \tag{15}$$

$$x_{n,j}(p) - x_{m+1,j}(p) \ge L_{m+1} + v_{m+1,j}(p)\,\Delta t - \frac{\left[v_{m+1,j}(p) - v_{min}\right]^2}{2a_{min,i}} \tag{16}$$

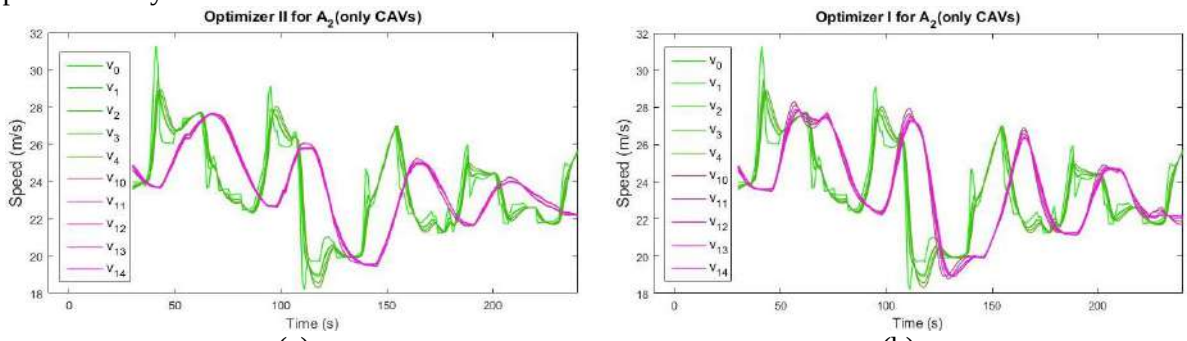
Important properties of the optimizer I such as sequential feasibility and nonempty interior of the constraint set and solution uniqueness have been proved through mathematical analysis in our previous study [3]. This study also proved these properties for the optimizer II by several key theorems and lemmas, which helped us establish the distributed algorithm.

3.3. Distributed algorithm, stability, and numerical experiments

This section develops distributed algorithms for solving the Optimizers I and II above. The main idea of this algorithm is to enable each CAV to iteratively compute and update its local control law at next one step or P steps according to other CAVs' control laws until the solution converges to the system optimal solution. Our experimental study shows that the dual-based distributed algorithm developed in [3] can solve Optimizer-I nicely but hits huge difficulty to solve the Optimizer II with P steps of prediction since it leads to very small step size at the step of solving the primal problem. Therefore, this study develops an improved dual-based distributed algorithms according to the features of the problem formulations. Mainly, the previous gradient projection algorithm for solving the primal problem in [3] is replaced by a more efficient distributed algorithm which converges to the global optimal solution of the primal problem. The convergence of the distributed algorithms is proved by both the mathematical analysis and numerical experiments. Furthermore, we analyze the asymptotic string stability under the unconstrained closed-loop condition and provide important guidance to select the penalty weights used in the objective function of the optimizers in MPC.

The numerical experiments based on the field data were conducted to validate the effectiveness and efficiency of the platoon control scheme. Mainly, the experiments compared the performance of a mixed flow platoon under two different control strategies. (i) Applying the cooperative platoon control. Namely, the experiments conduct one-step MPC for the CAV platoon A_1 and P-step MPC for the CAV platoon A_2. (ii) Applying independent one-step MPC for each CAV platoon. Namely, the experiments conduct one-step

MPC for both A_1 and A_2 excluding the movement anticipation of HDVs. One of the results is shown in Figure 3 due to the limit of the space. Clearly, the proposed control scheme can mitigate traffic oscillation more efficiently. Our other experiment results also demonstrate that the cooperative platoon control can mitigate speed fluctuation along the platoon more effectively under most of the frequencies. Meanwhile, the proposed distributed algorithm converges to the optimal solution quickly, thus it fits the online application very well.



(a) (b)

Figure 3. Speed fluctuation under different control schemes (green/purple lines for platoon A_1/A_2)

4. Summary

This research works on a platoon mixed with CAVs and HDVs on a straight roadway. Without loss of generality, we consider that the sample mixed platoon consists of three sub-platoons, including two CAV platoons separated by a HDV platoon. We also assume that all CAVs in the mixed platoon are well connected. The car-following behavior of the human drivers is modeled by using the Newell car-following model with the parameters: reaction time and minimum stop distance. These two unknown parameters are anticipated by the curve matching learning algorithm using online vehicles trajectories data and the features of the Newell's car-following model. Built upon the learning algorithm, this study proposes cooperative MPC for the two CAV platoons to ensure both transient smoothness and asymptotic stability of the entire platoon, while still keep the mobility of individual vehicles. Specifically, while the downstream CAV platoon A_1 is controlled by the one-step MPC, this study uses the P-step constrained MPC for the upstream CAV platoon A_2, in which P is determined by the anticipation of HDVs' real-time reaction time. According to the mathematical features of the problem formulations, an improved dual-based distributed algorithm is developed to solve the MPCs. Its convergence is proved by both the mathematical analysis and the experimental results. Extensive numerical experiments based on the field data indicate that the distributed algorithm can solve the one-step and P-step MPCs efficiently, and the cooperative platoon control scheme outperforms the control strategy which independently using one-step MPC developed in [3] to each individual CAV platoons. The cooperative MPC can dampen traffic oscillation propagation and stabilize the traffic flow more efficiently for the entire mixed flow platoon. Overall, this paper presents a major advancement in addressing an important transportation problem by well integrated rigorous optimization, control, and traffic flow techniques. It will significantly contribute to CAV technologies in both intellectual merit and practical applications.

Acknowledgments

This research is partially supported by the National Science Foundation awards CMMI-1436786 and CMMI-1554559.

Reference

[1] Zhang, Z. (1994). Iterative point matching for registration of free-form curves and surfaces. International journal of computer vision, 13(2), 119-152.

[2] Camacho, E. F., Alba, C. B. (2013). Model predictive control. Springer Science & Business Media.

[3] Siyuan Gong, Jinglai Shen, Lili Du (2016). Distributed computation based car-following control integrating optimal system performance for a platoon of autonomous vehicles. *Transportation Research Part B: Methodological* 94, 314-334.

Traffic & Mobility

FB2: Network Design and Location Decisions

Friday 1:00 – 2:30 PM

Session Chair: Yanfeng Ouyang

1:00 Emission-controlled Pavement Management Scheduling
Umit Tursun, Hasan Ozer*
University of Illinois

1:30 Scheduling Work Zones in Transportation Service Networks
Dening Peng, Pitu Mirchandani*
Arizona State University

2:00 Reliable Sensor Location for Object Positioning and Surveillance via Trilateration
Kun An, Siyang Xie, Yanfeng Ouyang
Affiliation?

Emission-controlled pavement management scheduling

Management of highway pavement requires considering an extensive set of structural, performance and environmental impacts simultaneously from both agency and user perspectives. Framework for employing mathematical decision modeling to address analytical approach need addressing life cycle impacts. This talk describes an analytic approach that can be used to evaluate and propose rehabilitation schedules based on economic, performance, and environmental considerations for various types of pavements managed by the Illinois State Tollway Authority. A mixed-integer nonlinear program (MINLP) is formulated to model the agency's life-cycle cost and environmental impacts where the decision set consists of the maintenance overlay type and the thickness of the overlay proposed on a temporal scale over a planning horizon. The problem requires interaction of the integer and continuous variables that leads to MINLP formulation. Considering practical implications of the problem, the continuous variables are bounded into a finite and discrete set, while the integrality constraints are relaxed. The objective and constraints of the problem can be alternated to suit the needs of the agency, which may be interested in minimizing environmental impacts and restricting the cost to the agency, or vice versa, over the pavement life-cycle.

Scheduling Work Zones in Transportation Service Networks

Dening Peng *

School of Computing, Informatics & Decision Systems Engineering

Arizona State University

699 S. Mill Ave., Suite 553, Tempe, AZ, USA, 85281

Phone: 1-480-559-4378, Email: Dening.Peng@asu.edu

Pitu Mirchandani

School of Computing, Informatics & Decision Systems Engineering

Arizona State University,

699 S. Mill Ave., Suite 553, Tempe, AZ, USA, 85281

Phone: 1-480-965-2758, Email: Pitu@asu.edu

Submitted on Jan. 4, 2016

*Corresponding Author

Abstract

The repair and maintenance of road network results in “work zones”, where some lane segments of a link are out of commission for a predicted period of time until the work is completed. Temporary link capacity reductions caused by lane closure can result in significant delays of commuters and transport services.

Road construction companies and transportation management agencies do a reasonable job of coordinating work zone activities after the work zone is initiated through appropriate scheduling and staging of day-to-day and week-to-week operations, so that the overall cost is contained, while safety and traffic congestion is not overly affected during peak periods. While a single, or few widely scattered concurrent work zones, will not have a large effect upon travel patterns, several work zones that are spatially and temporally close together, and which affect large flows of traffic, may lead to traffic patterns that are both costly to the travelers and vehicle-based services, resulting in significant negative environmental and safety consequences. In 2010, there were 87,606 crashes in work zones and 526 of these crashes were fatal (FHWA, 2013). While large trucks accounted for only 4% of all registered vehicles in the United States, 27% of work zone fatal crashes involved at least one large truck (FHWA, 2013). Through proper scheduling of work zones with respect to the spatial locations in the network and the time periods of the workzones, a reduction of negative impacts is expected.

In this paper, a network that provides transport logistics services (e.g., freight and parcel delivery) from origins to destinations is considered, where the overall objective is to minimize system transportation costs. The traveling cost of a link is treated as the cost in general sense, which can be interpreted as combinations or functions of travel time, monetary cost, and road unsafety. Origin-Destination (OD) flows of trucks are reactive to the network topology changes resulted from temporary lane closures, which means the routes chosen by each vehicle could change in response to work zone activities to minimize the total flow costs. The goal is to find the optimal lane closure schedule that has the minimum negative impact (e.g. gas usage, congestion) on the trucks’ flows in the network.

In work zone scheduling in transportation service network (WZS-TS) model developed, it is assumed that there are Origin-Destination (OD) flow demands of trucks every time period (e.g., peak period of a day). Each truck can choose its own route to minimize its travel cost and is treated as a unit of flow. The WZS-TS model possesses the features of both scheduling models and multi-commodity flows models. For each feasible schedule on each day, there is a multi-commodity flows problem over the network based on links’ available capacities after the scheduled lane closures. With lane closure schedules fixed, scheduling variable become parameters that define the links’ available capacities for each day. And the remaining problem is a multi-commodity flows problem for each day over the planning time horizon. In the multi-commodity flows subproblem of WZS-TS, the available link

capacity can be exceeded with a very high extra flow cost. This is different from the traditional multi-commodity flows problem, where link capacity constraints are hard constraints.

The travel cost on a link contains two parts: the regular flow cost for all the flow units traveling on the link, plus the extra flow cost for the flow units on links where the flow exceeds capacities. When a link is under maintenance, one or more lanes are closed, resulting in the temporary link capacity reductions, and thus the link may not be fully serve the flows without additional cost. Suppose a link has three lanes and all three lanes have the same "flow capacity" u, when two lanes are closed for maintenance, the available capacity of the link is u. If the units of flows on the link is more than u, then the extra flow cost will be incurred. Similarly, in the case when one lane is closed for maintenance, the available capacity of the link is $2u$ and the extra flow cost will be incurred if the flow units on the link are more than $2u$. When some of the lanes in a link is closed for maintenance, some of the flow that is originally on this link could divert to other paths and links to reach the destination and avoid the extra flow cost, so as to achieve the system-wide lower total cost, that is, the network flows are *reactive* to the maintenance schedules.

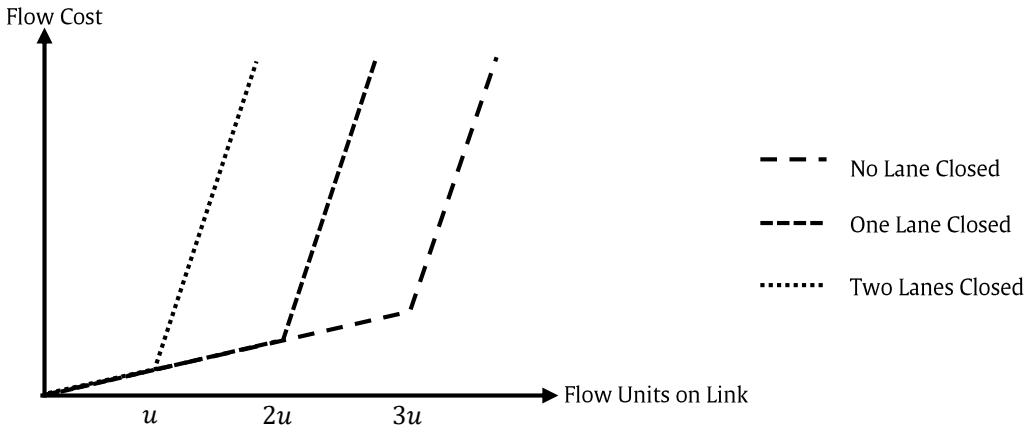


FIGURE 1: Three-Lane Link Flow Cost Curve.

Figure 1 illustrates this concept of approximating the congestion cost due to work zones as piecewise linear. In reality, the congestion cost would be nonlinear, to capture the gradual increase in cost due to service vehicles and commuters, and congestion functions for links need to be developed from data that is only available before the links are worked on, clearly not easily doable. This paper provides only an approach to get a reasonable schedule which does take into account gradual congestion, as opposed to neglecting it by using hard constraints. The results of this paper are being extended to the nonlinear case that more explicitly accounts for congestion due to all traffic.

Even et al. (1975) proved that the decision version of the multi-commodity flow problem is NP-

Complete even for only two commodities and unit capacities. By setting planning horizon one day and no link needs to repair, the problem discussed in this paper is converted to a multi-commodity flow problem, which proves that this problem can be reduced from multi-commodity flow problem in polynomial time. Thus the WZS-TS problem is strongly NP-Hard and needs an efficient heuristic to solve the problem quickly with satisfactory accuracy.

There are two layers of problems that constitute the problem of work zone scheduling in transportation service networks. The upper layer is the scheduling problem which decides the repair start date for each lane of the links that need maintenance. The lower layer is a series of multi-commodity flow problems based on the available capacities of links on each day, which is determined by the current lane closures. Once the schedule is set, solving the multi-commodity flow problems for each day is a relatively easier since the flow variables are all continuous variables. And thus the solution approach proposed in this paper focuses on the upper layer of obtaining good work zone schedules.

The solution method developed for this research problem is a decomposition technique named randomized fix-and-optimize (RFO) heuristic. Based on a feasible schedule, RFO groups work zones into subsets randomly, and iteratively optimizes the schedule of work zones in one subsets with the schedules of work zones in other subsets fixed. The number of subsets which the work zones are randomly grouped into can increase as needed to control the size of subproblems, so that good schedules can be found within short amount of time.

To motivate the heuristic, suppose at a point in the algorithmic process a feasible schedule is obtained which has some aspects similar to the optimal schedule. The feasible schedule probably has the schedule of some lane closures the same as those in the optimal schedule. Hence, if it is only needed to optimize the lane closure schedules that are different from the optimal schedule while keeping the schedules of all the other lanes fixed, the problem size will be much smaller and the time needed to solve the problem instance will reduce dramatically since there are much fewer integer variables to go through in a branch-and-bound process performed by solvers like CPLEX. This observation leads to the adoption of the *fix-and-optimize* heuristic as the core of the solution approach.

The basic process of the fix-and-optimize heuristic is to decompose the integer variables into subsets, based on an initial solution, and then optimize the values of a subset of integer variables together with all continuous variables while the values of the other integer variables in other subsets are fixed (this is a "subproblem" of the fix-and-optimize procedure). If the new objective function value is better than current best objective value, then the current candidate optimal values are updated. The RFO heuristic iterates through this process until a specified stopping criteria is met.

In the problem of scheduling work zones in transportation service networks, only the schedules that consider all work zones will have the lowest increase in total flow cost, because OD demands

happen over the whole network and each OD pair has network-wide minimum cost routing. This means applying fix-and-optimize heuristic with small subsets of work zones (one or two links) will hardly find satisfactory schedules since it only considers the maintenance of a few links at a time. However, if the size of the work zone subsets is large, the size of each fix-and-optimize subproblem will also be large and it would take a very long time to solve. To mitigate the conflict between solution quality and solving time length, the fix-and-optimize procedure with varying subset sizes and the use a truncated branch-and-bound method is developed.

The RFO uses CPLEX to solve the entire problem within a given time limit. If the problem is solved optimally, then the optimal schedule will be output and the program will terminate. If the problem is not solved optimally, the best feasible schedule obtained so far will be stored and used as the initial feasible solution for the fix-and-optimize procedure. A feasible schedule is that which is able to both complete all the maintenance work before the specified completion date and make sure each OD demands can be met.

The randomized fix-and-optimize (RFO) iteration starts with randomly dividing links that need maintenance into two subsets and solving each fix-and-optimize subproblem (FO subproblem) with a specified time limit. A RFO iteration is finished when the schedules of all the generated subsets of links are optimized. The RFO will be performed for a preset number of iterations and if any of the FO subproblems is not solved within the time limit in the last iteration, the RFO will enter a new stage where the number of subsets which the links to repair are randomly divided into is three. The RFO proceeds similarly in stages with more subsets of links and each RFO iteration is performed the same way as it is in the initial stage when there are only two subsets. And the number of subsets which the links are randomly decomposed into will increase as long as one of the FO subproblems is not solved within the time limit.

The reason of randomly grouping links that need maintenance into subsets is because the set of links with schedules that are different from the optimal schedule are not known since the optimal schedule is not known. Also, consideration of various OD demand patterns, and flows being reactive to network capacity changes, makes it formidable to pin-point the links that can have better schedule through classical network flows optimization models. Hence random grouping is applied to explore various combinations of links for better schedules. Both the decomposition of the links based on the required number of days to repair and decomposition based on links' unit flow cost were tested, but both of them had inferior performance compared to the random grouping approach. Through the iterative randomized fix-and-optimize process, the work zone schedule changes gradually towards the optimal schedule.

To illustrate the advantage RFO has over CPLEX, the performance comparison between RFO and CPLEX is made on various tests cases created based on three different networks. For the problem instances tested, RFO obtains optimal or near-optimal solutions with much less time compared to solving the WZS-TS problem solely with CPLEX. The authors note that optimal repair and maintenance of other flow carrying service networks, such as power networks, water distribution networks, and communication systems, may benefit from the model and solution method developed in this paper.

References

Even, S., Itai, A., Shamir, A. (Oct. 1975). *On the complexity of time table and multi-commodity flow problems.* Paper presented at the 16th Annual Symposium on Foundations of Computer Science, Washington, DC, USA.

FWHA. (September, 2013). Facts and Statistics - Work Zone Delay. *Facts and Statistics.* Retrieved fromhttp://www.ops.fhwa.dot.gov/wz/resources/facts_stats/delay.htm#fn1

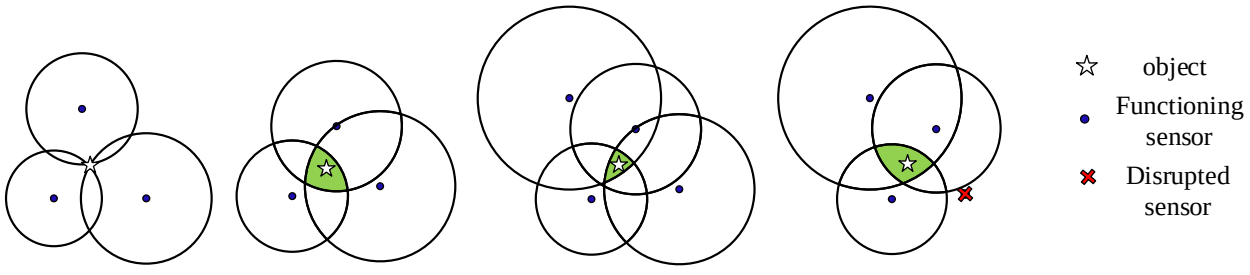
Reliable Sensor Deployment for Object Positioning and Surveillance in A Two-Dimensional Space

Kun An, Siyang Xie, Yanfeng Ouyang*
Department of Civil and Environmental Engineering, University of Illinois at Urbana - Champaign, Urbana, IL 61801, United States

Introduction

High-accuracy object positioning based on satellites (e.g. Global Positioning System (GPS)) or cellular phones has been playing a critical role in various application contexts such as vehicle navigation, driver guidance, aircraft tracking, extrasolar planets detection and missile and projectile guidance. Location detection is generally less accurate inside a building due to severe attenuation or blockage of satellite or phone signals. In recent years, massive availability of mobile devices has stimulated demand for indoor location-aware applications, including in-building guidance and location-based advertising in shopping malls, elderly navigation in nursing homes, hazardous materials detection in airports, and workforce tracking (policemen, firefighters). Indoor positioning systems utilize radio waves, Wi-Fi, Bluetooth, magnetic fields or other sensory information collected by the mobile devices. The effectiveness of such a positioning system highly depends on the quality (working range and precision level) and quantity of sensor coverages in the spatial area. Nevertheless, high-precision sensors are generally expensive to deploy, and installed sensors could be disrupted from time to time due to technological defects or deliberate sabotages. It remains a challenge to optimize sensor deployment locations that maximize the overall system-wide surveillance or positioning benefits under the risk of sensor failures.

Trilateration is one of the most popular mathematical techniques that many positioning systems have used to geographically position an object. For example, in a Wi-Fi positioning system, received signal strengths from all existing Wi-Fi access points are gathered and converted to distances using existing signal propagation models (each positioning system may have its own distance calculation method). With the distance information, a trilateration algorithm is used to pinpoint the user location, typically based on distances data from at least three different sensors, as shown in Figure 1(a). In effect, as shown in Figure 1(b), due to signal scattering and blockage, the collected or calculated distance information may be inaccurate such that a single location cannot be precisely identified, and the error increases with the distance between the sensor and the object. It is worth noting that the error can be reduced if information from more sensors is used, as shown in Fig. 1(c), but the system architecture becomes more complex and the trilateration algorithm requires more time. Hence the number of sensors used should be carefully determined such that a certain level of accuracy can be achieved in a reasonable calculation time. Moreover, if a sensor is disrupted as shown in Fig. 1(d), a more remote sensor can be used and yet with a larger error. Hence, sensors should be assigned to the object at different backup levels such that the overall expected error across various sensor failure scenarios is minimized.



(a) Ideal case (b) Three sensors (c) Multiple sensors (d) Disruption scenario

Fig. 1 Position error illustration in trilateration.

Extensive studies have been conducted to determine the optimal deployment of sensors in a discrete network. Among them, Geetla et al. (2014) studied the omnidirectional sensor location problem for emergency responses. Audio sensors are deployed to detect vehicle crashes on a road or at an intersection within the detection range. Li and Ouyang (2011; 2012) investigated reliable traffic sensor deployment to estimate OD flow volume, congestion state, and path travel time. Adjacent sensors along each flow path pair up to monitor the road segment in between. These studies consider sensor deployment in either a zero-dimensional space (i.e., sensors work independently to provide coverage) or one-dimensional space (i.e., sensors work in pairs to monitor line segments). Few studies have considered sensor location problem in a two-dimensional plane, where at least three sensors are required to work together to locate a target.

Methodology

We consider an area (e.g. airport, shopping mall) which contains a set of spatial neighborhoods $I := \{i\}$ that need surveillance coverage. In airports, such neighborhoods can be security check gates, boarding gates and restaurants where accidents are more likely to occur due to crowds' gathering. Each point $i \in I$ attracts v_i customers per day. Let J be the set of candidate locations for potential sensor installations. At most one sensor can be installed at each location $j \in J$ at a construction cost f_j. Let d_{ij} denote the distance from surveillance neighborhood i to sensor location j. A sensor located at j could be disrupted with a probability of p_j. We assume the receiver (can be the mobile device/object itself) always uses the nearest N, where $N \geq 3$, sensors to calculate the position of the object. Without loss of generality, for modeling convenience, N dummy sensors (located at $|J|+1,...,|J|+N$) are added to the system to ensure there are always at least N sensors available even under the worst case scenario in which all sensors are disrupted. Let $\tilde{J}$ be the set of dummy sensors and $\vartheta = J \cup \tilde{J}$ be the set of all sensors. The dummy sensors incur 0 installation cost and are not subject to failure, but make no contribution to object positioning; this is modeled by setting $d_{ij} = \infty, \forall j \in \tilde{J}, i \in I$. Let K be the set of candidate sensor combinations to locate customers. Each combination $k \in K$ contains exactly N sensors (including the dummy ones) and could monitor i with accuracy e_{ik}. We introduce incidence matrix $\{a_{kj}\}$ to represent the mapping relationships between combinations and sensors, where $a_{kj} = 1$ if combination k contains sensor j, or 0 otherwise. The maximum number of combinations is $\sum_{t=0}^{N} \binom{|J|}{t}$, where t indicates that t regular sensors and $N-t$ dummy sensors are used in the combination. Similarly, incidence matrix $\{b_{ikj}\}$ is introduced where $b_{ikj} = 1$ if sensor j is the most remote sensor from neighborhood i in combination k, or 0 otherwise.

As such, the receiver/object will search from its nearest sensor until N sensors have been found (or no more sensor is available). The key decision variables $\mathbf{X} := \{X_j\}$ determine sensor locations, where $X_j = 1$ if a sensor is installed at location j or $X_j = 0$ otherwise. For each surveillance neighborhood, the installed sensors are assigned to it at different levels. Variables $\mathbf{Z} := \{Z_{ijr}\}$ determine the relative sensor levels, where binary variable $Z_{ijr} = 1$ if sensor j is the r^{th} nearest installed sensor to neighborhood i, or 0 otherwise; $\mathbf{Y} := \{Y_{ikr}\}$ denote the sensor combination assignment to the customers, where $Y_{ikr} = 1$ if neighborhood i uses combination k whose most remote element sensor has level r, or 0 otherwise. Note that a combination k corresponds to only one level r, while there may exist multiple combinations corresponding to the same level r. P_{ikr} is the probability for Y_{ikr} =1 to happen, and $\mathbf{P} = \{P_{ikr}\}$.

The problem can now be formulated as the following mixed-integer non-linear program:

$$\min_{\mathbf{X,Y,Z,P}} \sum_{j\in J} f_j X_j - \sum_{i\in I}\sum_{k\in K}\sum_{r=1..|\vartheta|} v_i e_{ik} P_{ikr} Y_{ikr} \tag{1}$$

$$\text{s.t.} \sum_{r=1..|\vartheta|} Z_{ijr} \le X_j \quad \forall j\in J, i\in I \tag{2}$$

$$\sum_{r=1..|\vartheta|} Z_{ijr} = 1 \quad \forall j\in \tilde{J}, i\in I \tag{3}$$

$$Z_{ijr} = Z_{i,j+1,r+1} \quad \forall r = 1...|\vartheta|-1,\ j\in \tilde{J}\setminus|J|+N, i\in I \tag{4}$$

$$\sum_{j\in J} Z_{ijr} + \sum_{s=1}^{r} Z_{i,|J|+1,s} = 1 \quad \forall r = 1...|\vartheta|, i\in I \tag{5}$$

$$\sum_{j\in J} Z_{ijr} d_{ij} + \sum_{s=1}^{r} Z_{i,|J|+1,s} d_{i,|J|+1} \le \sum_{j\in J} Z_{ij,r+1} d_{ij} + \sum_{s=1}^{r+1} Z_{i,|J|+1,s} d_{i,|J|+1} \quad \forall r = 1...|\vartheta|-1, i\in I \tag{6}$$

$$Y_{ikr} \le \frac{1}{N}\sum_{j'\in\vartheta}\sum_{r'=1..|\vartheta|} a_{kj'} Z_{ij'r'} \quad \forall k\in K, r = 1...|\vartheta|, i\in I \tag{7}$$

$$Y_{ikr} \le \sum_{j\in\vartheta} b_{ikj} Z_{ijr} \quad \forall k\in K, r = 1...|\vartheta|, i\in I \tag{8}$$

$$P_{ikr} = \sum_{j\in\vartheta} p_j Z_{ijr} P_{ikr-1} \quad \forall k\in K, r = 1...|\vartheta|, i\in I \tag{9}$$

$$P_{ik0} = \prod\nolimits_{j\in\vartheta} \left(1-p_j\right)^{a_{kj}} \left(p_j\right)^{-a_{kj}} \quad \forall k\in K, i\in I \tag{10}$$

$$X_j, Z_{ijr}, Y_{ikr} \in \{0,1\} \quad \forall k\in K, j\in\vartheta, r = 1...|\vartheta|, i\in I \tag{11}$$

The objective function (1) presents the expected system cost including the sensor installation cost and the expected total inaccuracy penalty. Constraints (2) enforce that customers can only use installed sensors. Constraints (3) ensure for a certain surveillance neighborhood, each dummy sensor must be assigned to it at a certain backup level. The same dummy sensor could be assigned to other surveillance neighborhoods at different levels. Constraints (4) postulate that if a dummy sensor $j\in\tilde{J}$ is assigned to surveillance neighborhood i at level r, then dummy sensor $j+1$ must be assigned to i at level $r+1$. Constraints (5) require that at each level *r*, a surveillance neighborhood i either uses a regular sensor, or it has used the first dummy sensor at level $s\le r$. Constraints (6) state that for each surveillance neighborhood i, a nearer sensor must be assigned to an earlier level. Constraints (7) enforce that combination k is available to surveillance neighborhood i only if the N sensors in k are all installed. Constraints (8) require that combination k is available to surveillance neighborhood i when its element j serves at level r. Constraints (9) and (10) recursively define the assignment probability P_{ikr}. For modeling convenience, the failure probabilities of dummy sensors are set as a small constant ε (i.e. $p_j = \varepsilon \ \forall j\in\tilde{J}$). Given that the nearer sensors are used earlier, a combination k is used if and only if its element sensors are all functioning, and the other constructed sensors closer than its most remote sensor (which has the highest level in k) are all disrupted. The rigorous derivation of P_{ikr} is provided in the full paper. The current model is nonlinear due to the existence of nonlinear terms $P_{ikr}Y_{ikr}$ in (1) and $Z_{ijr}P_{ikr-1}$ in (9). Linearization techniques similar to those in Li and Ouyang (2012) are also described in the full paper.

Preliminary results

We use 3-by-3 and 3-by-4 networks to demonstrate the applicability of the model. The candidate sensor

locations (evenly distributed in the region), and the surveillance neighborhoods are shown in Fig. 2. Each combination uses $N=3$ sensors. Accuracy is computed based on $e_{ik}=\sum_{j\in J}\frac{a_{kj}}{\left(d_{ij}\right)^2+\varepsilon}\ \forall i\in I,k\in K$, where d_{ij} is the Euclidean distance. For simple illustration, we assume each facility is subject to an equal failure probability $p=0.1$. P_{ikr} can thus be simplified as $P_{ikr}=(1-p)^{\sum_{j\in J}a_{kj}}p^{r-3}$ where $\sum_{j\in J}a_{kj}$ is the number of regular sensors in combination k. This way, P_{ikr} becomes a constant for each r and can be preprocessed and the model becomes linear. The model is programed in C++ and is solved by CPLEX. The computation time for the 3-by-3 and 3-by-4 cases are 360 and 1820 seconds, respectively. The installed sensors are marked as green in Fig 2.

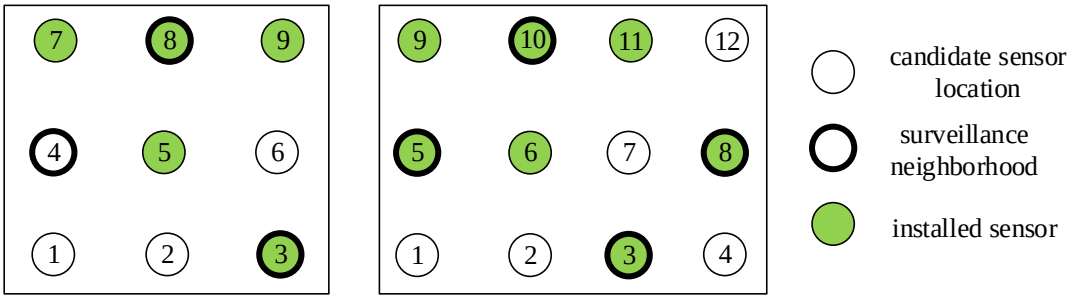


Fig. 2. Optimal sensor deployment for the 3-by-3 and 3-by-4 cases.

Fig. 3 illustrates how the sensor combinations are used by customers in neighborhood $i=3$ in the 3-by-3 case. The installed sensors 3, 5, 9, 8, 7 are assigned to levels 1 to 5 based on distance, while the dummy sensors are assigned at levels 6 - 8. Some representative combinations are illustrated in this figure. For example, the shaded combination (*k*=*u*) will be used to monitor this neighborhood if and only if sensors 5, 9 and 8 are functioning and sensor 3 has been disrupted. The most remote sensor in this combination is 8, which is ranked at level *r*=4. Hence combination *u* corresponds to backup level *r*=4 and it will be used with a probability of $P_{3u4}=(1-0.1)^3 0.1^{4-3}=0.0729$.

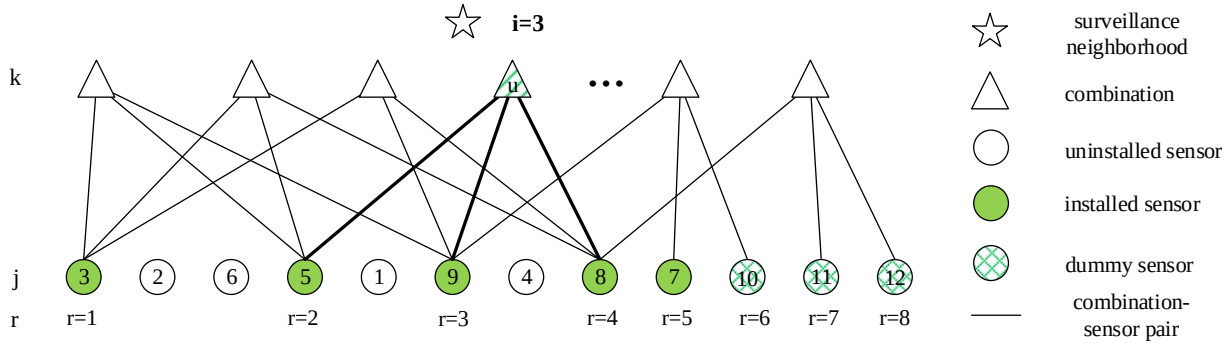


Fig. 3. Assignment scheme of sensor combinations.

Although CPLEX performs well in these 3-by-3 and 3-by-4 networks, its performance deteriorates rapidly as the problem size increases. It cannot provide an optimal solution for even a 4-by-4 network with an equal failure probability within several hours of computation time. In the full paper, we also present customized algorithms based on Lagrangian relaxation (LR) to improve the solution efficiency. Using LR, we relax constraints (2) to decouple the correlation among the surveillance neighborhoods, and this helps decompose the original problem into two sub problems, which can be solved separately. Also, a real-world application with site-dependent sensor failures will be presented in the paper to demonstrate

the applicability of the model, the efficiency of the proposed algorithms, and insights and practical recommendations.

Incomplete References

Geetla, T., Batta, R., Blatt, A., Flanigan, M., & Majka, K. (2014). Optimal placement of omnidirectional sensors in a transportation network for effective emergency response and crash characterization. *Transportation Research Part C: Emerging Technologies, 45*, 64-82.

Li, X., & Ouyang, Y. (2011). Reliable sensor deployment for network traffic surveillance. *Transportation research part B: Methodological*, 45(1), 218-231.

Li, X., & Ouyang, Y. (2012). Reliable traffic sensor deployment under probabilistic disruptions and generalized surveillance effectiveness measures. *Operations Research, 60*(5), 1183-1198.

Traffic & Mobility

FC2: Traffic Flow Modeling

Friday 2:45 – 4:15 PM

Session Chair: Hillel Bar-Gera

2:45 **The Role of Stochasticity in Traffic Flow Instabilities**

[1]Junfang Tian, [2]Rui Jiang, [3]Martin Treiber*

[1]Institute of Systems Engineering-College of Management and Economics-Tianjin University, [2]MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology-Beijing Jiaotong University, [3]Technische Universität Dresden-Institute for Transport & Economics

3:15 **A Kalman Filter Approach for Dynamic Calibration of A Simplified Lower-Order Car Following Model**

Kerem Demirtas, Pitu Mirchandani, Xuesong Zhou*

Arizona State University

3:45 **Representation Requirements for Perfect First-In-First-Out Verification in Continuous Flow Dynamic Models**

Hillel Bar-Gera, Malachy Carey*

Ben-Gurion University of the Negev, Ulster University

The role of stochasticity in traffic flow instabilities

Junfang Tian[1], Rui Jiang[2], and Martin Treiber[3]

[1]Institute of Systems Engineering, College of Management and Economics, Tianjin University
No. 92 Weijin Road, Nankai District, Tianjin 300072,China
e-mail address: jftian@tju.edu.cn
[2]MOE Key Laboratory for Urban Transportation Complex Systems Theory and Technology, Beijing Jiaotong University
No.3 Shangyuancun, Haidian District, Beijing 100044, China
e-mail address: jiangrui@bjtu.edu.cn
[3]Technische Universität Dresden, Institute for Transport & Economics, Würzburger Str. 35, D-01062 Dresden, Germany
e-mail address: treiber@vwi.tu-dresden.de

This paper investigates the role of stochasticity in traffic flow instabilities, which is a typical self-driven system far from equilibrium. We demonstrate that stochasticity is necessary to correctly describe the observed spatiotemporal dynamics of growing traffic oscillation in the car following process. A heuristic analysis qualitatively explains the concave growth of the oscillation amplitude along the vehicles of a platoon. Based on this analysis, we propose a simple car-following model containing indifference regions and acceleration noise described by Brownian motion which reproduces well the experimental and empirical observations.

Key words: car following; stochasticity

The emergence and evolution of oscillations as well as their development into jams is an intriguing pattern formation phenomenon in traffic flow, which is an undesired feature since stop-and-go movement is a nuisance to motorists, consumes more fuel, and likely causes more accidents. The true origin of these traffic waves, however, is still not clearly revealed. From the 1950s', scientists always explain this phenomenon via linear stability theory. In most of the traffic models named as two-phase models, it is assumed that in the steady state, there is a one-to-one bivariate relationship $u_e(\rho)$ between mean speed u and traffic density ρ, or relationship $V(\Delta x)$ between individual speed v and spacing Δx, defined as the gross distance between two neighboring vehicles (sum of bumper-to-bumper distance and vehicle length). Stability analysis shows that, in a certain density range, traffic flow is metastable or unstable. Oscillations emerge due to disturbances in the metastable and unstable traffic flows. The oscillations could grow and develop into jams via a subcritical Hopf bifurcation.

Nevertheless, this explanation has been questioned in the last two decades. Based on the analysis of empirical detector data, Kerner classified the congested traffic into synchronized flow and jam, and proposed the three-phase traffic theory. In this theory, the steady state of synchronized flow is assumed to occupy a two-dimensional region in the flow-density plane rather than a unique relationship. The formation and evolution of oscillations have been explained by the observed compression of traffic at the location of breakdown. However, the underlying mechanism of this so-called "pinch effect" is unclear.

Recently Jiang et al. carried out a car following experiment on a 3.2-km-long open road section, in which a platoon of 25 passenger cars has been studied. The leading vehicle was asked to move with constant speed. Stripe structure has been observed in the spatiotemporal evolution of traffic flow, which corresponds to the formation and development of oscillations. It has been found that the standard deviations of speed increase in a concave way along the 25-car-platoon. Later, this concave growth pattern has also been found in real traffic flow. Moreover, Jiang et al. have shown that (i) the simulation results of some representative two-phase models, including General Motor models (GMs), Gipps' Model, Optimal Velocity Model (OVM), Full Velocity Difference Model (FVDM), and Intelligent Driver Model (IDM), run against the experimental finding since the standard deviation initially increases in a convex way in the unstable density range; (ii) by removing the fundamental diagram in these two-phase models and allowing the traffic state to span a two-dimensional region in velocity-spacing plane (indifference region) and additionally introducing stochastic elements,, the growth pattern of disturbances has changed and becomes qualitatively or even quantitatively in accordance with the observations.

We argue that some form of noise (or stochasticity) in the car-following process is necessary for a correct description of the pattern formation of oscillation growth. Next, we demonstrate the argument from heuristic analysis and simulation results. Firstly, we add white noise to the acceleration equation of time-continuous two-phase car following models resulting in

$$a_n(t) = a\left(d_n(t), v_n(t), \Delta v_n(t)\right) + \psi_n(t) \tag{1}$$

$$<\psi_n(t)>=0, <\psi_n(t)\psi_{n'}(t')>=Q\delta(t-t')\delta_{nn'} \tag{2}$$

where, $a\left(d_n(t), v_n(t), \Delta v_n(t)\right)$ is the acceleration function of the underlying deterministic car following model for vehicle n as a function of the vehicle speed $v_n(t)$ at time t, the space gap $d_n(t) = x_{n+1}(t) - x_n(t) - L_{\text{veh}}$ and the relative speed $\Delta v_n(t) = v_{n+1}(t) - v_n(t)$ of the leading vehicle n+1. Here, L_{veh} is the vehicle length and x_n is the position of vehicle n and x_{n+1} is the position of its leader, respectively. Furthermore, $\psi_n(t)$ is the acceleration noise, <y> denotes the expectation value of y, and $\delta(t-t')$ is Dirac's delta function which is defined by $\delta(t-t')=0$ for any $t' \neq t$, $\int \delta(t-t')dt = 1$. δ_{ij} is the Kronecker symbol, $\delta_{ij}=1$ for i=j, and zero, otherwise. Q is the noise intensity with the unit m^2/s^3. Notice that (1) and (2) can alternatively be written as a stochastic differential equation.

$$dv_n(t) = a\left(d_n(t), v_n(t), \Delta v_n(t)\right) + \sqrt{Q}dB_t \tag{1}$$

where dB_t denotes a Wiener process (Standard Brownian motion). Particularly, in the limit $\Delta t \to 0$, the variance of the speed change of a given vehicle is that of a Brownian motion,

$$Var\left(v_n(t+\Delta t) - v_n(t)\right) = Q\Delta t \tag{2}$$

An example of this model class is the stochastic IDM (SIDM), whose deterministic acceleration function (i.e., that of the normal IDM) is defined by

$$a_n(t) = a_{\max}\left(1 - \left(\frac{v_n(t)}{v_{\max}}\right)^4 - \left(\frac{d_n^*(t)}{d_n(t)}\right)^2\right) \tag{3}$$

$$d_n^*(t) = \max\left(v_n(t)T - \frac{v_n(t)\Delta v_n(t)}{2\sqrt{a_{\max} b}}, 0 \right) + d_0 \qquad (4)$$

where $v_{\max}$ is the maximum speed, $a_{\max}$ the maximum acceleration, b the comfortable deceleration, T the desired time gap and d_0 is the jam gap.

Simulations of the platoon configuration by the SIDM (Figure 1) show that with the increase of stochasticity amplitude, the initially convex growth pattern gradually changes into concave growth pattern. Nevertheless, although noise is able to make the growth pattern of SIDM qualitatively agree with the experimental observations, there exist quantitative deviations (Figure 1). We also fitted the deterministic IDM parameters to the driving experiments of Figure 2 but this did not lead to a simultaneous quantitative agreement for all the experiments displayed in this figure. While this is not a proof that a two-dimensional region of steady states is necessary, it points to the difficulties of two-phase models in describing these observations.

We will now show that introducing indifference regions in combination with noise can reproduce the observations.

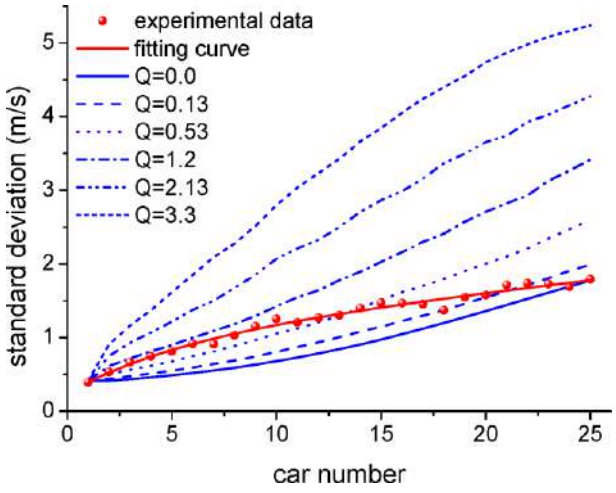


Figure 1. Speed standard deviation of a platoon of 25 vehicles as simulated with the SIDM with the parameters, $a_{\max}$ = $0.5m/s^2$, b = $2.0m/s^2$, T = $1.2s$, $v_{\max}$=$30m/s$, L_{veh}=$5m$, and d_0=$1.5m$. The speed of the leading vehicle $v_{leading}$= $30km/h$.

In order to clarify the mechanism, we first present a heuristic analysis in form of a stylized model with following behavior assumptions: (i) each driver has an indifference region, i.e., a range of gaps to the preceding car where there is no response to changes in the gap; (ii) a driver is able to adapt his/her vehicle's speed to that of the preceding vehicle instantaneously. (iii) the adaptation is subject to independent noise contributions leading to an additional speed stochasticity ξ of mean $\langle\xi\rangle=0$ and variance $\mathrm{Var}\langle\xi\rangle=\sigma^2$.

Assuming there is a platoon consisting of N vehicles on the road, where vehicle 1 is the leader and vehicle N is the end of the platoon. The speed of vehicle n is v_n. Supposing that $v_1= v + \xi_1$, where v is constant speed and ξ_1 represents the stochasticity of vehicle 1. According to the above assumptions, vehicle 2 will move with the speed $v_2 = v_1+ \xi_2$ and vehicle n will move with the speed $v_n = v_{n-1} + \xi_n$, i.e.

$$v_n = v_{n-1}+\xi_n = v_1 +\xi_2+\ldots+\xi_n = v+\xi_1+\ldots+\xi_n \quad (5)$$

Since all $\xi_i (i = 1,..,n)$ are independent, we obtain

$$\langle v_n\rangle = v+n\langle\xi\rangle = v \quad (6)$$

$$\mathrm{Var}(v_n) = n\mathrm{Var}(\xi) = n\sigma^2 \quad (7)$$

Thus, the standard deviation of the speed of vehicle n increases in a concave way along the platoon.

$$\sigma_n = \sqrt{n}\sigma \quad (8)$$

Figure 2 (a)-(e) compares the standard deviation of the time series of the speed of each car in the platoon of Jiang's car following experiments with the results predicted by Equation (8), which demonstrates that the predicted results are highly consistent with the experimental results.

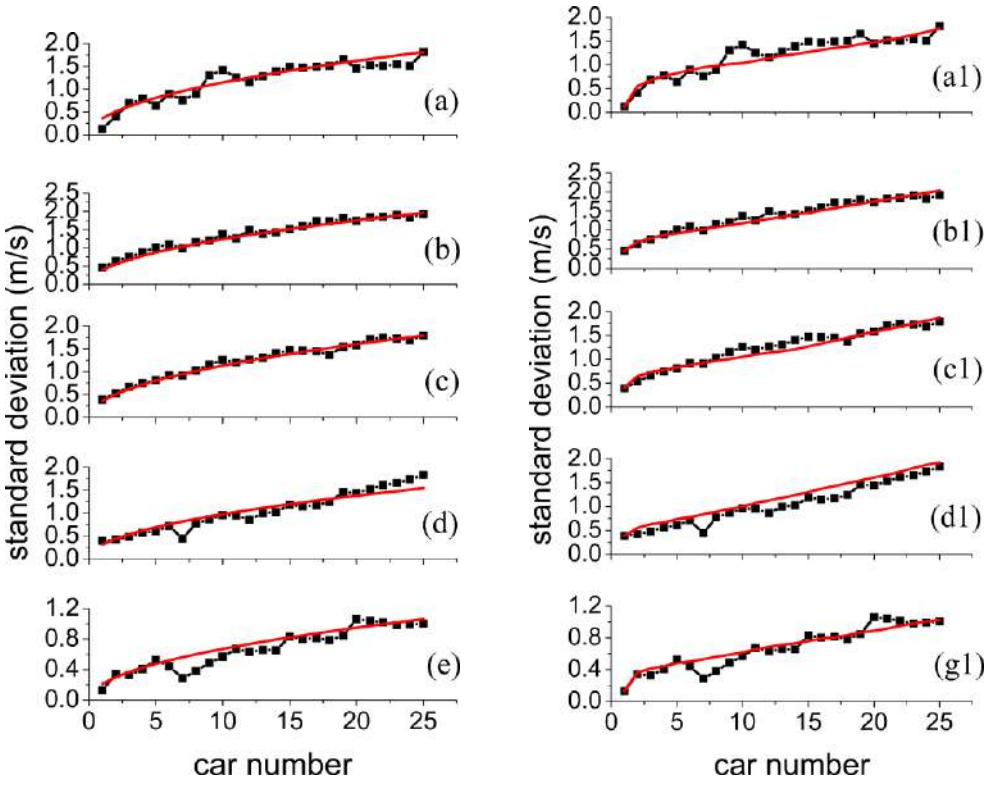


Figure 2. The standard deviation of the vehicle speed in the platoon. The car number 1 is the leading car of the platoon. The symbol solid black lines are the experiment results. In (a)-(e) the solid red lines are obtained by Equation (8), where σ^2 = 0.362, 0.391, 0.359, 0.308 and 0.212$[m/s]^2$. In (a1)-(g1) the solid red lines are the simulation results of the BIDM, where a = 0.53m/s^2, b = 1.7m/s^2, T_1 = 0.5s, T_2 = 1.8s, $v_{\max}$ = 120km/h, A = 0.042 m^2/s^4, d_0 = 1.5m, L_{veh}=5m and Δt = 0.1s. From top to bottom, the leading vehicle of the platoon is required to move with $v_{leading}$=50, 40, 30, 15 and 7km/h, respectively.

To integrate the above "toy model" into the full car-following process, we propose a very simple car following model as follows: (i) when the spacing between vehicles is in the indifference region $[d^1, d^2]$, the acceleration of vehicles will only change stochastically, i.e. their accelerations obey the Brownian movement; (ii) when the spacing is outside the indifference region, the normal deterministic car following acceleration applies. We will denote models incorporating these rules as models with Brownian movement. Taking the IDM as an example, the IDM with

Brownian motion (BIDM) is specified in discrete time with time step Δt, $t=j\Delta t$, as follows:

$$\begin{aligned}
&\text{if } d_n^1(t) < d_n(t) < d_n^2(t) \\
&\quad a_n(t) = \left[a_n(t-\Delta t) + \tilde{a}_{nj} \right]_{-b}^{a_{\max}} \\
&\text{else} \\
&\quad a_n(t) = a_{\max}\left(1 - \left(\frac{v_n(t)}{v_{\max}}\right)^4 - \left(\frac{d_n^3(t)}{d_n(t)}\right)^2\right)
\end{aligned} \tag{9}$$

where $\tilde{a}_{nj}$ denotes the stochastic acceleration part obeying

$$< \tilde{a}_{nj} >= 0 \text{ , and } < \tilde{a}_{nj}\tilde{a}_{n'j'} >= A\delta_{nn'}\delta_{jj'} \tag{10}$$

$\left[y\right]_{y_{\min}}^{y_{\max}} := \max\left(\min\left(x, y_{\max}\right), y_{\min}\right)$ is the bounded function. A is the noise intensity with the unit m^2/s^4. The critical space gaps are defined as follows:

$$d_n^i(t) = \max\left(v_n(t)T_i - \frac{v_n(t)\Delta v_n(t)}{2\sqrt{a_{\max}b}}, 0\right) + d_0 \tag{11}$$

where the index i = 1, 2 and 3. T_i is the time gap and $T_3=(T_1+T_2)/2$ for simplicity.

Figure 2 (a1)-(g1) show that the growth rate of the disturbances is consistent with the experimental data. Therefore, it can be concluded that the BIDM can reproduce Jiang's experiments very well.

Conclusion

Stochasticity has been neglected in traffic flow theory for a long time. In many models, stochasticity is not considered at all or is considered only as a kind of marginal effect, such as acceleration noise. However, this paper demonstrates that stochasticity in connection with indifference zones plays a decisive role in the pattern formation in traffic flow.

We believe that clarifying the role of the stochasticity in the formation and development of oscillations has significant implications in that: (1) it reveals the underlying mechanism of traffic flow instability, i.e. the origin of traffic jam formation; (2) it allows constructing correct models and accordingly right testing bed for traffic technologies alleviating traffic jams, such as adaptive cruise control (ACC) and vehicle-to-X (V2X) technologies.

A Kalman Filter Approach for Dynamic Calibration of A Simplified Lower-Order Car Following Model

Kerem Demirtas
School of Computing, Informatics & Decision Systems Engineering, Arizona State University
Email: Kerem.Demirtas@asu.edu

Pitu Mirchandani
School of Computing, Informatics & Decision Systems Engineering, Arizona State University
Email: pitu@asu.edu

Xuesong Zhou
School of Sustainable Engineering and the Built Environment, Arizona State University
Email: xzhou74@asu.edu

Extended Abstract

In this study, we are interested in dynamic calibration of car following parameters in order to explore both inter-driver and intra-driver heterogeneity. Specifically, we offer an augmented state space system for a lower order linear spacing car following model, and implement a modified Kalman filter algorithm in order to track the leader-follower pairs and simultaneously predict and estimate the parameters related with the behavior of the following drivers.

Microscopic traffic models consider the interaction of individual vehicles with nearby vehicles as well as the transportation infrastructure in a very detailed resolution to calculate the next position of the vehicles under consideration for a short term horizon (typically around 0.1 to 1 second) that gets updated as the simulation runs. Car Following (CF) models that describe the longitudinal motion of a vehicle under the laws of motions of physics along with a car following behavioral aspect and Lane Changing (LC) models that probabilistically define the lateral movements of a vehicle as instantaneous decision models are the two types of models applied in the context of microscopic traffic simulation.

Newell [New02] proposes a simplified car following theory where, in free flow, the vehicles travel at their desired speeds. However, in congestion, the follower adjusts his speed to the speed he observes from his leader according to a spacing of his preference at that speed, which results in a copy of the leader's trajectory, but shifted by a translation in time and space given by the following equation.

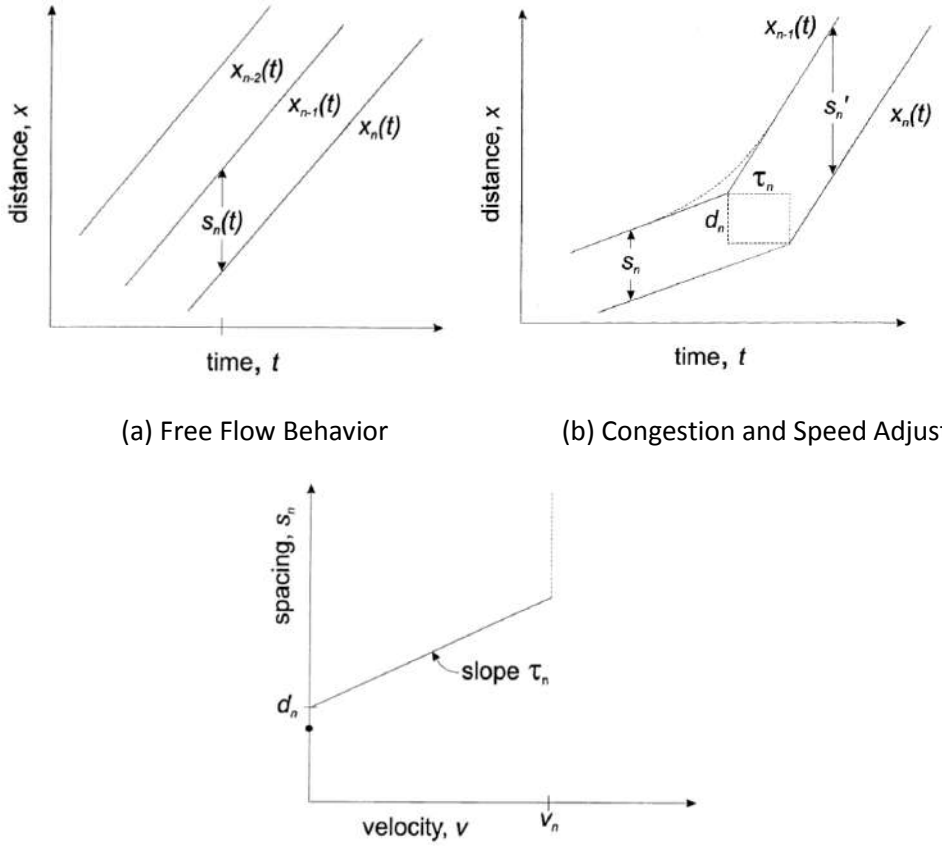
$$x_n(t+\tau_n) = x_{n-1}(t) - d_n \quad (1)$$

In equation (1), *x* denotes the position of the leading and the following vehicles, subscripted by *n*-1 and *n*, respectively at time points given inside the parenthesis. The parameters τ_n and d_n describe the simplified car following behavior of the follower *n*, and it is further assumed that τ_n and d_n come from a bivariate normal distribution which are also independent from the other vehicles [New02].

Equation (1) is also equivalent to the linear spacing model given in equation (2) below

$$s_n(t+\tau_n) = d_n + v_{n-1}(t)\tau_n \quad (2)$$

where s_n represents the spacing the follower *n* chooses to have at time $t+\tau_n$ with respect to the speed $v_{n-1}(t)$ of its leader observed at time t, by also adopting the exact same speed at $t+\tau_n$, so that $v_n(t+\tau_n) = v_{n-1}(t)$. The illustration of Newell's car following model can be found in Figure 1.



(a) Free Flow Behavior (b) Congestion and Speed Adjustment

(c) Linear Speed-Spacing Relation

FIGURE 1 Newell's Linear Spacing Car Following Model, adapted from [New02]

An advantage of this model is that the disturbances that are caused by driver inhomogeneity propagate according to rules that govern the macroscopic traffic flow, which represent the aggregate behavior of the vehicles on the road as an average rather than getting amplified [ACL04].

Daganzo [Dag06] show the equivalence between the cellular automata and kinematic wave models that also explains the bridge between the macroscopic models and Newell's simplified car following model [New02] under the assumption of a triangular fundamental diagram. Ahn et al. [ACL04] validate Newell's car following model by analyzing field data of vehicles discharging from long queues and verify that the parameter pairs are independent between drivers, and thus justify that variation was due to driver inhomogeneity. Ossen et al. [OHG06] studies the behavior of following vehicles under several car following models by a simulation approach and conclude that inter-driver behavior differences have to be considered instead of purely focusing on a single parameter setting that gives the best fit between the trajectories of the leader and the follower. More recently, Ossen and Hoogendoorn [OH08] study the effects of measurement noise on microscopic car following model parameter calibration by introducing several types of noise to synthetic data, and discover that the optimal calibration models under certainty assumption considerably fail due to the not so robust tendency of the objective function to small changes that result from noisy data, and therefore, conclude that some kind of smoothing or filtering is required in such situations. Kesting and Treiber [KT08], develop a genetic algorithm that integrates multiple car following

models in a nonlinear optimization framework in order to minimize the deviations of the estimated follower dynamics from those of the measurements by trying to reproduce the trajectory of the following vehicle. They also apply the parameters that are calibrated to a specific following vehicle to another one, and reveal that intra-driver heterogeneity is as important as inter-driver heterogeneity in terms of calibration errors. Duret et al. [DCB09] show that variability between different vehicles can be explained by estimating the parameters that induce a linear speed-spacing relation for the congested region of the fundamental diagram. Using NGSIM data, they find the best values of the parameters for Newell's car following model so that the reconstructed trajectories are as close as possible to the original ones. Chiabaut et al. [CLB10] propose an efficient method to estimate the parameters of Newell's simplified car following model by using the relative standard deviation measure rather than a classical criterion such as the mean square error as an objective to minimize the variations in measurements along the sampled trajectories. Taylor et al. [Tay15] develop a dynamic time warping (DTW) algorithm in order to estimate time dependent parameter values for Newell's car following model. Through numerical experiments on NGSIM data, they are able to show that intra-driver inhomogeneity, which can be described as a change of single driver's behavior in time during a trip, is also an issue in addition to inter-driver inhomogeneity.

Since the purpose of the study is to dynamically calibrate τ_n and d_n, introducing them into the state definition gives the augmented state $\mathbf{x} = [x_{n-1}, x_n, \tau_n, d_n]^T$, where the bold **x** represents vector notation. Further assuming that the transition of the new states τ_n and d_n towards the next time step will be an identity transform corrupted with Gaussian noise only, we can rearrange the whole state transition, which results in the matrix form given in the below equation.

$$\begin{bmatrix} x_{n-1}(t+\Delta t) \\ x_n(t+\Delta t) \\ \tau_n(t+\Delta t) \\ d_n(t+\Delta t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & v_{n-1}(t) & 0 \\ 0 & 1 & v_n(t) & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_{n-1}(t) \\ x_n(t) \\ \tau_n(t) \\ d_n(t) \end{bmatrix} + \boldsymbol{\psi}(t) \tag{5}$$

Given that both the leader and follower positions (x_{n-1}, x_n) are available through somewhat noisy GPS measurements, and the spacing (s_n) between them is provided through certain sophisticated technologies in the following vehicle, we can write the measurement equation in matrix form as below.

$$\begin{bmatrix} x_{n-1}(t+\Delta t) \\ x_n(t+\Delta t) \\ s_n(t+\Delta t) \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & v_{n-1}(t) & 1 \end{bmatrix} \begin{bmatrix} x_{n-1}(t+\Delta t) \\ x_n(t+\Delta t) \\ \tau_n(t+\Delta t) \\ d_n(t+\Delta t) \end{bmatrix} + \boldsymbol{\xi}(t) \tag{6}$$

Note that $\boldsymbol{\psi}(t)$, and $\boldsymbol{\xi}(t)$ represent zero mean state and measurement noises with covariance matrices **Q** and **R**, respectively.

Typically, in order for both the state and the measurement equations to hold for the underlying linear spacing model, the continuous increment Δt should be equal to $\tau_n(t)$. However, discretization is required for implementation purposes, so one should find the discrete time step value through rounding $\tau_n(t)$ to the nearest integer multiple of 0.1 second since the discretized data is available in 0.1 second resolution. An illustration of the Kalman filter algorithm with discrete timestamps indexed by *j* is given in Figure 2.

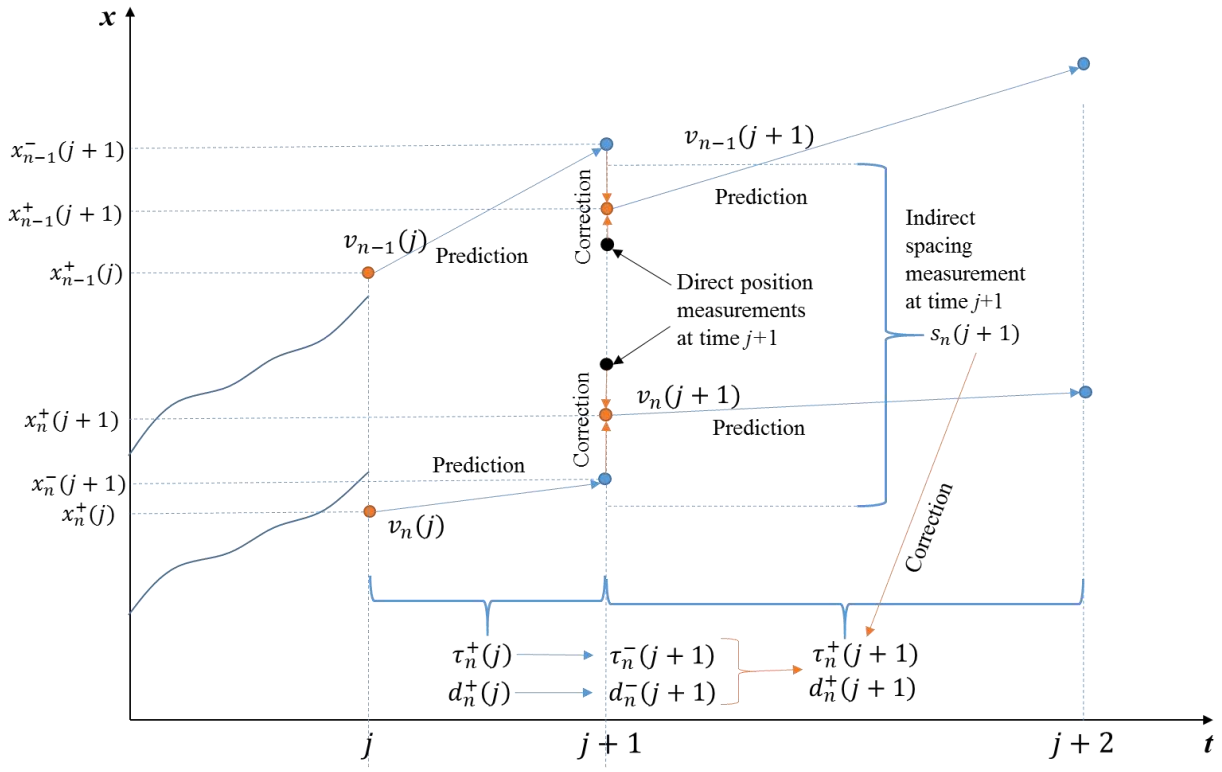


FIGURE 2 Illustration of an iteration of the Kalman filter algorithm

The proposed approach is tested on the trajectories from the Next Generation Simulation (NGSIM) project data and shows satisfactory results. This study also points that intra-driver inhomogeneity is an inevitable phenomenon that should not be overlooked in short term traffic state estimation and prediction. A computationally efficient and structurally appropriate algorithm such as a Kalman filter can also improve real-time estimation of macroscopic traffic features like density, flow and speed through just-in-time measures of backward wave speed as an outcome of this study. As a future study, application of this framework on a platoon level is under development, which is intended to give more aggregate information about the overall traffic state.

References

[ACL04] Ahn, Soyoung, Michael J. Cassidy, and Jorge Laval. "Verification of a simplified car-following theory." *Transportation Research Part B: Methodological* 38.5 (2004): 431-440.

[BKW05] Brockfeld, Elmar, Reinhart Kühne, and Peter Wagner. "Calibration and validation of microscopic models of traffic flow." *Transportation Research Record: Journal of the Transportation Research Board* 1934 (2005): 179-187.

[CBL09] Chiabaut, Nicolas, Christine Buisson, and Ludovic Leclercq. "Fundamental diagram estimation through passing rate measurements in congestion."*Intelligent Transportation Systems, IEEE Transactions on* 10.2 (2009): 355-359.

[CLB10] Chiabaut, Nicolas, Ludovic Leclercq, and Christine Buisson. "From heterogeneous drivers to macroscopic patterns in congestion."*Transportation Research Part B: Methodological* 44.2 (2010): 299-308.

[Dag06] Daganzo, Carlos F. "In traffic flow, cellular automata= kinematic waves."*Transportation Research Part B: Methodological* 40.5 (2006): 396-403.

[DCB09] Duret, Aurélien, Christine Buisson, and Nicolas Chiabaut. "Estimating individual speed-spacing relationship and assessing ability of Newell's car-following model to reproduce trajectories." *Transportation Research Record: Journal of the Transportation Research Board* 2088 (2008): 188-197.

[Kal60] Kalman, Rudolph Emil. "A new approach to linear filtering and prediction problems." *Journal of basic Engineering* 82.1 (1960): 35-45.

[KT08] Kesting, Arne, and Martin Treiber. "Calibrating car-following models by using trajectory data: Methodological study." *Transportation Research Record: Journal of the Transportation Research Board* 2088 (2008): 148-156.

[New02] Newell, Gordon Frank. "A simplified car-following theory: a lower order model." *Transportation Research Part B: Methodological* 36.3 (2002): 195-205.

[OH08] Ossen, Saskia, and Serge Hoogendoorn. "Validity of trajectory-based calibration approach of car-following models in presence of measurement errors." *Transportation Research Record: Journal of the Transportation Research Board* 2088 (2008): 117-125.

[OHG06] Ossen, Saskia, Serge Hoogendoorn, and Ben Gorte. "Interdriver differences in car-following: a vehicle trajectory-based study." *Transportation Research Record: Journal of the Transportation Research Board* 1965 (2006): 121-129.

[Tay15+] Taylor, Jeffrey, et al. "Method for investigating intradriver heterogeneity using vehicle trajectory data: a dynamic time warping approach." *Transportation Research Part B: Methodological* 73 (2015): 59-80.

Representation requirements for perfect first-in-first-out verification in continuous flow dynamic models

Hillel Bar-Gera[1] and Malachy Carey[2]

1. Introduction

Continuous flow dynamic network loading (DNL) models have been studied extensively in recent decades, and proved useful in many practical applications. Popular examples are the cell transmission model – CTM (Daganzo 1994) and the link transmission model – LTM (Yperman et al. 2006). One of the challenging issues in these models is the way priorities of different traffic streams (commodities) are managed. In many previous studies, as early as Carey (1986) and Smith (1993), flow propagation priority rules aimed to satisfy the first-in-first-out (FIFO) condition; but most practical continuous flow dynamic models do not satisfy FIFO perfectly. Many of these models are very useful, despite their deviation from FIFO. Following previous publications (Blumberg and Bar Gera, 2009; Carey *et al*., 2014), the purpose of this study is to explore the efforts that might be needed if full adherence to network-wide FIFO is desired, as well as whether it should be pursued and for what purposes.

It is well known that in real traffic adherence to FIFO is not perfect. Some may argue that if our models deviate from FIFO and reality deviates from FIFO then there is no problem. However, there is no a-priori reason to assume that the deviations from FIFO in models (stemming from computational reasons) should bear resemblance to the deviations from FIFO in real traffic (stemming from driver behavior). In addition, there is no reason to assume that alternative priority rules will be easier than FIFO to implement computationally, especially if such rules will be derived from quantitative empirical evidence to reflect actual traffic behavior.

The full paper (Bar-Gera and Carey, 2017) considers six different ways to represent dynamic flow solutions over a network. For each representation one may ask whether it is sufficient for verifying FIFO, whether the verification process is finite, and whether proving FIFO can be directly implied. The evaluation involves eight alternative definitions of FIFO, seven of them are shown to be essentially equivalent, while the last definition is not, and may therefore be considered as "weak" FIFO. In this extended abstract a few key findings are highlighted, primarily using illustrative examples. Section 2 describes the limitations of pre-determined time steps. An example for the limitations of history lists is provided in Section 3. Section 4 shows why certain straight forward approaches for FIFO verification may lead to infinite processes. Conclusions and future potential research directions are presented in section 5.

[1] Ben-Gurion University of the Negev, Department of Industrial Engineering and Management, P.O.Box 653 Beer-Sheva, 84105 Israel. bargera@bgu.ac.il

[2] Ulster University Business School, Ulster University, Northern Ireland BT37 0QB, UK. m.carey@ulster.ac.uk

2. The limitations of predetermined time steps

The original CTM and LTM models (Daganzo 1994; Yperman et al. 2006) assume a predetermined grid, in which time is divided into regular "steps" of fixed duration. The following example illustrates why perfect FIFO may require consideration of additional points in time. Consider two routes, r_1 and r_2, sharing one cell (or link) a with entrance capacity of 6 vehicles per time step and exit capacity of 4 vehicles per time step. In the first time step 6 vehicles from route r_1 enter the cell, and in the following time step 2 vehicles from route r_2 enter the cell. The cumulative volumes at the cell entrance and exit by route and time are given in table 1, using the regular pre-determined time step resolution.

Label	Location	t=0	t=1	t=2	t=3
Route r_1	Entrance	0	6	6	6
Route r_1	Exit	0	0	4	6
Route r_2	Entrance	0	0	2	2
Route r_2	Exit	0	0	0	2
Total (a)	Entrance	0	6	8	8
Total (a)	Exit	0	0	4	8

Comment: the value in each cell represent the cumulative volume by route, location and time.

Table 1: Illustration of the limitations of pre-determined time steps for FIFO satisfaction

The most straightforward interpretation of this solution is obtained by linear interpolation, suggesting that the last two vehicles on route r_1 exit in parallel with the two vehicles on route r_2. This is a FIFO violation, since all vehicles on route r_1 entered the cell before all vehicles on route r_2. The problem can be easily fixed by adding a breakpoint at $t = 2.5$, yet such breakpoint is not part of the original grid. The general question is how many breakpoints are needed, and how can these be determined.

It is worth noting that solutions with predetermined time-steps may also cause inaccuracies with respect to other DNL conditions (e.g. spillback). Several studies used grid-free methods to compute exact solutions in terms of other DNL conditions (e.g. Mazaré et al., 2011; Raadsen et al., 2015); but these studies did not address the issue of perfect network-wide FIFO in a multi-commodity context.

3. Time step histories

In CTM the problem presented in table 1 is typically solved by consideration of entrance time within the current cell. But if the "entrance" and "exit" considered above relate to a longer segment shared by the two routes, possibly containing several cells or links, then tracking entrance time within one cell (or one link) will not be sufficient. Smith (1993) proposed to track the entire history of entrance

times throughout the route. This is probably one of the most elaborate schemes for handling network-wide FIFO in finite-dimensional solutions, and may seem as far as one would wish to go in an effort to satisfy FIFO. However, as we shall see, history lists are not sufficient for perfect FIFO verification.

Consider the network in figure 1 and the time-step-histories listed in table 2. Four units of traffic travel along two routes. In both cases (4A and 4B) there are no direct violations of FIFO, i.e. in both solutions the exit order at node 4 is consistent with the time-step-histories. Yet an attempt to elaborate the interpretation by adding breakpoints of time, leads to a conclusion that the solution in 4A does not satisfy FIFO, while the solution in 4B does satisfy FIFO. This example shows why time-step-histories are not sufficient for perfect network-wide FIFO verification.

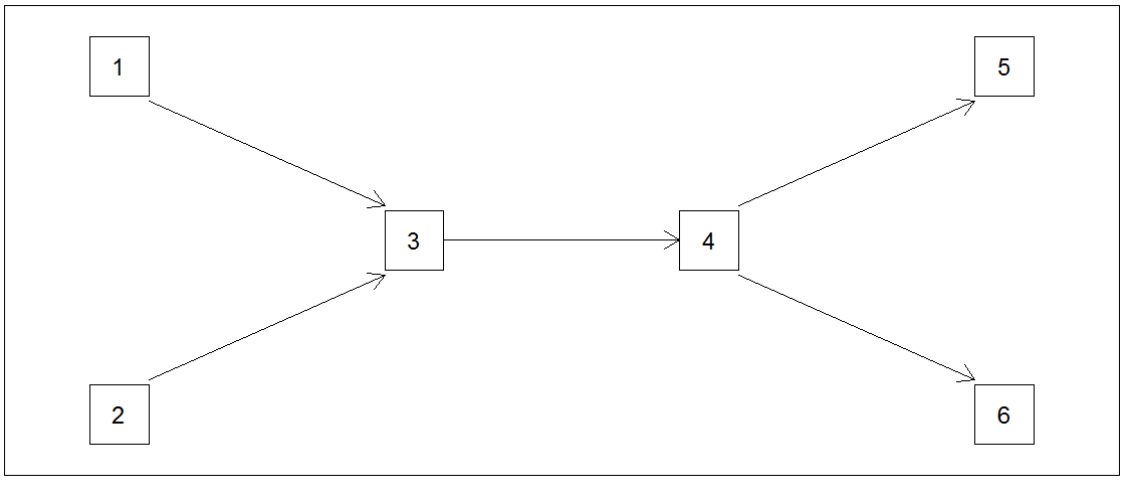


Figure 1: an example of a network with one merge and one diverge

Volume	Route	$T_s(1)$	$T_s(2)$	$T_s(3)$	$T_s(4)$	$T_s(5)$	$T_s(6)$
1	[1,3,4,5]	1		3	4	6	
1	[1,3,4,6]	2		3	4		5
1	[2,3,4,5]		2	3	4	5	
1	[2,3,4,6]		1	3	4		6

A. Time-step-histories that do not contain direct FIFO violation, but **do not** satisfy perfect FIFO.

Volume	Route $r(c)$	$T_s(c,1)$	$T_s(c,2)$	$T_s(c,3)$	$T_s(c,4)$	$T_s(c,5)$	$T_s(c,6)$
1	[1,3,4,5]	1		3	4	5	
1	[1,3,4,6]	2		3	4		6
1	[2,3,4,5]		2	3	4	6	
1	[2,3,4,6]		1	3	4		5

B. Time-step-histories that do not contain direct FIFO violation, and **do** satisfying perfect FIFO.

Table 2: An illustration of the difference between time-step-histories FIFO and perfect FIFO

4. Verifying FIFO in a network with cycles

The problems shown in the previous sections could be solved by adding breakpoints in time. One straightforward approach to FIFO verification (and satisfaction) is to add any breakpoint of time that could be needed, traced forward as well as backwards along routes, so that the same set of breakpoints at a node applies to all routes, and the implied breakdown of traffic into components (or "cohorts") remains the same throughout each route. This section illustrates why this could lead to an infinite (or extremely large) number of breakpoints. For the sake of conciseness, the exposition here is less formal mathematically; further details can be found in (Bar-Gera and Carey, 2017).

Consider the 6-nodes 8-links network in figure 1. Suppose that the length of each outer link is 20 m, the length of links [3,5] and [6,2] is 40 m, the free flow speed is 20 m/s, and the capacity is 10 v/s. In addition to the overall link capacities, there are bottlenecks at the exits from links [3,5] and [6,2] with a capacity limitation of 1 v/s. Four routes are used: $r_1 = [1,2,3,4,5,6]$; $r_2 = [1,2,3,5,6]$; $r_3 = [4,5,6,1,2,3]$; $r_4 = [4,5,6,2,3]$. The initial flow on all routes is 1 v/s, so there is no congestion and all traffic moves at free flow speed. At a certain time, t_0, the flow on route r_2 increases to $1 + \delta_1$ for one second, and the flow on route r_4 increases to $1 + \delta_2$ for one second. As a result queues of δ_1 and δ_2 vehicles are created on links [3,5] and [6,2] respectively, and the travel times on these links increase from 2 to $2 + \delta_1$ and $2+\delta_2$ respectively. Subsequently, flows return to their original values, so the new travel times remain unchanged for the remainder of the simulated period. The solution in this case could be described with a very small number of breakpoints, but if these breakpoints are traced as described above, the resulting set of breakpoints contains any sum of integer products of δ_1 and δ_2. Depending on the specific values of δ_1 and $\delta_2$this set could be very large, or even infinite.

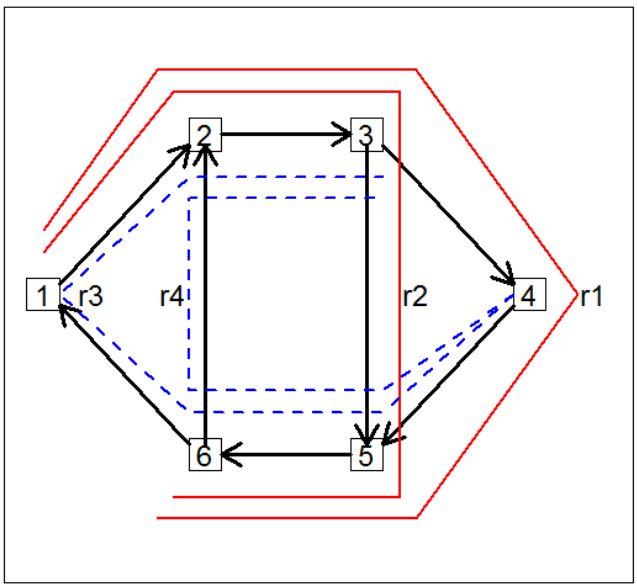


Figure 1: an example of a network with cycles illustrating FIFO challenges

5. Conclusions and future research

In a nutshell, the key conclusion of this analysis is that in order to satisfy perfect network-wide FIFO the number of discretized elements of flow should probably be allowed to grow quickly and unboundedly with model duration, and it cannot be determined a-priori. The implication is that practical usage of perfect FIFO solutions is probably impossible. If breakpoints of time are added more cautiously (see details in Bar-Gera and Carey, 2017), methods for computing perfect FIFO solutions could perhaps be developed in future research, at least for specific limited problems. Such solutions can then be used for the evaluation of FIFO violations in other, more practical methods, and thus contribute to the understanding of the properties of such methods.

References

1. Bar-Gera, H., Carey, M., 2017. Representation requirements for perfect first-in-first-out verification in continuous flow dynamic models. Accepted for publication in *Transportation Research Part B: Methodological.*
2. Blumberg, M., Bar-Gera, H., 2009. Consistent node arrival order in dynamic network loading models. *Transportation Research Part B: Methodological* 43(3), 285–300.
3. Carey, M., Bar-Gera, H., Watling, D., Balijepalli, C., 2014. Implementing first-in-first-out in the cell transmission model. *Transportation Research Part B: Methodological* 65, 105-118.
4. Carey, M., 1986. A constraint qualification for a dynamic traffic assignment model. *Transportation Science*, 20(l), 55-58.
5. Daganzo, C.F., 1994. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological* 28(4), 269–287.
6. Mazaré, P.E., Dehwah, A.H., Claudel, C.G., Bayen, A.M., 2011. Analytical and grid-free solutions to the Lighthill–Whitham–Richards traffic flow model. *Transportation Research Part B: Methodological* 45 (10), 1727–1748.
7. Raadsen, M.P.H., Bliemer, M.C.J., Bell, M.G.H. 2015. An efficient and exact event-based algorithm for solving simplified first order dynamic network loading problems in continuous time. *Transportation Research Part B: Methodological*, http://dx.doi.org/10.1016/j.trb.2015.08.004.
8. Smith, M.J., 1993. A new dynamic traffic model and the existence and calculation of dynamic user equilibria on congested capacity-constrained road networks. *Transportation Research Part B: Methodological* 27(1), 49-63.
9. Yperman, I., Logghe, S., Tampere, C.M.J., Immers, L.H., 2006. The multi-commodity link transmission model for dynamic network loading. In: *Proceedings of the 85th Annual Meeting of the TRB, Washington, DC*.

Traffic & Mobility

FD2: Ride Sharing and Parking

Friday 4:30 – 6:00 PM

Session Chair: Yafeng Yin

4:30 **Optimal Two-Sided Pricing Strategies of Shared E-Parking Platform with Elastic Demand**
Chaoyi Shao, Hai Yang, Fangni Zhang*
Hong Kong University of Science and Technology

5:00 **Integration of an Aggregated Dynamic Traffic Model with Advanced Optimization Techniques for Strategic Transit-Parking Planning**
Joana Cavadas, António Pais Antunes, Nikolas Geroliminis*
University of Coimbra, École Polytechnique Fédérale de Lausanne

5:30 **Optimal Parking Provision for Ride-Sourcing Services**
Zhengtian Xu, Yafeng Yin, Liteng Zha*
University of Michigan, University of Florida

Optimal two-sided pricing strategies of shared E-parking platform with elastic demand

Chaoyi Shao[a], Hai Yang[a] and Fangni Zhang[a,b]

[a]*Department of Civil and Environmental Engineering, The Hong Kong University of Science and Technology, Hong Kong, P.R. China*

[b]*Centre for Transport Studies, Department of Civil and Environmental Engineering, Imperial College London, London SW7 2AZ, United Kingdom*

Introduction

Parking in downtown areas has usually been a headache for both commuters and traffic managers in many metropolises because of the limitation of parking slots in center areas. Cruising for a parking spot and the walking time to the working place will also be an emphasis part of the commuter's decision whether he/she travel by auto or transit. Levying parking fee is not a prospective way to decrease the parking demand since in some cases they may exceed the total vehicle operating cost. Due to the lack of land in CBD areas, we cannot simply construct new parking facilities to satisfy the commuters. Especially in cities like Hong Kong, the high density of traffic and limited road capacity will be the restrictions to its development. Shoup (2006) has found that 30% of traffic congestion in road networks is caused when people are circulating around to find a parking spot, and about 8.1 min is spent in finding a parking spot. Another research (Ayala et al., 2011) also introduces in Chicago, in order to find a parking slot, people will suffer extra 63 million miles driving each year, which generates 48,000 tons of carbon dioxide to the environment.

Parking management is normally considered as an integrated part of travel demand management. From this perspective, levying some road toll on the vehicle (Verhoef et al., 1995; Arnott and Inci, 2006; Zhang et al., 2008 and Qian et al., 2011 and 2012) is considered to be effective for mitigating traffic congestion and regulating parking demand. Zhang et al. (2011), Yang et al. (2013) and Liu et al. (2014a, b) found that parking reservation through parking permits distribution and trading is efficient in traffic management. Particularly, it was found that an appropriate combination of reserved and unreserved parking spots can temporally relieve traffic congestion at the bottleneck and hence reduce the total system cost, because commuters without a reserved parking spot are compelled to leave home earlier in order to secure a public parking spot. A latest review of the economic analysis and modeling of

parking was given by Inci (2015). It is also noteworthy that there has been a number of recent studies that examined parking slot allocation and pricing using game theory or incorporated parking searching process as part of the traffic network equilibrium problem in a network context (Boyles et al., 2015; Liu and Geroliminis, 2016; Zou et al., 2015; He et al., 2015; Cao and Menendez, 2015; Du and Gong, 2016; Zheng and Geroliminis, 2016).

To reduce the wastage of cruising for parking, some smart parking systems have been set up to make guidance for people to park their vehicles. In San Francisco, SFPark (http://sfpark.org/) puts all the information online for people to park and also uses demand-dependent parking fees to coordinate the traffic congestion in different areas. Geng and Cassandras (2013) propose a novel 'smart parking' system which can be downloaded as a smart phone app to help individuals reserve or find both on street and off street parking slots and the result shows quite positive, when people are informed the details of parking slot, they can save a slot of time and the parking slots will be further utilized as well.

Based on improving the utilization of existing parking slots, shared parking will be a new notion of making more efficient use of parking facilities. It uses the private parking slots when the owners are not using. This can be feasible for most of the private parking slots are not used during the working hours since their owners go to work in other locations, and also their pattern of availability remains the same in a long term. Moreover, shared parking can be a win-win strategy since on the one side the rigid parking demand can be satisfied, on the other side the parking slot owners can make additional income by suffering little inconvenience.

Considering the above new business opportunity and model particularly in an area with mixed commercial and residential land use developments, parking slot management companies intend to temporarily repurchase some private parking slots and sell them to public users during certain time of a day. Guo et al. (2016) was the first one to develop a simulation-optimization based decision method to determine the repurchase strategy. A Gaussian mixture model was proposed to describe the time-varying arriving/departing behavior of drivers, and the expected optimal repurchase amounts and parking time were estimated via simulation optimization. Shao et al. (2016) considered advanced booking and allocation of shared parking slots and proposed a simple binary integer linear programming model to allocate the requests to specific parking slots so as to maximize the parking slot utilization or accommodate as many requests as possible under parking space and time constraints. Xu et al. (2016) addressed the private parking slot sharing problem during regular working hours in a big city by using market design theory. Their experimental results showed that the proposed mechanisms would result in remarkable social welfare gain and budget balance for the e-parking platform in cities with a large population.

A two-sided market is a meeting place for two groups of agents who interact through an intermediary or platform and provide each other with network benefits. Particular two-sided markets, like credit card services that connect cardholders and affiliated merchants and operating systems that connect application developers and clients, have emerged over a long period, but the theory of two-sided markets was developed after the year 2000. Rochet and Tirole (2004) analyze the price allocation in both side of the market under several different situations and also compare the optimal pricing structures between private and social cases. Caillaud and Jullien (2003) determine the equilibrium of two-sided market structures which emerge and characterize the efficiency properties. Armstrong (2006) gives three models to illustrate the competition in two-sided market which are monopoly platform, two-sided single-homing and competitive bottleneck. Economides and Tåg (2012) discuss the network neutrality regulation in two-sided market model and focus on the cross-group externalities heterogeneity among consumers and content providers. Diverse issues related to the two-sided markets were reviewed and addressed by Roson (2005) under various assumptions about timing, price instruments and externalities.

In this paper, we consider the optimal price strategy of shared e-parking platform with elastic demand and supply. Combined with the allocation function which is similar to the matching function given by Yang and Yang (2011) in taxi market in aggregate analysis, our model is based on the model of Armstrong (2006) and Economides and Tåg (2012) to analyze the price strategies of both sides under three different scenarios. We find out that when the platform serves as a reseller, the profit of monopoly platform can be maximized when the ratio of prices is less than a specific value and the platform is in deficit all the time when system optimal is achieved. While when the platform serves as an agent, the contract curve is a linear segment with a constant slope. By comparing with the two roles the platform serves, we find that when that the platform serves as an agent is profitable, the efficiency of the platform will always be greater than the reseller case. Moreover, both selling and buying prices are also greater than reseller case.

References

Armstrong, M., 2006, Competition in two-sided markets. The RAND Journal of Economics 37, 668-691.

Arnott, R., Inci, E., 2006. An integrated model of downtown parking and traffic congestion. Journal of Urban Economics 60 (3), 418-442.

Ayala, D., Wolfson, O., Xu, B., DasGupta, B., Lin, J., 2011. Parking slot assignment games. In: Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. http://dx.doi.org/10.1145/2093973.2094014.

Boyles, S.D., Tang, S., Unnikrishnan, A., 2015. Parking search equilibrium on a network. Transportation Research Part B: Methodological 81, 390-409.

Caillaud, B., Jullien, B., 2003. Chicken & egg: Competition among intermediation service providers. The RAND Journal of Economics 34, 309-328.

Cao, J., Menendez, M., 2015. System dynamics of urban traffic based on its parking-related-states. Transportation Research Part B: Methodological 81, 718-736.

Du, L., Gong, S., 2016. Stochastic Poisson game for an online decentralized and coordinated parking mechanism. Transportation Research Part B: Methodological 87, 44-63.

Economides, N., Tåg, J., 2012. Network neutrality on the Internet: A two-sided market analysis. Information Economics and Policy 24, 91-104

Geng, Y., Cassandras, C., 2013. New "smart parking" system based on resource allocation and reservations. IEEE transactions on Intelligent Transportation Systems 14 (3), 1129-1139.

Guo, W., Zhang, Y., Xu, M., Zhang, Z., Li, L., 2016. Parking spaces repurchase strategy design via simulation optimization. Journal of Intelligent Transportation Systems: Technology, Planning, and Operations 20 (3), 255-269.

He, F., Yin, Y., Chen, Z., Zhou, J., 2015. Pricing of parking games with atomic players. Transportation Research Part B: Methodological 73, 1-12.

Inci, E., 2015. A review of the economics of parking. Economics of Transportation 4 (1–2), 50-63.

Liu, W., Geroliminis, N., 2016. Modeling the morning commute for urban networks with cruising-for-parking: An MFD approach. Transportation Research Part B: Methodological 93, 470-494.

Liu, W., Yang, H., Yin, Y., 2014a. Expiable parking reservations for managing morning commute with parking space constraints. Transportation Research Part C: Emerging Technologies 44, 185-201.

Liu, W., Yang, H., Yin, Y., Zhang, F., 2014b. A novel permit scheme for managing parking competition and bottleneck congestion. Transportation Research Part C: Emerging Technologies 44, 265-281.

Qian, Z., Xiao, F., Zhang H.M., 2011. The economics of parking provision for the morning commute. Transportation Research Part A: Policy and Practice 45 (9), 861-879.

Qian, Z., Xiao, F., Zhang H.M., 2012. Managing morning commute traffic with parking. Transportation Research Part B: Methodological 46 (7), 894-916.

Rochet, J., Tirole, J., 2004. Platform competition in two-sided markets. Journal of the European Economic Association 1. 990-1029.

Roson, R., 2005. Two-sided markets: A tentative survey. Review of Network Economics 4, 142-160.

Shao, C., Yang, H., Zhang, Y., Ke, J., 2016. A simple reservation and allocation model of shared parking slots. Transportation Research Part C: Emerging Technologies 71, 303-312.

Shoup, D.C., 2006. Cruising for parking. Transport Policy 13, 479-486.

Verhoef, E., Nijkamp, P., Rietveld, P., 1995. The economics of regulatory parking policies: The (im)possibilities of parking policies in traffic regulation. Transportation Research Part A: Policy and Practice 29 (2), 141-156.

Xu, S.X., Cheng, M., Kong, X., Yang, H., Huang, G.Q., 2016. Private parking slot sharing. Transportation Research Part B: Methodological 93, 596-617.

Yang, H., Liu, W., Wang, X., Zhang, X., 2013. On the morning commute problem with bottleneck congestion and parking space constraints. Transportation Research Part B: Methodological 58, 106-118.

Yang, H., Yang, T., 2011. Equilibrium properties of taxi markets with search frictions. Transportation Research Part B: Methodological 45, 696-713.

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. Transportation Research Part B: Methodological 42 (1), 38-56.

Zhang, X.N., Yang, H., Huang, H.J., 2011. Improving travel efficiency by parking permits distribution and trading. Transportation Research Part B: Methodological 45 (7), 1018-1034.

Zheng, N., Geroliminis, N., 2016. Modeling and optimization of multimodal urban networks with limited parking and dynamic pricing. Transportation Research Part B: Methodological 83, 36-58.

Zou, B., Kafle, N., Wolfson, O., Lin, J., 2015. A mechanism design based approach to solving parking slot assignment in the information era. Transportation Research Part B: Methodological 81, 631-653.

Integration of an aggregated dynamic traffic model with advanced optimization techniques for strategic transit-parking planning

Joana Cavadas [a,*], António P. Antunes [a], Nikolas Geroliminis [b]

[a] *CITTA, Department of Civil Engineering, University of Coimbra, Portugal*
[b] *Schoool of Architecture, Civil and Environmental Engineering, Laboratory of Urban Transport Systems, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland.*

Public transit systems are essential for the mobility in urban areas. These systems play a crucial role not only in providing those who cannot own or drive a car with an acceptable level of mobility (Preston and Rajé, 2007), but also with respect to the pursuit of all the dimensions of sustainable development. As a matter of fact, by helping to reduce fuel consumption, pollutant emissions and traffic congestion in comparison to the automobile, urban transit is clearly advantageous from the environmental and economic standpoints (Schiller et al., 2010). On the other hand, at least in Europe and North America, transit systems are a source of financial problems for local governments, as their revenues are rarely enough to cover their expenses (APTA, 2010).

A possible approach to circumvent transit financial problems is to manage transit systems and parking systems in an integrated manner, using parking fees to fund transit deficits. For this purpose, we developed an optimization-simulation modeling approach for assisting local governments in the establishment of pricing policies for the transit networks and parking areas under their control, directly or indirectly (through concessions). This approach has two components: a mixed-integer nonlinear optimization model and a network level aggregate traffic model based on the macroscopic fundamental diagram (MFD). The latter model provides a relationship between network space-mean vehicle density and flow in urban areas with small spatial density heterogeneity taking into account the number of vehicles (accumulation) in a zone, the zone's average speed and congestion conditions (Geroliminis and Daganzo, 2008).

The context for the application of the proposed modeling approach is a city divided into zones, which are grouped into regions taking into account their travel patterns and traffic characteristics. Trips between each pair of zones can be made either by car (for people who can use this mode) or by bus, or not made at all if (generalized) travel costs are considered too high by the traveler. Modal choice in the city is described by two logit models. The first describes the behavior of people who cannot use a car, and distinguishes the transit and the non-travel modes. The second is a nested model that applies to people who can use a car. The first level of this model distinguishes between car, transit and non-travel choices, and the second level, which only concerns car users, relates to parking options. Specifically, it distinguishes between parking in the desired destination zone and in nearby zones, and between free and paid parking. For both logit models, choices depend on the travel costs of both modes. In the case of car, these costs consist of vehicle depreciation and maintenance costs, fuel costs, maintenance costs, walking costs, cruising for parking costs (if parking is free) and parking fees (if parking is paid). In the case of transit, fares, time costs, access costs, and discomfort costs are taken into account.

The objective function of the proposed optimization model aims at maximizing the net revenues made from transit and parking (R), being written as follows:

$$R = \sum_{o \in Z} \sum_{d \in Z} q_{od}^{A} \left(p_{d}^{A} - C_{d}^{A} \right) + \sum_{o \in Z} \sum_{d \in Z} q_{od}^{B} \left(p_{od}^{B} - C_{od}^{B} \right)$$

where $\boldsymbol{Z}$ is the number of city zones; q^A_{od} is the number of car trips between zones o and d (that need to pay for parking); p^A_d is the parking fee in zone d; C^A_d is the unit parking supply cost in zone d; q^B_{od} is the number of transit trips between zones o and d; p^B_{od} is the transit fare for trips between zones o and d; C^B_{od} is the unit transit supply cost between zones o and d.

The decision variables of the model are the number of car and transit trips (which are related to the supply levels for parking and transit), the parking fees, and the transit fares. With respect to parking, fees are applied in a zone if and only if the number of cars willing to park there exceeds a given proportion of the zone's total parking capacity. In the case of transit, the OD zones of the city are grouped into transit zones defined according to OD pairs; that is, the amounts paid by transit users depend on the origin and destination of their trips.

The MFD dynamics model is embedded in the optimization model in order to determine the system dynamics according to speeds, congestion and cruising for parking costs taking into account the demand for car trips. In this model, a unimodal, continuous and concave MFD function is considered to relate traffic flow, traffic density and speed for each region (set of zones with similar traffic characteristics). These functions are estimated for each region taking into account characteristics of the road network such as free flow speed, jam density and congested wave speed, and how this network is influenced by the transit network. Two different time frames are considered in this MFD model, one to update the modal shares according to their generalized costs and the zone-based routes followed by car drivers for each OD-pair, and the other to set the different speeds, congestion and cruising for parking effects based on the MFD functions. Three levels of heterogeneity are created in this model to accommodate route choice, mode choice, estimation of the average speed for each OD trip and parking choices. The first level takes into account the car route followed by each car user according to OD pairs and departure time periods, where the choice lies on the least cost route that travelers might followed. The accumulation and flows given by this level are then used to upgrade the dynamics of the network for the zone level. This second level is used to relate the flow with parking capacity, allowing to achieve the levels of cruising for parking according to the vacant free parking places and the demand assigned to them. In this level, aspects linked to the direction followed by users is also considered so that the characteristics of a two-way network are also included and dealt with in the third level of the model. This last level is the one where the MFD function is applied. Taking into account the flows and the cruising for parking in each zone, the connection between each zone and region is now considered to integrate the MFD function and determine the speed (which includes congestion aspects) for each zone allowing to estimate the expected generalized travel costs for each user.

The solutions to the optimization model (where the MFD traffic model is embedded) are calculated through hybrid simulated annealing and cross entropy SA-CE algorithm, as the model is too complex to be solved through mathematical programming methods.

We illustrate the results that can be obtained through the model with an application to a hypothetical two-region, 25-zone city inspired on midsize cities in Portugal (Figure 1). According to the best solution found through the SA-CE, the parking fee and the transit fare for this city should be 3€/hour and 1.75€/trip, respectively. As shown in Table 1, this would lead to a positive net revenue of about 35.52 10^3€ if parking and transit are managed together, and to modal shares for car and transit of 61% and 39%. The table also shows that, if the parking fee is increased, the number of trips made by car decreases and the average speed increases. The change of speed over the period under analysis obtained through the MFD model can be seen in Figure 2.

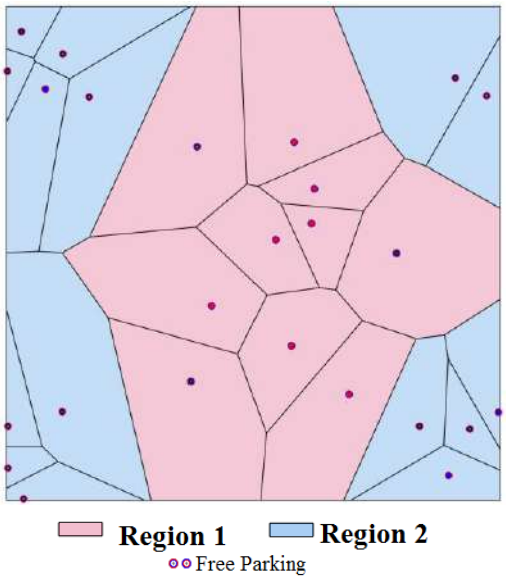


Figure 1 - Hypothetical two-region twenty-five-zone city

Table 1 – Key performance indicators for different parking fees and transit fares

Parking Fee (€)	Transit Fare (€)	Net Revenues (10^3 €)			Number of car trips (10^3)			Number of transit trips (10^3)	Average speed (km / h)
		Parking	Transit	Total	Free parking	Paid parking	Total		
3.25	1.25	17.03	-25.97	-8.94	147.94	18.57	166.51	125.45	26.05
3.50	1.25	15.85	-25.58	-9.73	149.35	16.66	166.01	125.77	26.05
3.00	1.25	17.91	-26.46	-8.55	146.53	20.69	167.22	125.06	26.07
3.25	1.50	18.32	-3.77	14.55	151.85	19.14	170.99	119.34	25.81
3.25	1.00	16.20	-51.11	-34.91	143.95	18.21	162.16	131.67	26.29
3.00	1.75	20.28	15.24	35.52	154.11	21.82	175.93	113.16	25.57

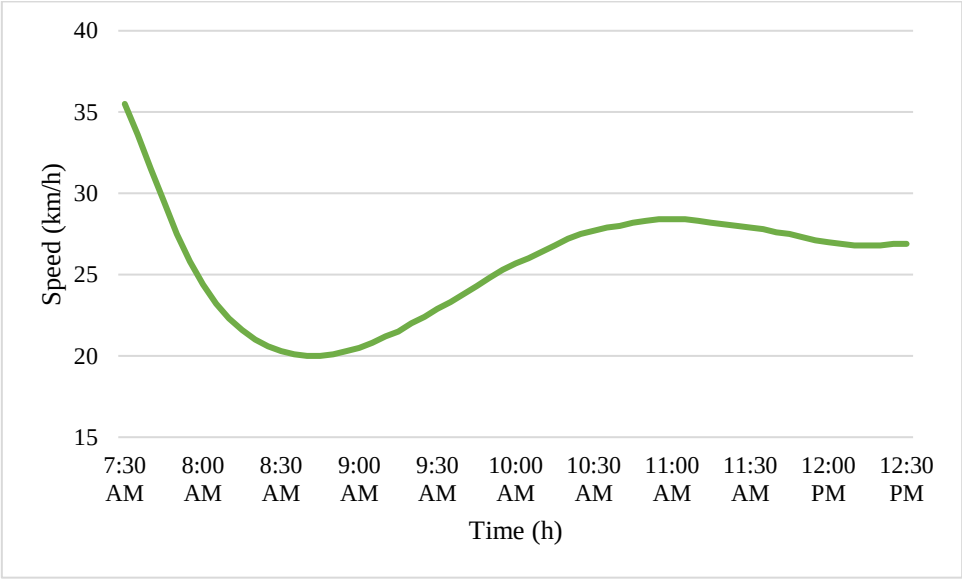


Figure 2 - Car speed in Region 1 of the hypothetical city with a parking fee of 3€ and a transit fare of 1.75€.

The proposed integrated transit-parking planning model provides a good understanding of how pricing schemes affect the city mobility dynamics. It is also able to provide insights into how the

deficits of a transit system can be coped with by managing transit fares and parking fees while ensuring reasonable levels of service. Furthermore, aspects such as congestion, parking capacity or transit system supply can also be positively affected by adjusting transit fares and parking fees.

References

APTA, (2003). Public Transportation Fact Book, 53th Edition. FHWA, Highway Statistics.

Geroliminis, N., and Daganzo, C.F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B: Methodological 42, 759–-70.

Preston, J., and Raje, F. (2007). Accessibility, mobility and transport-related social exclusion. Journal of Transport Geography 15(3), 151-160.

Schiller, P.L., Bruun, E., and Kenworthy, J.R. (2010). An Introduction to Sustainable Transportation: Policy, Planning and Implementation. Earthscan Ltd., London, UK. 368 p.

Optimal Parking Provision for Ride-sourcing Services

Zhengtian Xu [a], Yafeng Yin [a,*], Liteng Zha [b]

[a]*Department of Civil and Enviromental Engineering, University of Michigan, 2350 Hayward Street, Ann Arbor, MI 48109-2125, United States*
[b]*Department of Civil and Coastal Engineering, University of Florida, 365 Weil Hall, Gainesville, FL 32611-6580, United States*

Abstract

Ride-sourcing services have become increasingly important in meeting travel needs in metropolitan areas. However, the cruising of vacant ride-sourcing vehicles generates additional traffic demand that may worsen traffic conditions. This paper investigates the allocation of a certain portion of road space to on-street parking for vacant ride-sourcing vehicles. A macroscopic conceptual framework is developed to capture the trade-off between capacity loss and the reduction of cruising. Considering a hypothetical matching mechanism adopted by the platform, we further materialize the framework and then apply it to study the interactions between the ride-sourcing system and parking provision under various market structures.

Keywords: Ride-sourcing services; Cruising; Parking; Matching; Parking provision

Extended Abstract

1. Introduction

The proliferation of smartphones in recent years has catalyzed rapid growth of ride-sourcing services (also known as transportation network companies or TNCs) such as Uber and Lyft (Rayle et al., 2016). By requesting rides via mobile applications, customers of these ride-sourcing services are matched efficiently to affiliated drivers who drive their own non-commercial vehicles to provide for-hire rides. Such on-demand e-hailing services significantly reduce the search frictions and bring together riders and drivers with very low transaction costs (Anderson, 2014; He and Shen, 2015; Wang et al., 2016). As an evidence for the rapid expansion of ride-sourcing services, the number of active drivers on Uber increased exponentially in the U.S. from almost zero in 2012 to over 160,000 in 2014 (Hall and Krueger, 2015), and globally passed over the one-millionth landmark in 2015 (Lazo, 2015). According to a recent travel survey (the San Francisco Municipal Transportation Agency, 2015), ride-sourcing services have been used by 23% of residents in the San Francisco Bay Area at least monthly. Among the travel mode share in San Francisco, ride-souring services account for 2%. In comparison, the traditional taxi's mode share has declined to be less than 0.5%.

However, the great success of ride-sourcing services imposes a considerable vehicular travel demand to urban road systems, especially for the centralized downtown area. Since there are only a limited number of spaces in downtown areas for ride-sourcing vehicles (RVs) to park, vacant RVs can only cruise on roads while waiting to be matched with the next customer. In a road network filled with RVs, e.g., the San Francisco downtown area served by many Uber or Lyft vehicles, such cruising contributes additional traffic demand and slows down other traffic, making already-congested streets even worse. Such a phenomenon is expected to become much more severe when emerging shared-use mobility services play a major role in meeting travel needs in metropolitan areas. One intuitive solution is to segment a certain portion of road space to provide on-street parking for RVs to reduce cruising; however, such a plan also reduces road capacity and possibly yields more congestion. The research question is therefore how to find appropriate balance to determine the optimal provision of parking spaces for a given road network. To answer such a "big picture" question (Daganzo et al., 2012) or make a "sketchy decision" at a highly aggregated level (Nie et al., 2016), this paper develops a macroscopic modeling framework that uses only a few parameters but attempts to capture the major underlying mechanisms of the ride-sourcing system under parking provision.

In summary, focusing on aggregate behaviors in an e-hailing ride-sourcing system, this paper aims to propose an effective macroscopic modeling framework to describe the operations of the system and capture the impacts of parking provisions. A conceptual framework is firstly established integrating components concerning each aspect of a ride-sourcing system. We then employ a deductive approach to further develop the framework based on a simplified matching mechanism between RVs and customers. The derived model is applied to formulate the optimal parking provision strategies as mathematical programs. Numerical experiments are generated to compare the properties and performance of ride-sourcing systems and the corresponding parking provision strategies under two different market structures.

* Corresponding author. Tel.: +1-734-764-8249. *E-mail address:* yafeng@umich.edu

2. A Conceptual Framework for E-hailing Ride-Sourcing Systems

We consider an e-hailing ride-sourcing system operating under the provision of dedicated parking spaces to RVs. In the system, after serving a customer, an RV starts searching for an empty parking space (shown as Fig. 1). The process terminates with the RV either successfully finding a parking space or being matched to a waiting customer by an online platform. More specifically, if a new dispatch order arrives before the RV finding an open parking space, it stops parking search and starts serving the customer; otherwise, the RV will cruise until finding a vacant space, and then park there and wait for the next dispatch.

To facilitate the presentation, we decompose the framework into six components including three processes and three characteristic functions. The three processes capture respectively the aggregate matching between waiting RVs and customers, the meeting of matched RVs and customers, and the search for empty parking spaces by cruising RVs. The first characteristic function is the customer demand function relating the customer hourly arrival rate to the average cost perceived for each trip. The second one implies the RV fleets conservation. And the third characteristic function delineates the relationship between the system's average speed and density under a network or macroscopic fundamental diagram (e.g., Mahmassani et al., 1984; Geroliminis and Daganzo, 2008). The following presents each component. A complete notation list is included in *Appendix A* of this paper.

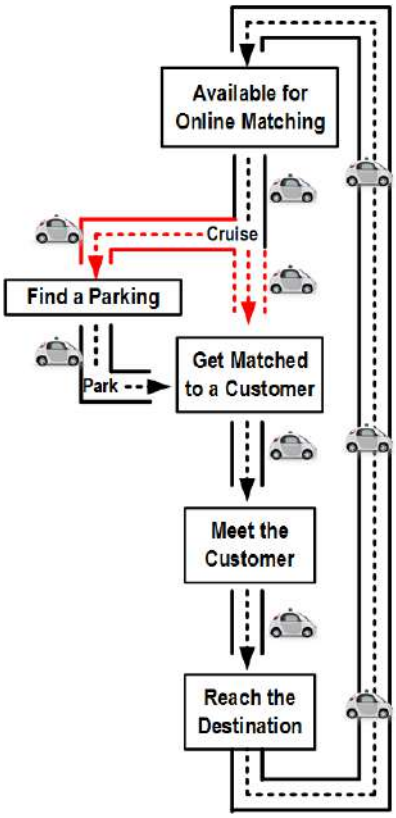


Figure 1. The workflow of an RV with e-hailing

2.1. Online matching process of unmatched RVs versus customers

An aggregate matching function can be used to characterize the online matching frictions between unmatched RVs and customers (e.g., Douglas, 1972). To capture the competition among drivers and customers, we apply the following matching function suggested by Yang and Yang (2011):

$$m^{c-t} = M^c(J^u, N^u)$$

where m^{c-t} represents the rate of matchings per hour; J^u and N^u denote the number of unmatched customers and RVs, respectively.

2.2. Meeting process of matched RVs versus customers

After the online match, the time before the physical meeting, denoted as w^m, is identical for the matched RV and customer. Intuitively, the average meeting distance r^m is a function of the densities of both unmatched RVs and customers. Assuming the total area served by RVs is constant, we have:

$$w^m \cdot v = D^m(J^u, N^u)$$

where the function D^m depends on the specific matching mechanism adopted by the online platform, and v denotes the average network running speed.

2.3. Search process of cruising RVs versus vacant parking spaces

The search process of cruising RVs for vacant parking spaces can be aggregately viewed as another matching process:

$$m^{p-t} = M^p(K^v, N^c, v, w^t)$$

where m^{p-t} is the rate of matchings per hour; K^v is the number of vacant parking spaces; N^c is the number of cruising RVs; and w^t is average waiting time for an RV to be matched with a customer. Different from online matching, the friction for a cruising RV to meet with a vacant parking space also depends on travel time between them (and thus the network running speed v) (Liu and Geroliminis, 2016). Moreover, the parking search process can be terminated if the RV gets matched first and thus the matching rate of parking m^{p-t} will depend on how quickly the online customer matching can take place, i.e., w^t.

2.4. Customer demand function

Let Q denote customer demand per hour and w^c be the average waiting time for customers getting matched; L_t, F and C are the travel distance, fare and total monetary cost for each trip, respectively. Define β and γ respectively as the out-of-vehicle and in-vehicle value of time (\$/h). Then, the customer demand Q is given as a monotonically decreasing function of the total trip cost, i.e.

$$Q = f\left(F + \beta(w^c + w^m) + \gamma\frac{L_t}{v}\right), \text{ where } f' < 0.$$

2.5. Fleet size conservation

The conservation of RV fleets requires

$$N = N^u + N^m + N^o$$

where N denotes the total RV fleet size; N^u, N^m and N^o respectively represent the instantaneous number of RVs under matching, meeting and serving processes.

2.6. Speed versus density

Instead of assuming a constant travel time, this study applies a macroscopic or network fundamental diagram to relate the average running speed with the average vehicular density within the network. Denote N^p, N^b respectively as the number of parked RVs and background vehicles, then $N - N^p + N^b$ represents the total number of running vehicles within the network. Then, let the constants k_j and A_u respectively indicate the jam density and the total utilizable area for roads and curbside parking. And denote variable K as the total number of parking spaces. Assuming k_j is also identical to the density in a fully-occupied parking lot, we have the term $k_j A_u - K$ equals to the effective road spaces for running vehicles.

By replacing the density with the division of the number of vehicles over road spaces, the network speed-density relation can thus be written as,

$$v = V\left(\frac{N - N^p + N^b}{k_j A_u - K}\right), \text{ where } N^b = g\left(\frac{L_b}{v}\right)\cdot\frac{L_b}{v} \text{ and } V',\ g' < 0.$$

The constant L_b denotes the average travel distance of the background traffic, while the function g relates background traffic demand with the average travel time cost.

2.7. Stationary states of a ride-sourcing system under parking provision

By integrating the aforementioned six decomposed components and substituting all the excessive variables therein with their equivalences under the stationary condition, we establish a conceptual framework (shown as the second column in Table 1) for a ride-sourcing system under parking provision. This conceptual framework is flexible enough to model systems under various operating mechanisms and conditions.

In this study, to facilitate the investigation of interactions between e-hailing ride-sourcing and parking systems, a simplified matching mechanism is introduced to analytically establish the relationships between variables in the matching processes. Subsequently, a deductive realization of the conceptual framework is further derived (shown as the third column in Table 1), and then applied to examine the parking provision strategies in different scenarios.

Table 1. Summary of conceptual framework and its deductive realization

Ride-sourcing System (⁕)	**Conceptual Framework**	**Deductive Realization**
Online Matching	$Q = M^c(w^c Q,\ w^t Q)$	$w^t = \frac{\tau}{1-\exp\left(-\frac{w^c}{w^t}\right)}$

Meeting Distance	$w^m \cdot v = D^m(w^c Q, w^t Q)$	$w^m \cdot v = \frac{\mu \cdot \left[\frac{1}{2}\sqrt{\frac{A_W}{Qw^c}} \cdot \mathrm{erf}\left(\sqrt{\frac{w^c}{w^t}}\right) - \frac{1}{\sqrt{\pi}}\sqrt{\frac{A_W}{Qw^t}} \cdot \exp\left(-\frac{w^c}{w^t}\right)\right]}{1-\exp\left(-\frac{w^c}{w^t}\right)}$
Parking Search	$\frac{N^p}{w^t} = M^p(K - N^p, Qw^t - N^p, w^t, v)$	$Q \cdot w^{t\,2} = N^p \cdot \left(w^t + \frac{L_r}{(K-N^p)\cdot v}\right)$
Fleet Conservation	$N = Q \cdot \left(w^t + w^m + \frac{L_t}{v}\right)$	$N = Q \cdot \left(w^t + w^m + \frac{L_t}{v}\right)$
Demand Function	$Q = f\left(F + \beta(w^c + w^m) + \gamma\frac{L_t}{v}\right)$	$Q = Q_0 \cdot \exp\left[-\kappa\left(F + \beta(w^c + w^m) + \gamma\frac{L_t}{v}\right)\right]$
Speed-Density Relation	$v = V\left(\frac{N - N^p + g\left(\frac{L_b}{v}\right)\cdot\frac{L_b}{v}}{k_j A_u - K}\right)$	$v = v_f \cdot \left(1 - \frac{N - N^p + Q_b^0 \cdot \exp\left(-\kappa_b \cdot \frac{L_b}{v}\right)\cdot\frac{L_b}{v}}{k_j A_u - K}\right)$

We use the label $(*)$ to represent the entire nonlinear modeling system. Within Sys. $(*)$, there are nine unknowns in total, i.e., the customer and RV waiting time for getting matched w^c and w^t, the meeting time after being matched w^m, the number of parked RVs N^p, the customer demand Q, the network average speed v, the fare F, the total fleet size N and the total number of parking spaces K. If the latter three variables are specified by stakeholders, e.g., a government agency determines the parking provision K while a ride-sourcing service provider decides its price F and fleet size N, the solution to the nonlinear system prescribes an estimate of the performance for the ride-sourcing system and road network.

2.8. Numerical example for impacts of parking provision

In this numerical example, we consider a Manhattan-like area of 20 square miles served by 50,000 RVs. The average trip distance of ride-sourcing orders is 2.5 miles, and each trip on average costs $15. Other parameters used in the example can be found in *Appendix A*. We then vary the number of parking spaces K in the area and investigate its impacts on system performance by solving the deductive model. Figure 2 displays the changes of two performance measures, including the realized hourly ride-sourcing and background traffic demand Q and Q^b (Fig. 2a) and the network average speed v (Fig. 2b). Both figures present a clear trade-off between the reduction of cruising and capacity drop caused by the allocation of road space to parking. As shown in the figures, the average speed v and customer serving rate Q increase initially and peak at the parking supply K being approximately 3.8×10^4, then decrease. This suggests that $K = 3.8\times10^4$ is a threshold where the system performance improves with respect to the increase in parking provision on the left side and gets worse on the right side.

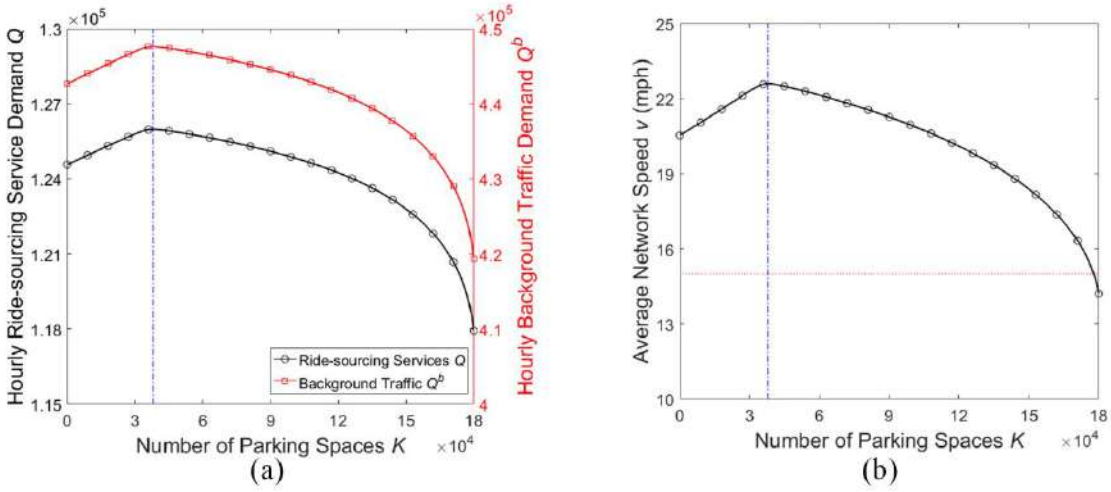


Figure 2. The impacts of the parking provision on the ride-sourcing system

Interestingly, Figure 2a shows that under a fixed fare and RV supply, parking provision imposes the same effect on the demand for ride-sourcing services and the background traffic. While better traffic condition will undoubtedly benefit the background traffic, reduced trip-serving time for a ride-sourcing system also means less travel cost for customers and higher demand to serve for drivers.

3. Time-Varying Optimal Parking Provision

Based on the modeling system (∗), we investigate the optimal parking provision for the ride-sourcing market. Two market structures are compared in this paper: 1) a system optimum case, where the government controls the fare F, the RV fleet size N and the parking spaces K to maximize the social benefit; 2) a monopoly case, where a private ride-sourcing company controls the fare F and the fleet size N to maximize its revenue, while the government only determines the size of parking spaces K to maximize the social benefit.

Ride-sourcing companies do not own the fleets themselves, but offer online matching services for those RVs/drivers registered to their platform. Thus, the fleet size in service N is endogenously determined depending on the real-time market condition. Nevertheless, the commission fee charged by the ride-sourcing company empowers it to indirectly control the number of vehicles in service. Denote F^c as the commission fee that drivers pay to the ride-sourcing company for each order or transaction. Then, the total hourly profit for the ride-sourcing company and RV drivers, respectively denoted as Π^c and Π^d, can be calculated as

$$\Pi^c = \left(F^c - c_p\right) \cdot Q$$

$$\Pi^d = (F - F^c) \cdot Q - (c_v N - c_s N^p)$$

where c_p is the unit-order operation cost of the e-hailing platform; c_v represents the hourly cost of operating a running RV; c_s denotes the hourly savings for a single parked RV. The term $c_v N - c_s N^p$ above represents the total hourly expenses for RV drivers operating their vehicles. Combined with the customers' benefits, the hourly social benefits S can be formulated as:

$$S = \int_0^Q f^{-1}(\omega|Q^0)d\omega - Q\left[\beta(w^c + w^m) + \gamma\frac{L_t}{v}\right] - \left(c_p Q + c_v N - c_s N^p\right) + c_b\left[\int_0^{Q^b} g^{-1}(\omega|Q_b^0)d\omega - Q^b\frac{L_b}{v}\right]$$

where c_b denotes the hourly cost for each vehicle and its passengers in the background traffic. The first integral term represents consumers' surplus; the second and third terms respectively indicate the total hourly time costs for customers and operational expenses of the ride-sourcing system; the last term represents the total benefit of travelers in background traffic.

Assume the labor market of drivers is perfectly competitive and the RV supply N is realized at a level such that $\Pi^d = 0$. Then, the system optimum (SO) case of the ride-sourcing market solves the nonlinear optimization model below:

$$\max\left\{S\middle|F, N, K \geq 0; \Pi^{\mathrm{d}} = 0;\ \text{Sys.}\,(*)\right\}$$

And the monopoly (MP) case solves the following bi-level program:

$$\max\left\{S\middle|K \geq 0; (F^*, N^*) = \operatorname{argmax}\{\Pi^c|F, F^c \geq 0;\ \Pi^d = 0;\ \text{Sys.}\,(*)\}\right\}$$

4. Major Findings

Comprehensive numerical experiments are conducted to compare the parking provision strategies and their impacts under the above two market structures. It is evident that a time-varying optimal parking provision can balance between reduction of cruising for RVs and capacity drop of a road system. Here, we highlight two major findings regarding the parking provision strategies for the government:

- *For non-peak hours, the optimal number of parking spaces is inversely proportional to the instantaneous customer demand. No space is needed for peak hours with demand surge.* The company under both SO and MP raises the fare and commission fee at peak hours, then lowers them to encourage ridership when demand drops. The system during peak hours behaves like a seller's market where RVs receive an instant match while customers endure longer waiting times; conversely, at non-peak hours, the system switches to a buyer's market and RVs must wait. Thus, the need for parking spaces considerably increases for non-peak hours, especially after midnight.
- *Even though there are fewer RVs in service under the monopoly scenario, a larger supply of parking spaces should be provided as compared to the system optimal case*. Because the customer demand under MP shrinks even more than the vehicle supply in comparison to SO, the customer's waiting time for online matching w^c is shorter. Moreover, since the network average speed v is mostly higher under MP, the service time is also much less. Consequently, the customers in MP can enjoy a better level of service but with a higher fare, while the longer waiting time of RVs in MP yields a greater need for parking spaces.

Acknowledgements

The work described in this paper was partly supported by the U.S. National Science Foundation (CMMI-1724168; CMMI-1740865).

Appendix A. Nomenclature

A.1. System state variables

Notation	Interpretation
F	Trip fare for an average distance L_t (\$)
N	RV fleet size
K	Number of provided parking spaces
w^c	Average customer waiting time for getting matched (h)
w^t	Average RV waiting time for getting matched (h)
w^m	Average meeting time for customers and RVs after being matched (h)
N^p	Number of parked RVs
Q	Hourly ride-sourcing customer demand
$\boldsymbol{v}$	Average network travel speed (mph)

A.2. Other state variables

Notation	Interpretation
m^{c-t}	Hourly matching rate between unmatched customers and RVs
J^u, J_d^u	Total number and density of unmatched customers
N^u, N_d^u	Total number and density of unmatched RVs
m^{p-t}	Hourly matching rate between cruising RVs and vacant parking spaces
K^v	Number of vacant parking spaces
N^c	Number of cruising RVs
N^b	Number of vehicles in background traffic
S	Total social benefits (\$)
Π	Total profit of ride-sourcing company/drivers (\$)
F^c	Commission fee that the ride-sourcing company charged per order

A.3. Parameters and their default values in the numerical examples

Notation	Interpretation	Default Value
τ	Interval of matching steps (h)	3×10^{-3}
A_w	Area of the whole region under RV service (mi^2)	20
L_r	Total length of the road system (mi)	500
μ	Parameter adjusts Euclidean distance to Manhattan distance	$\sqrt{\pi/2}$
Q^0	Potential ride-sourcing customer demand per hour	1.5×10^7
Q_b^0	Potential background traffic demand per hour	1.5×10^7
κ	Demand sensitivity parameter for ride-sourcing services (1/\$)	0.25
κ_b	Demand sensitivity parameter for background traffic (1/h)	1
β	Out-of-vehicle value of time (\$/h)	40
γ	In-vehicle value of time (\$/h)	20
v_f	Free-flow travel speed (mi/h)	30
k_j	Jam density (veh/mi · lane)	250

A_u	Total utilizable area for roads and parking facilities (lane·mi)	2000
L_b	Average travel distance in background traffic (mi)	5
L_t	Average distance of trips served by ride-sourcing services (mi)	2.5
c_b	Hourly cost for each single vehicle and its passengers in background traffic ($)	30
c_p	Operational cost for the e-hailing platform per order ($)	2
c_v	Hourly operational cost for a running RV ($)	25
c_s	Hourly savings for a single parked RV ($)	10

References

Anderson, D. N. (2014). "Not just a taxi"? For-profit ridesharing, driver strategies, and VMT. *Transportation*, *41*(5), 1099-1117.

Arnott, R. (1996). Taxi travel should be subsidized. *Journal of Urban Economics*, 40, 316-333.

Daganzo, C. F., Gayah, V. V., & Gonzales, E. J. (2012). The potential of parsimonious models for understanding large scale transportation systems and answering big picture questions. *EURO Journal on Transportation and Logistics*, *1*(1-2), 47-65.

Douglas, G.W. (1972). Price regulation and optimal service standards: The taxicab industry. *Journal of Transport Economics and Policy* 6, 116-127.

Geroliminis, N., & Daganzo, C. F. (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. *Transportation Research Part B: Methodological*, 42(9), 759-770.

Hall, J. V., & Krueger, A. B. (2015). An analysis of the labor market for Uber's driver-partners in the United States. *Princeton University Industrial Relations Section Working Paper*, 587.

He, F., & Shen, Z. J. M. (2015). Modeling taxi services with smartphone-based e-hailing applications. *Transportation Research Part C: Emerging Technologies*, *58*, 93-106.

Lazo, L. (2015). Uber turns 5, reaches 1 million drivers and 300 cities worldwide. Now what? https://www.washingtonpost.com/news/dr-gridlock/wp/2015/06/04/uber-turns-5-reaches-1-million-drivers-and-300-cities-worldwide-now-what/?utm_term=.26e3be6faefe. Accessed Dec. 4, 2016.

Liu, W. & Geroliminis, N. (2016) Modeling the morning commute for urban networks with cruising-for-parking: An MFD approach. *Transportation Research Part B: Methodological*, 93, 470-494.

Mahmassani, H.S., Williams, J.C., Herman, R. (1984). Investigation of network-level traffic flow relationships: some simulation results. *Transportation Research Record: Journal of the Transportation Research Board*, No. 971, 121-130

Nie, Y., Ghamami, M., Zockaie, A. and Xiao, F. (2016). Optimization of incentive policies for plug-in electric vehicles. *Transportation Research Part B: Methodological*, 84, 103-123.

Rayle, L., Dai, D., Chan, N., Cervero, R., & Shaheen, S. (2016). Just a better taxi? A survey-based comparison of taxis, transit, and ridesourcing services in San Francisco. *Transport Policy*, 45, 168-178.

San Francisco Municipal Transportation Agency, 2015, Travel Decisions Survey 2015. https://www.sfmta.com/sites/default/files/reports/2016/Travel%20Decision%20Survey%202015%20Report_Accessible.pdf. Accessed December 18, 2016.

Wang, X., He, F., Yang, H., & Gao, H. O. (2016). Pricing strategies for a taxi-hailing platform. *Transportation Research Part E: Logistics and Transportation Review*, 93, 212-231.

Yang, H., & Yang, T. (2011). Equilibrium properties of taxi markets with search frictions. *Transportation Research Part B: Methodological*, 45(4), 696-713.

Traffic & Mobility

SA2: Simulation-Based Analysis and Optimization

Saturday 9:00 – 11:00 AM

Session Chair: Ronghui Liu

9:00 **Traffic Management Strategies for Trucks in Urban Environments Based on a Fast Traffic Simulation Algorithm**
Michele Simoni, Christian Claudel*
University of Texas at Austin

9:30 **Simulation Based Quantification of the Potential Impacts of Incidents on Connected Vehicle Applications**
Abdullah Kurkcu, Fan Zuo, Jingqin Gao, Kaan Ozbay*
New York University

10:00 **An Efficient Sampling Method for Stochastic Simulation-based Transportation Optimization**
Timothy Tay, Carolina Osorio*
Massachusetts Institute of Technology

10:30 **A Within-Day Microscopic Dynamical Model of Route Choice and Responsive Traffic Signal Control**
[1]Ronghui Liu, [2]Mike Smith*
[1]University of Leeds, [2]University of York

Traffic management strategies for trucks in urban environments based on a fast traffic simulation algorithm

Michele D. Simoni[a], Christian G. Claudel[b]

[a]Dept. of Civil, Architectural, and Environmental Engineering, University of Texas at Austin, 301E Dean Keeton, St., Austin, 78712 TX, USA
[b]Dept. of Civil, Architectural, and Environmental Engineering, University of Texas at Austin, 301E Dean Keeton, St., Austin, 78712 TX, USA

Keywords: Urban freight transportation; Traffic simulation; LWR model; Real-time traffic management; Memetic algorithms;

Introduction

Urban freight distribution and traffic congestion are two sides of the same coin. In fact, it is well known that urban freight movements have a negative impact on the transportation network. Besides safety and environmental externalities, trucks generate congestion because of their lower speed, their reduced maneuverability, and their frequent stops for deliveries (Dablanc, 2007). Conversely, high levels of traffic delay truck deliveries and compromise the performance and reliability of freight distribution, ultimately increasing the overall costs of the carriers (McKinnon, 1999).

During the last decade, technologic improvements in wireless communication, computational and sensing technologies, have paved the way to a series of applications aimed at providing a more efficient transportation system (e.g. dynamic route guidance, real time traffic management, and electronic toll collection) labeled as Intelligent Transportation Systems (ITS). Moreover, automobile and hi-tech companies have made significant progress in computerizing vehicles. While the potential benefits of these strategies and technologies have often been praised for passenger transportation, very little is known about their applications and impact in the area of freight transportation.

Real time traffic management (TM) represents a relatively straightforward solution to optimize the usage of available network space without requiring significant infrastructure investments. Perhaps the most challenging tasks to achieve effective solutions, consist in: quickly exchanging communication (inbound and outbound); properly modeling the impacts of trucks on traffic flows in urban settings, such as arterial roads and local streets; rapidly identifying possible optimization strategies.

While the first issue can be addressed nowadays by technological advances such as GPS, and wireless communications, little research has been done in deriving fast and reliable computational methods to describe the effects of this typology of slow vehicles. Microscopic traffic models (car-following) can do that very well, but they are usually time consuming and can be difficult to calibrate. In this study, the dynamics of traffic is described using the classical Lighthill-Whitham-Richards (LWR) macroscopic traffic model. Slow vehicles (such as trucks) that do not behave like the rest of traffic are described as "moving bottlenecks". The moving bottleneck theory has been described for a long time, for example in Newell (1993; 1998), Munoz and Daganzo (2002), Leclerq et al. (2004) and Daganzo and Laval, (2005). However, simulating multiple moving bottlenecks and their interaction with traffic is complex, since bottlenecks both influence and are influenced by the surrounding traffic. To date, no algorithm that can endogenously simulate an arbitrary number of moving bottlenecks has been proposed in the literature. Our present objective is to show that fast semi-analytic algorithms such as the one proposed by Mazaré et

al. (2011) can compute the solutions associated with very large numbers of interacting moving bottlenecks extremely quickly, and almost exactly. This is extremely useful in the context of logistics, since it allows us to quickly solve large-scale optimization problems in which the objective depends on both the state of traffic and the trajectories of the trucks

As a main contribution of this paper we present and evaluate two alternative TM strategies for trucks in urban settings based on a macroscopic traffic flow model. The first one consists in joint coordination of traffic lights and trucks departures on an arterial corridor to maximize its throughput. The second one consists in a parking-loading curbside management strategy for reducing delays associated with trucks' deliveries. The first strategy could be employed in the surroundings of large urban freight traffic generators, such as airports, marine ports and container terminals, but also facilities like shopping malls, hospitals, colleges and universities and government offices (Jaller et al., 2015). The second strategy could be implemented in locations characterized by a significant number of deliveries like central business districts and commercial areas (Nourinejad and Roorda, 2016). In both situations the proposed traffic management algorithms would require some degree of vehicle connectivity and positioning (for example through GPS), which can be achieved using smartphone apps.

The optimization problems are solved by using the Memetic Algorithm (MA) meta-heuristic, an extension of the population-based hybrid genetic algorithms (GAs) coupled with a local search procedure that allows for refinements of the solutions. While, several applications of genetic algorithm have been proposed in the field of traffic control, to the best of our knowledge, MAs have not been utilized yet for this typology of problems. This particular technique is particularly suitable to situations where the objective function cannot be derived analytically, but only by means of simulations. Furthermore, combining an extensive search of the best zones in the search space (exploration) with a more detailed search in zones with potential better solutions (exploitation) seems to work well for large problems (Cotta 2012) and they can often provide better results than other well known approaches like Genetic Algorithm, Tabu Search, and Simulated Annealing (Garg, 2010).

Fast semi-analytical traffic solver

In traffic flow theory, different typologies of "slow" vehicles (or platoons) can be modeled as moving bottlenecks. These obstructions in traffic streams are usually associated with the presence of buses in urban traffic, and trucks or simply slower vehicles on highways. All these situations, indeed, are characterized by a partial blockage the road, causing a capacity reduction. Several studies have highlighted the importance of the effects of moving bottlenecks on traffic and have developed methodologies to include them into existing traffic models (Gazis and Herman, 1992; Newell, 1993; Newell, 1998; Munoz and Daganzo, 2002; Leclerq, 2004; Daganzo and Laval, 2005). The main difficulty associated with moving bottlenecks is the fact that moving bottlenecks are following an hybrid ordinary differential equation (ODE), which is coupled with the Partial Differential Equation (PDE) corresponding to the traffic flow model, making it particularly difficult to solve in the general case (multiple moving bottlenecks with complex initial and boundary conditions).

In this paper, we adopt the model by Simoni and Claudel (2017) who tackle this problem with an efficient formulation that compute the parameters associated with moving and fixed bottlenecks without computing the complete solution to the LWR model to derive their trajectories and their effects on surrounding traffic. Thanks to that, computational times are considerably improved compared to classical

numerical schemes, without affecting the computational accuracy. The algorithm adopted is conceptually illustrated in Figure 1, but additional explanations can be found in Simoni and Claudel (2017). The solution approach in a nutshell, is based on the fact that moving bottlenecks can be inactive due to presence of congestion (the average speed is lower than the maximum velocity of the moving bottleneck) or to lack of traffic (capacity reduction does not affect the flow). In these cases, since traffic downstream and upstream the moving bottleneck is not affected, no particular change in the solution is considered. Only when the regular traffic is travelling at higher speed and the capacity is not sufficient to allow everyone to overtake, the moving bottleneck becomes active. In this case, an internal boundary condition is created to account for that. In order to model several active moving bottlenecks with different features, interacting with each other, the algorithm uses the inf-morphism property of solutions and the existence of domains of influence.

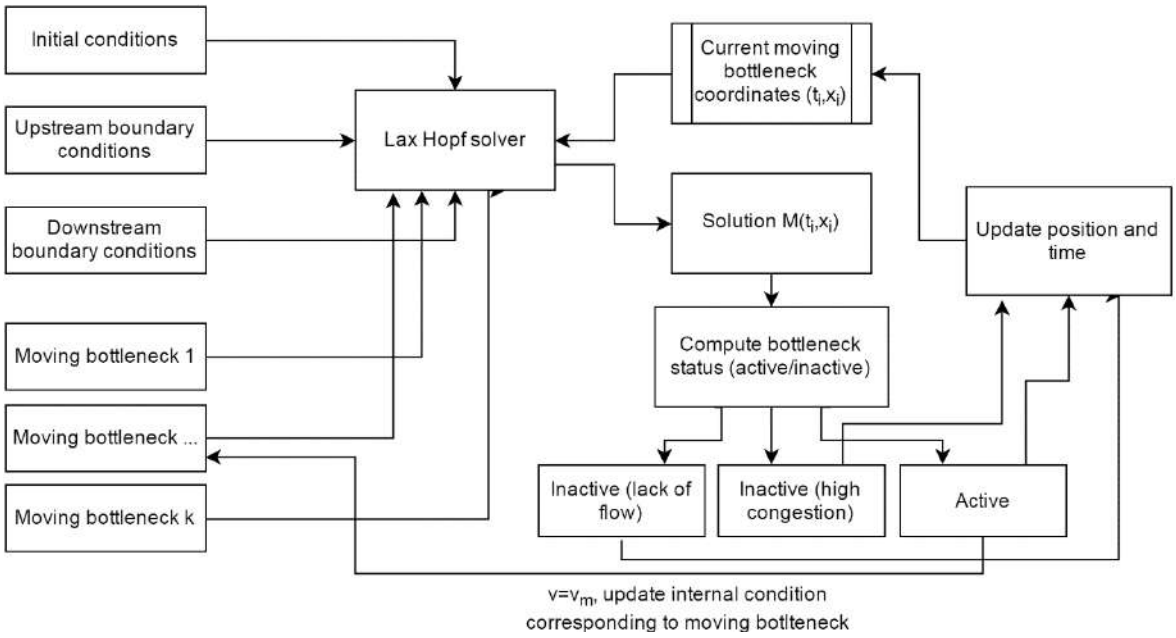


Figure 1: Overview of the traffic simulation algorithm for multiple moving bottlenecks

Optimization problems

Joint coordination of traffic lights and trucks departures problem

The objective function corresponds to the total outflow, that is, the cumulative number of vehicles at the downstream end of the link, N_d. In this problem, the total outflow depends on the entry times of m trucks (t_i) and the on the signal settings $\varphi_j = (c, \phi)$, where c and ϕ correspond respectively to the cycle time length and the green time length of traffic signal j. Each solution corresponds to an array specifying the entry times of the trucks, and the cycle length and green time of the traffic light. Hence the problem can be expressed as:

$$\max_{t_i, \varphi_j} N_d(t_i, \varphi_j)$$

subject to:

1. $t_{i,min} \leq t_i \leq t_{i,max} \; \forall i \in I$ moving bottlenecks' entry times constraints

2. $\varphi_j(c, \phi) \begin{cases} c_{min} \leq c \leq c_{max} \\ \phi_{min} \leq \phi \leq \phi_{max} \end{cases} \quad \forall j \in J$ cycle times and green times constraints

The road stretch studied corresponds to a two-lanes link of 300 meters with a traffic light is situated at the downstream end of the link. A time period of 900 s is simulated, during which 5 trucks enter from a location upstream the link (which represents the freight traffic generator), each characterized by different maximum speed and entry time.

Parking-loading curbside management strategy

In this case, the problem is formulated again, as a maximization of total outflow. However, in this case, the objective function depends only the delivery locations d_j of trucks *j*. The available locations for deliveries are determined by discretizing the link in "slots" of length *l* each. Hence, the problem turns into a combinatorial optimization problem. Each solution corresponds to a array specifying the delivery locations of each truck. No specific constraint is applied, except for the exclusion of already occupied locations from the range of available ones when assigning subsequent trucks.

We investigate this problem on the same road stretch analyzed before and for the same period of time. In this case 10 trucks enter at random times and stop at the curbside for a random periods of time between 60 s and 300 s.

Memetic Algorithm

Similarly to GAs, MAs are very flexible as they do not they do not require any knowledge of the gradient and they can avoid getting stuck in local optima (Teklu et al., 2007). In addition, MAs are more suitable to simulation-based frameworks as derivative-based optimization methods require the knowledge of the analytical form of the problem. For these reasons, GAs have been already employed to solve different joint traffic control and assignment problems (Foy et al., 1992; Lee and Machemehl, 1998; Yin, 2000; Ceylan and Bell, 2004; Teklu et al., 2007). In this study, the MA adopted consists of the following steps:

Step 0 (Initial Population): Generating a population of n random solutions satisfying the problem constraints (if any). The fitness value of each solution (according to the optimization problem) is calculated by performing a simulation with the fast algorithm presented in the previous section.

Step 1 (Parents selection): Parents are chosen by means of the tournament selection procedure, which consists in the selection of best performing solutions among a restricted pool of randomly chosen solutions from the population

Step 2 (Breeding): Children are derived by means of a two-point crossover procedure. In case of combinatorial optimization, "Order 1 Crossover" is used, which is an ordered crossover method applicable where direct swap is not feasible.

Step 3 (Mutation-Local Search): The traditional mutation operators like swapping or moving some nodes are replaced by a local search (LS) procedure LS. The LS scan a series of decision variables and performs alternative operations to check whether the solution can be improved. If so, the mutated solution replaces the original one. The LS procedure occurs with a certain predefined probability.

Results

Table 1: Computational performance of the algorithm for different population sizes (average between 20 tests) in the first problem

	Maximum outflow (veh)	Solution improvement (%)	Computation time (s)
Population 20	437	13.1	8.5
Population 30	439	13.5	12.6
Population 50	439	13.7	21.6

Table 2: Computational performance of the algorithm for different population sizes (average between 20 tests) in the second problem

	Maximum outflow (veh)	Solution improvement (%)	Computation time (s)
Population 20	193.7	12.6	42.7
Population 30	195.9	13.8	60.6
Population 50	197.8	15.0	102.0

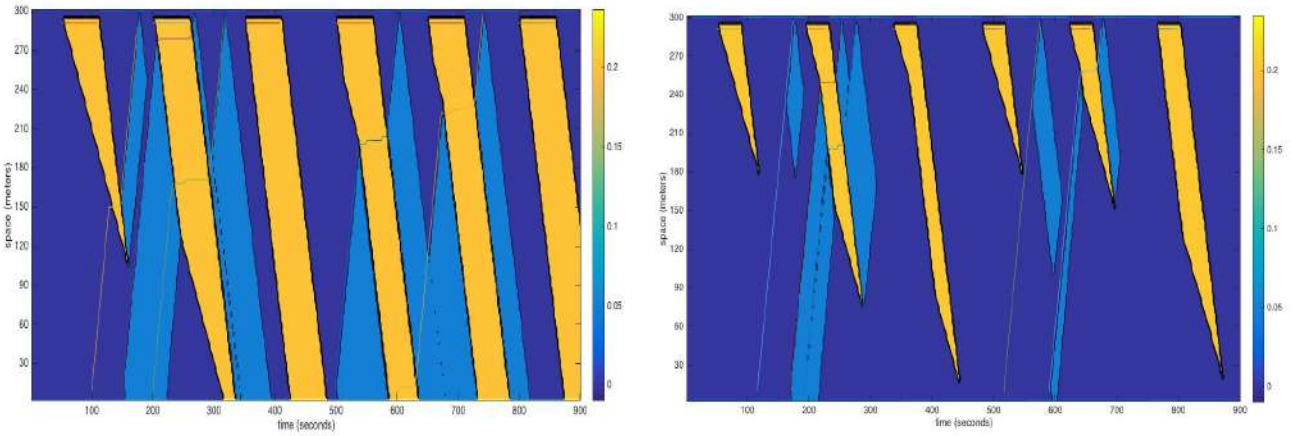


Figure 2: Space-time density diagram representing traffic on the link before and after the optimization in the first problem

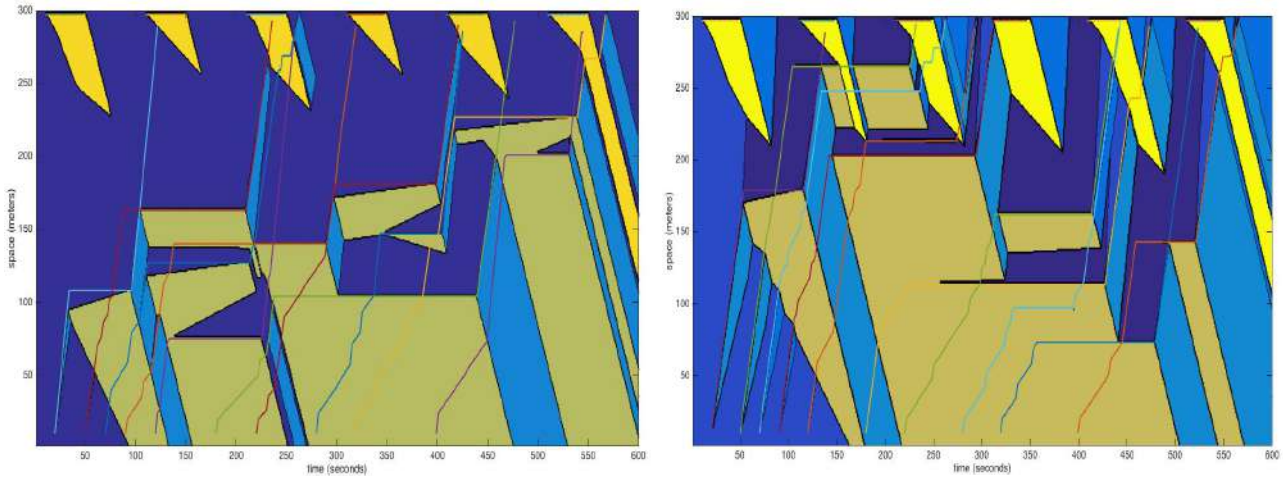


Figure 3: Space-time density diagram representing traffic on the link before and after the optimization in the second problem

References

Ceylan, H. & Bell, M. G. H. (2004). Traffic signal timing optimisation based on genetic algorithm approach, including drivers' routing. Transportation Research, 38B (4),329–342.

Cotta, C. (2012). Handbook of memetic algorithms (Vol. 379). F. Neri, & P. Moscato (Eds.). Heidelberg: Springer.

Dablanc, L. (2007). Goods transport in large European cities: Difficult to organize, difficult to modernize. Transportation Research Part A: Policy and Practice, 41(3), 280-285.

Daganzo, C. F., & Laval, J. A. (2005). Moving bottlenecks: A numerical method that converges in flows. Transportation Research Part B: Methodological, 39(9), 855-863.

Foy, M. D., Benekohal, R. F., & Goldberg, D. E. (1992). Signal timing determination using genetic algorithms. Transportation Research Record, (1365), 108.

Garg, P. (2010). A Comparison between Memetic algorithm and Genetic algorithm for the cryptanalysis of Simplified Data Encryption Standard algorithm. arXiv preprint arXiv:1004.0574.

Gazis, D. C., & Herman, R. (1992). The moving and "phantom" bottlenecks. Transportation Science, 26(3), 223-229.

Jaller, M., Wang, X. C., & Holguin-Veras, J. (2015). Large urban freight traffic generators: Opportunities for city logistics initiatives. Journal of Transport and Land Use, 8(1), 51-67.

Leclercq, L., Chanut, S., & Lesort, J. B. (2004). Moving bottlenecks in Lighthill-Whitham-Richards model: A unified theory. Transportation Research Record: Journal of the Transportation Research Board, 1883(1), 3-13.

Lee, C. & Machemehl, R. B. (1998). Genetic algorithm, local and iterative searches for combining traffic assignment and signal control, Traffic and Transportation Studies: In: Proceedings of ICTTS 98, 489-497.

Mazaré, P. E., Dehwah, A. H., Claudel, C. G., & Bayen, A. M. (2011). Analytical and grid-free solutions to the Lighthill–Whitham–Richards traffic flow model. Transportation Research Part B: Methodological, 45(10), 1727-1748.

McKinnon, A. (1999). The effect of traffic congestion on the efficiency of logistical operations. International Journal of Logistics: Research and Applications, 2(2), 111-128.

Munoz, J. C., & Daganzo, C. F. (2004). Moving bottlenecks: a theory grounded on experimental observation. University of California Transportation Center.

Newell, G. F. (1993). A simplified theory of kinematic waves in highway traffic, Part II: Queueing at freeway bottlenecks. Transportation Research Part B: Methodological, 27(4), 289-303.

Newell, G. F. (1998). A moving bottleneck. Transportation Research Part B: Methodological, 32(8), 531-537.

Nourinejad, M., & Roorda, M. J. (2016). Parking enforcement policies for commercial vehicles. Transportation Research Part A: Policy and Practice.

Simoni and Claudel (2017). A fast semi-analytic algorithm for computing solutions associated with multiple moving or fixed bottlenecks: Application to joint scheduling and signal timing. arXiv:1701.01026

Teklu, F., Sumalee, A., & Watling, D. (2007). A genetic algorithm approach for optimizing traffic control signals considering routing. Computer - Aided Civil and Infrastructure Engineering, 22(1), 31-43

Yin, Y. (2000). Genetic-algorithms-based approach for bilevel programming models. Journal of Transportation Engineering

SIMULATION BASED QUANTIFICATION OF THE POTENTIAL IMPACTS OF INCIDENTS ON CONNECTED VEHICLE APPLICATIONS

Abdullah Kurkcu, M.Eng.
Graduate Research Assistant,
Department of Civil & Urban Engineering,
CitySmart Laboratory at UrbanITS Center,
Tandon School of Engineering,
Center for Urban Science + Progress (CUSP),
New York University (NYU)
One Metrotech Center, 19th Floor
Brooklyn, NY 11021, USA
ak4728@nyu.edu

Fan Zuo, M.Sc.
Graduate Research Assistant,
Department of Civil & Urban Engineering,
CitySmart Laboratory at UrbanITS Center,
Tandon School of Engineering,
Center for Urban Science + Progress (CUSP),
New York University (NYU)
Six MetroTech Center, 4th Floor,
Brooklyn, NY 11021, USA
fz380@nyu.edu

Jingqin Gao, M.Sc.
Center for Urban Science + Progress (CUSP),
New York University (NYU)
Six MetroTech Center, 4th Floor,
Brooklyn, NY 11021, USA
Center for Urban Science + Progress (CUSP),
New York University (NYU)
Six MetroTech Center, 4th Floor,
Brooklyn, NY 11021, USA
jingqin.gao@nyu.edu

Kaan Ozbay, Ph.D.
Professor,
Department of Civil & Urban Engineering,
CitySmart Laboratory at UrbanITS Center,
Tandon School of Engineering,
Center for Urban Science + Progress (CUSP),
New York University (NYU)
Six MetroTech Center, 4th Floor,
Brooklyn, NY 11021, USA
kaan.ozbay@nyu.edu

Keywords: Connected Vehicles, Traffic Simulation, Calibration, Validation, Basic Safety Messages (BSM)

1 INTRODUCTION

The proliferation of mobile wireless communication technologies has made possible the connectivity between vehicles (V2V), vehicle to infrastructure (V2I) and vehicle to any other agents (V2X) of the transportation network environment. The emerging Connected Vehicle (CV) platform based on the mobile and wireless technologies has the potential to transform the way people travel while promising to be an invaluable real-time data source that can provide information required to estimate traffic conditions on a network. Such real-time information -including speed, position, and acceleration- may also be used to understand, evaluate, and improve the existing transportation infrastructure in a way that was not possible just a few years ago. CV applications reviewed in the literature can be broadly divided into safety (*1-4*), mobility (*5-7*), environmental (*8; 9*) and other applications (*10-12*). However, the ability to effectively use the CV data relies on the market penetration rate of CV deployments and the prevailing traffic conditions. Therefore, it is imperative to develop a comprehensive simulation-based methodology for the accurate estimation of critical transportation performance measures for different market penetration rates of the CV technologies under various traffic conditions. Although it is really expensive and technologically challenging to conduct field tests for CV applications, such situations can be simply simulated using a calibrated traffic simulation network.

Evaluation of traffic operations under different conditions is a challenging task due to the difficulty in modeling drivers' complex tactical and strategic decision-making process. Clearly, the performance and validity of a microscopic traffic simulation model in assessing traffic management strategies under disruption scenarios depend on its ability to simulate actual traffic conditions (13) stochastically. The fundamental stage in the development and use of traffic simulation models as major prediction and decision-making tools is the calibration and validation of the parameters and models that characterize the traffic flow, such as lane changing, car following, gap acceptance and route choice (*13*). Calibration is the process of adjusting the simulator to maintain a statistically reasonable correspondence between the simulated and the actual traffic conditions and measurements. It is an iterative process until obtaining the desired confidence level where the model and its results are suitable for the objective. Validation is the final process to evaluate if each element in the model sufficiently simulates actual traffic conditions and the overall performance of the model is acceptable (*14*). The term "calibration" is used in this paper to describe the validation and calibration process together in this study. The input for simulation contains two essential groups of data (*15*):

- Physical Input Data (IS): Volume counts, capacity, and physical roadway features
- Calibration Parameters (CS): Adjustable components of driver behaviorz

$$S_S : f(I_S, C_S) \rightarrow O_{sim} \mid I_S, C_S + \varepsilon : O_{obs} \tag{1}$$

where

$f(I_S, C_S)$	=	Function specification of the internal models in a simulation system
I_S	=	Physical input data observed in the field
C_S	=	Set of calibration parameters for a simulation system
O_{sim}	=	Simulation output data, given the input data and calibrations
ε	=	Acceptable margin of error between simulation outputs
O_{obs}	=	Observed field data

In this study, five adjacent intersections in Brooklyn, NY are selected for the case study. These intersections on Jay Street are coded into the simulation network. Figure 1 below shows the location of the intersections and the generated traffic simulation model in Simulation of Urban Mobility (SUMO) -

an open-source simulation package (*16*). The model is then calibrated using the ground truth data collected for a morning peak period.

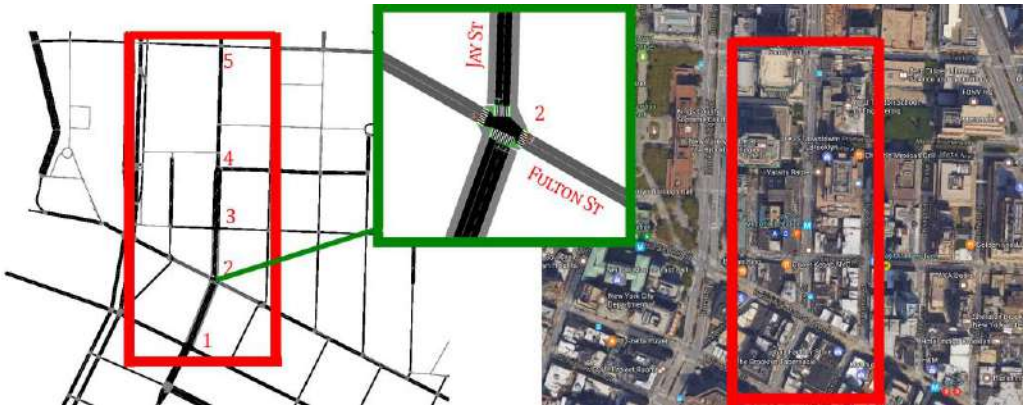


Figure 1: The Simulation Network

Data sources used to calibrate the network include real-time travel time information from online mapping services, video recordings of the Jay Street and Fulton Street intersection on different days; the data collected by fixed road-side wireless sensors to track vehicles, and manually collected turning movement counts. Once the simulation network is calibrated and validated, CV messages can be generated using a software called the Trajectory Conversion Algorithm (TCA) built by NYU CitySMART Transportation group and Noblis (*17; 18*). This software is designed to test different scenarios for producing, transmitting, and storing CV information. The fundamental contribution of this paper is to illustrate and quantify the effects of random incidents on the performance of CV applications using a well-calibrated traffic simulation model. The impact of different types of incidents with varying severity and duration on critical CV applications will be investigated to understand the effect of incidents on the accuracy of on and off line performance measures at various market penetration levels. The results of this simulation based study will be reported in the final version of this paper. The following section will briefly explain the proposed methodology, calibration efforts and the developed TCA tool for simulating CV messages (*17*).

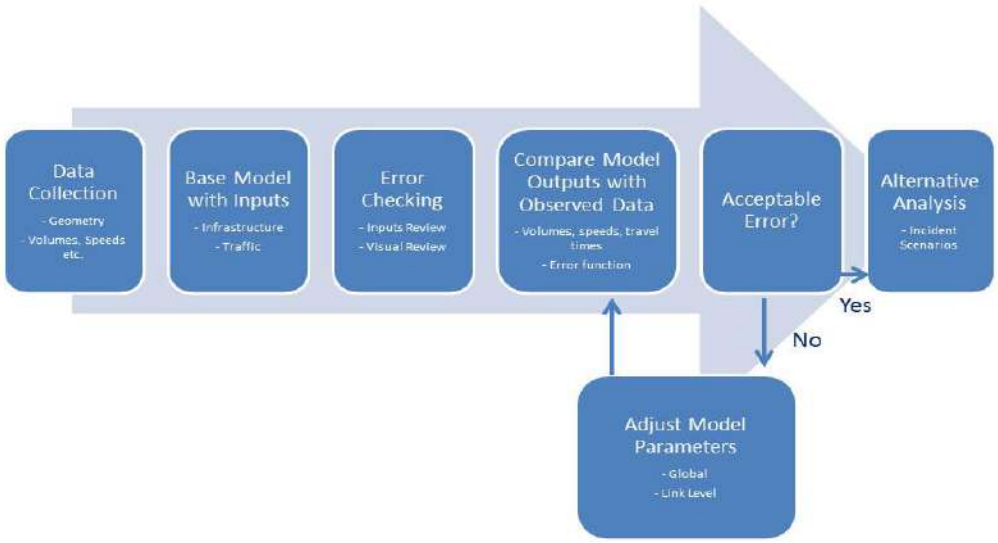


Figure 2: The Calibration Process

2 METHODOLOGY

2.1 Calibration and Validation of the Simulation Model

The process of simulation calibration requires adjusting the calibration parameters to minimize the difference between the output of the simulation and observed conditions to a predetermined acceptable level. Steps involved in building, calibrating and using a microscopic simulation model for scenario are shown in Figure 2.

2.1.1 Data Needs for Calibration

Traffic speeds, travel times and volumes are required to calibrate the network. In this paper, vehicle speeds, headways, and traffic volumes are extracted from video files recorded at the study location. Travel times are collected by using web-based real-time traffic resources such as Bing Maps (*19*) and fixed Bluetooth wireless sensors (*20; 21*) that capture MAC addresses of mobile devices traveling at the study location. The data collection methodology used in this paper integrates real-time sensor data from online traffic reporting websites and wireless sensors with historical traffic information gathered from video files to generate detailed snapshots of actual traffic conditions. Both data sources are used to build a calibrated time-dependent demand matrix for an average workday. This demand matrix is used to generate results that closely match real-time traffic conditions. Table 1 below summarizes data sources used in the calibration process.

Table 1: Data used in the Simulation Calibration Process

Data Type	Description
Travel Time	Extracted 5 minute travel times from videos recorded during a typical peak hour period
Travel Time	Collected through Bing Maps API
Travel Time	Collected using wireless sensors deployed at fixed locations
Events Data	Incidents and special events data around the study area
Turning Counts	Turning movement counts extracted from video files
Traffic Signals	Actual Traffic Signal Timings
Pedestrian Counts	Pedestrian counts and the number of jaywalkers extracted from the video files

All of the calibration/validation results including traffic speed, volume and travel time comparisons will be reported in the final version of the paper. In addition, as an additional performance measure, vehicle headways at the North Bound approach will be compared to the observed headways. After the traffic simulation is calibrated to reflect base case conditions, the simulation analyses will be performed for two different incident configurations. The comparative results of each scenario with respect to a base scenario without incidents will be reported in the final paper.

2.2 Generating CV Messages with the TCA Software

The developed open-source the TCA tool is one of the important functionalities of the BSM Data Emulator. The TCA can work in an online mode which will generate CV messages while the traffic simulation is running, or in an offline mode which will use the simulation outputs and convert them to CV messages. Figure 3 graphically illustrates the overall approach for converting vehicle trajectory files to CV messages using the offline mode of operation. In this method, vehicle trajectory files reporting vehicles' speed and position every 0.1 seconds are used as inputs to the TCA tool.

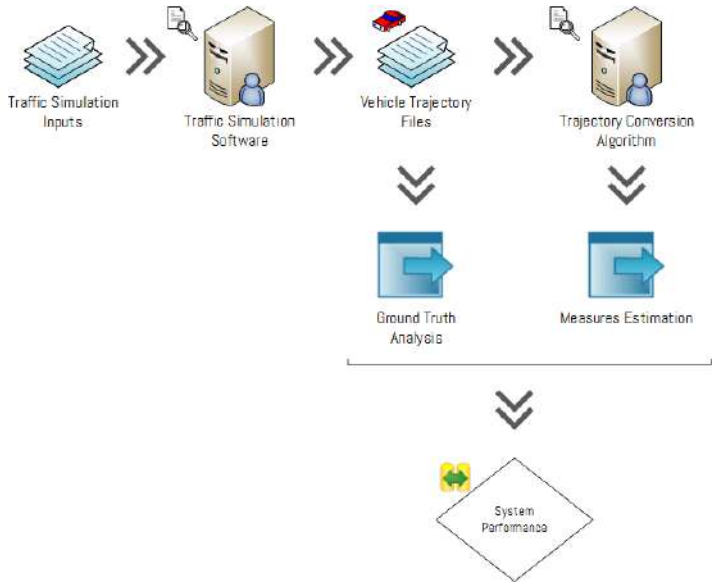


Figure 3: TCA Software - The Offline Mode of Operation

Although the existing version of TCA only supports PARAMICS (*22*) and VISSIM (*23*), vehicle trajectories generated by SUMO can be manipulated to be in the same output format with PARAMICS or VISSIM. A python code is written to convert SUMO's vehicle traces to the PARAMICS format and TCA-P offline version (*18*) is used to generate CV messages. Vehicles in the microscopic simulation can be equipped to generate and transmit Probe Data Messages (PDMs), Basic Safety Messages (BSMs), ITS Spot messages (SPOT), and/or European Cooperative Awareness Messages (CAM) with the TCA tool (*24*). In the extended abstract, we used only 100% market penetration level BSM messages which are generated every 0.1 seconds for analysis. After acquiring CV messages at various market penetration levels, transportation measures such as speed and travel time can be calculated using such messages. The accuracy of the estimated transportation measures can then be compared to the ground truth data retrieved from the simulation itself. In the final version of the paper, we will attempt to quantify the effects of different types of incidents on the accuracy of estimations by running the traffic simulation model for various scenarios and market penetrations.

3 EXPERIMENTAL SETUP

These scenarios are generated to understand the impact of incidents on the accuracy of estimated transportation measures using CV messages.

3.1 Base Scenario

The base scenario is simulated for the morning peak period from 8AM – 9AM. Simulation statistics are generated every 5 minutes for model calibration and validation. Although there are no incidents simulated in the base scenario, there are some incident-like events that take place. For example, there is a bus stop after the first intersection and buses usually double park for approximately 30-60 seconds to allow passengers to get on board. These disturbances as well as pedestrians jaywalking are included in the base scenario to reflect real-world daily situations. 3830 pedestrians per hour are observed at the first intersection from the video recording, and 21.6% of them are jay-walkers.

3.2 Double Parking Scenario

Two incident scenarios, double parking and work-zone, namely, are setup for testing the accuracy of CV applications under different traffic conditions. Double parking is a common phenomenon in urban areas. For this scenario, It has been observed that on average, 20 double parking events occur per hour on a section of Jay Street in Brooklyn, NY. Double parking events are randomly assigned to road sections with a mean interval of 180 seconds and mean duration of 190 seconds. It has been observed from the videos that 85% of double parking vehicles are passenger cars, 10% of them are vans, and 5% of them are light trucks. These observations are coded into the model accordingly.

3.3 Work-zone Scenario

Work-zone scenario used information from an existing work zone permit data collected from the New York City permit management system (*25*) that includes work-zone information such as the geo-location, permit start/end date, permittee name, and roadway type. Using this real work-zone permit information, a work zone scenario – an electric repair case - is simulated. This incident has been observed to affect the traffic flow significantly at this section. Vehicle trajectory files are generated for both scenarios and feed into the TCA tool.

4 PRELIMINARY RESULTS AND DISCUSSION

The TCA tool in the BSM software is only used to generate 100% market penetration rate BSM messages in the extended abstract. The format of the generated messages can be seen in Figure 4 below. Although 100% market penetration rate is used for the initial tests, it is still possible to lose some of the data due to the transmission distance of the messages. If messages are not received by other vehicles or road-side equipment (RSE), they are lost. The default transmission distance is assumed to be 20 ft. without any delays.

psn	acceleration	brakepressure	brakestatus	hardbreaking	heading	instantacc	lane	link	link_x	localtime	spd	transTo	transtime	trasnmission	x	y	index
49	-9999.0000	0.0000	0.0000	0.0000	-9999.000	0	0	4:3	0.00	99.99	0.00	RSE1	514.0	99.99	2316.79	4415.32	0
70	-9999.0000	0.0000	0.0000	0.0000	-9999.000	0	0	4:3	0.00	99.99	0.00	RSE1	2238.0	99.99	2316.79	4415.32	0
121	-9999.0000	0.0000	0.0000	0.0000	-9999.000	0	0	4:3	0.00	99.99	0.00	RSE1	3388.0	99.99	2316.79	4415.32	0
169	-9999.0000	0.0000	0.0000	0.0000	-9999.000	0	0	4:3	0.00	99.99	0.00	RSE1	2469.0	99.99	2316.57	4408.13	0
198	-9999.0000	0.0000	0.0000	0.0000	-9999.000	0	0	3:4	0.00	99.99	0.00	RSE1	228.0	99.99	2257.83	4092.14	0

Figure 4: The Generated BSM Messages using the TCA Tool

From the initial simulation tests conducted, it was found out that for a scenario of normal traffic conditions speed and travel time estimations are at an accuracy level of approximately 90% or above regardless of the level of market penetration. However, these estimations are heavily affected when there is an incident. This impact on the accuracy of the transportation measure estimations will be investigated with more simulation tests and quantified in the final paper. For example, if the most of the CVs are stuck behind the double parking car then, the performance measures that are calculated using the messages coming from these connected vehicles are found to be less accurate due to biased spatial distribution of the vehicles along the affected roadway section. This is an important problem in terms using CVs as real-time probes for making real-time decisions to re-route traffic to reduce congestion and warn incoming vehicles to reduce the likelihood of secondary accidents.

REFERENCES

[1] So, J. J., G. Dedes, B. B. Park, S. HosseinyAlamdary, and D. Grejner-Brzezinsk. Development and evaluation of an enhanced surrogate safety assessment framework. *Transportation Research Part C: Emerging Technologies*, Vol. 50, 2014, pp. 51-67.

[2] Jeong, E., C. Oh, G. Lee, and H. Cho. Safety impacts of intervehicle warning information systems for moving hazards in connected vehicle environments. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2424, 2014, pp. 11-19.

[3] Songchitruksa, P., and L. Zha. Advancing Safety Performance Monitoring at Signalized Intersections Through Use of Connected Vehicle Technology. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2432, 2014, pp. 99-109.

[4] Talebpour, A., H. Mahmassani, F. Mete, and S. Hamdar. Near-crash identification in a connected vehicle environment. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2424, 2014, pp. 20-28.

[5] Bekiaris-Liberis, N., C. Roncoli, and M. Papageorgiou. Highway Traffic State Estimation with Mixed Connected and Conventional Vehicles. *arXiv preprint arXiv:1504.06879*, 2015.

[6] Paikari, E., L. Kattan, S. Tahmasseby, and B. H. Far. Modeling and simulation of advisory speed and re-routing strategies in connected vehicles systems for crash risk and travel time reduction.In *Electrical and Computer Engineering (CCECE), 2013 26th Annual IEEE Canadian Conference on*, IEEE, 2013. pp. 1-4.

[7] Park, H., C. Bhamidipati, and B. Smith. Development and evaluation of enhanced intellidrive-enabled lane changing advisory algorithm to address freeway merge conflict. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2243, 2011, pp. 146-157.

[8] Jin, Q., G. Wu, K. Boriboonsomsin, and M. Barth. Improving traffic operations using real-time optimal lane selection with connected vehicle technology.In *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, IEEE, 2014. pp. 70-75.

[9] Kari, D., G. Wu, and M. J. Barth. Eco-friendly freight signal priority using connected vehicle technology: a multi-agent systems approach.In *Intelligent Vehicles Symposium Proceedings, 2014 IEEE*, IEEE, 2014. pp. 1187-1192.

[10] Reis, A. B., S. Sargento, F. Neves, and O. Tonguz. Deploying roadside units in sparse vehicular networks: what really works and what does not. *Vehicular Technology, IEEE Transactions on*, Vol. 63, No. 6, 2014, pp. 2794-2806.

[11] Su, P., J. Lee, and B. B. Park. Calibrating Communication Simulator for Connected Vehicle Applications. *Journal of Intelligent Transportation Systems*, 2014, pp. 1-11.

[12] Goodall, N. J., B. Park, and B. L. Smith. Microscopic estimation of arterial vehicle positions in a low-penetration-rate connected vehicle environment. *Journal of Transportation Engineering*, Vol. 140, No. 10, 2014, p. 04014047.

[13] Brockfeld, E., R. Kühne, and P. Wagner. Calibration and validation of microscopic traffic flow models. *Transportation Research Record: Journal of the Transportation Research Board*, No. 1876, 2004, pp. 62-70.

[14] Chu, L., H. X. Liu, J.-S. Oh, and W. Recker. A calibration procedure for microscopic traffic simulation.In *Intelligent Transportation Systems, 2003. Proceedings. 2003 IEEE, No. 2*, IEEE, 2003. pp. 1574-1579.

[15] Ozbay, K., H. Yang, E. F. Morgul, S. Mudigonda, and B. Bartin. Big data and calibration and validation of traffic simulation models. *Traffic and Transportation Simulation: Looking Back and Looking Ahead: Celebrating*, Vol. 50, 2014, pp. 92-122.

[16] Krajzewicz, D., G. Hertkorn, C. Rössel, and P. Wagner. SUMO (Simulation of Urban MObility)-an open-source traffic simulation.In *Proceedings of the 4th Middle East Symposium on Simulation and Modelling (MESM20002)*, 2002. pp. 183-187.

[17] Vasudevan, M., J. O'Hara, S. Mudigonda, K. Ozbay, and A. Kurkcu. Estimating Key Transportation Measures Using Connected Vehicle Messages.In, Work Performed for U.S. DOT, 2016.
[18] USDOT. *TCA 2.3.3*. OSADP https://www.itsforge.net/index.php/community/explore-applications#/38/67. Accessed January 4, 2016.
[19] Morgul, E., H. Yang, A. Kurkcu, K. Ozbay, B. Bartin, C. Kamga, and R. Salloum. Virtual Sensors: Web-Based Real-Time Data Collection Methodology for Transportation Operation Performance Analysis. *Transportation Research Record: Journal of the Transportation Research Board*, No. 2442, 2014, pp. 106-116.
[20] Shlayan, N., A. Kurkcu, and K. Ozbay. Exploring pedestrian Bluetooth and WiFi detection at public transportation terminals.In *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, IEEE, 2016. pp. 229-234.
[21] *Trafficast BLUETOAD*. http://www.trafficcast.com/bluetoad.html. Accessed January 1, 2017.
[22] *Quadstone Paramics*. http://www.paramics-online.com/. Accessed January 1, 2017.
[23] *PTV Vissim*. http://vision-traffic.ptvgroup.com/en-us/products/ptv-vissim/. Accessed January 1, 2017.
[24] J2735, S. o. A. E. S. s. Dedicated Short Range Communications (DSRC) Message Set Dictionary.In, November 2009.
[25] *NYCStreets Permit Management System*. https://nycstreets.net/public/permit/search. Accessed January 1, 2016.

An Efficient Sampling Method for Stochastic Simulation-based Transportation Optimization

Timothy Tay, Carolina Osorio

Department of Civil and Environmental Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139 [tayt@mit.edu, osorioc@mit.edu]

December 9, 2016

1 Introduction

Stochastic microscopic simulators are popular tools for estimating the performance of predetermined transportation strategies, such as traffic signal control in cities. In many transportation problems, the performance measures are highly nonlinear, with no exact analytical expressions available. This is due to the numerous interactions between the elements in the problem. As such, simulation becomes a suitable way to estimate the performance of a given strategy.

Furthermore, microscopic simulators track individual vehicles and travelers in the network, with the decisions of each traveler being determined by the many disaggregate behavioral models embedded in the simulator. For instance, car-following models are used to represent how a driver behaves in response to the car in front of him, route choice models determine which route a traveler uses to get from the origin to destination, as well as models to describe how a traveler responds to current traffic conditions. In addition, since the microscopic simulator tracks every vehicle, it is also possible to obtain estimates of vehicle-related performance measures, such as fuel consumption (Osorio and Nanduri, 2015a) and emissions (Osorio and Nanduri, 2015b). Coupled with the network supply information, the microscopic simulators are able to provide detailed representation of the outcome on the roads. Through the complex interactions between travelers' decisions and infrastructure, the simulators can recreate complex emergent phenomena on the macroscopic scale, such as congestion propagation.

However, the intricate level of detail brings with it its own challenges. Due to the large number of interactions between elements in the simulation, there is a high computational cost associated with evaluating the simulator. This becomes a major problem for simulation-based optimization (SO) with a limited computational budget, especially in high-dimensional problems.

In this paper, we consider SO problems with continuous, nonlinear and general (eg. non-convex) objective functions, where the analytical form is unknown, and constraints with analytical and differentiable expressions. There are several model-independent experimental designs (for reviews, see Chen et al., 2006 and Kleijnen, 2015), which can be used with various metamodels. However, it is also possible to make use of available information on the model to design an improved sampling strategy.

Our past work in the area of SO has focused on the formulation of metamodels, which are analytical approximations to the SO objective function. Our focus has been on the formulation of metamodels that combine information from analytical network models and from the simulators. In this paper, we propose to use information from analytical network models to design an efficient sampling strategy. Such a

strategy can be used with any SO algorithm. The sampling strategy uses the analytical approximation of the SO objective function as a sampling distribution. The idea is to increase the probability of sampling areas in the feasible region, where the analytical network model predicts good performance (i.e., good objective function values). This design can be incorporated into a metamodel approach that has shown promising results for large-scale simulation-based transportation optimization (Osorio and Chong, 2015).

In the following section, we describe the problem and proposed sampling method in further detail, and present an algorithm framework. In section 3, we discuss the empirical set up used to validate the proposed approach.

2 Methodology

2.1 Problem Formulation

The objective function for the transportation SO problems can be written as

$$\min_{x \in \Omega} f(x, z; p) \equiv E[F(x, z; p)] \tag{1}$$

where F denotes the stochastic performance measure obtained through simulation, x represents the deterministic decision variable, z denotes the endogenous variables and p denotes the deterministic exogenous parameters. In this study, we consider a traffic signal control problem, with the average trip travel time as the performance measure F and the green times of the traffic signal phases as the decision variable x. Then, z would account for route choice decisions, departure times etc.; p would account for external traffic demand, vehicle attributes, traffic network topology etc.

Since the objective function f has no closed form, we use a metamodel proposed by Osorio and Bierlaire (2013) to approximate it. The functional form is given by

$$m(x, y; \alpha, \beta, q) = \alpha T(x, y; q) + \phi(x; \beta) \tag{2}$$

where T is the physical component described by an analytical queueing network model, ϕ is the functional component, often represented by a quadratic polynomial in x with diagonal second-derivative matrix, y are endogenous macroscopic variables, q are exogenous macroscopic parameters and α and β are parameters of the metamodel. The functional component ϕ can be seen as a local correction term for the approximation of the objective function given by T.

2.2 Queueing Network Model

The queueing network model used in the physical component of the metamodel combines ideas from both urban traffic theory and queueing theory, such that each lane in the network is represented by a finite capacity $M/M/1/k$ queue. Since the model contains the topology of the entire traffic network, it is thus able to function as an analytical global approximation of the objective function. The detailed formulation of the model can be found in Osorio and Bierlaire (2009), while a more tractable version for large-scale networks can be seen in Osorio and Chong (2015).

The notation to be used in the queueing model is described below, with the index i referring to a given queue:

γ_i	external arrival rate
λ_i^{eff}	effective arrival rate
μ_i	service rate
μ_i^{eff}	effective service rate (accounts for both service and eventual blocking)
ρ_i	traffic intensity
ρ_i^{eff}	effective traffic intensity
k_i	upper bound of the queue length
N_i	total number of vehicles in queue i
$P(N_i = k_i)$	probability of queue i being full, also known as the blocking or spillback probability
p_{ij}	transition probability from queue i to queue j
$\mathcal{D}_i$	set of downstream queues of queue i

The formulation of the highly tractable queueing network model (Osorio and Chong, 2015) is shown below:

$$\begin{cases} \lambda_i^{eff} = \gamma_i\left(1 - P(N_i = k_i)\right) + \sum_j p_{ij}\lambda_j^{eff} & \text{(3a)} \\ \rho_i^{eff} = \frac{\lambda_i^{eff}}{\mu_i} + \left(\sum_{j\in\mathcal{D}_i} p_{ij}P\left(N_j = k_j\right)\right)\left(\sum_{j\in\mathcal{D}_i}\rho_j^{eff}\right) & \text{(3b)} \\ P(N_i = k_i) = \frac{1-\rho_i^{eff}}{1-\left(\rho_i^{eff}\right)^{k_i+1}}\left(\rho_i^{eff}\right)^{k_i} & \text{(3c)} \end{cases}$$

Equation (3a) is a flow conservation equation relating the flow transmission between upstream and downstream queues. Equation (3b) defines the effective traffic intensity while accounting for spillback resulting from the downstream queues. Equation (3c) provides the expression for the spillback probability in terms of ρ_i^{eff}. We refer the reader to Osorio and Chong (2015) more details on the queueing network model.

The analytical approximation of the average travel time, denoted by $T(x, y; q, \theta)$, can then be obtained using Little's Law (Little, 2011) as follows:

$$T(x, y; q, \theta) = \frac{\sum_i \theta_i E[N_i]}{\sum_i \gamma_i(1-P(N_i=k_i))} \quad (4)$$

where the numerator represents the number of vehicles in the network and the denominator represents the average arrival rate to the network, while accounting for blocking. θ_i are link-specific parameters for fitting the approximation, which can be thought of as weights indicating the importance of each link in determining $T(x, y; q, \theta)$. The expected number of vehicles in queue i can be expressed in terms of the traffic intensity (Osorio and Chong, 2015):

$$E[N_i] = \rho_i^{eff}\left(\frac{1}{1-\rho_i^{eff}} - \frac{(k_i+1)\left(\rho_i^{eff}\right)^{k_i}}{1-\left(\rho_i^{eff}\right)^{k_i+1}}\right) \quad (5)$$

2.3 Improved Sampling Method

The analytical expression for $T(x, y; q, \theta)$ can serve as a means to weight the sampling distribution, so as to increase the probability of sampling in regions with values of small $f(x, z; p)$. The proposed method can be embedded in most SO algorithms.

At any given iteration k, when the SO algorithm attempts to pick a new point x for sampling, the sampling method conducts the following two steps:

Step 1. Sample a new point x that has a small $f(x, z; p)$ value using a distribution based on $T(x, y; q, \theta)$.

Step 2. Improve accuracy of $T(x, y; q, \theta)$ by fitting θ_i using simulation observations from the last $k - 1$ iterations.

In step 1, we use an inverse transform sampling method to obtain a sampling point x. Given a uniform random vector $U \in \mathbb{R}^d$ such that $\forall i, U_i \in [0,1]$, the point to be sampled is

$$x = F_T^{-1}(U) \tag{6}$$

where $F_T(x, y; q)$ is the normalized cumulative sum of $T(x, y; q, \theta)$, and d is the dimension of the decision variable x. By using the inverse transform sampling method, this effectively weighs the sampling distribution more heavily in regions where $T(x, y; q, \theta)$ contains minima. Hence, if $T(x, y; q, \theta)$ provides a good approximation of $f(x, z; p)$, then the computational budget can be used more efficiently to potentially find lower minima at a faster pace.

Since this method is highly contingent on $T(x, y; q, \theta)$ being a good approximation of $f(x, z; p)$, step 2 attempts to parametrically correct the approximation using the data collected from simulation observations from the current and past iterations. This can be done by regression, where the parameters θ_i are fitted at every iteration by solving a least squares problem, similar to that in Osorio and Bierlaire (2013). θ_i are assigned to take initial value one, which leads to the initial approximation of the average travel time taking on the original form given by Little's Law.

3 Empirical Analysis

We evaluate the performance of the proposed sampling method by combining it with several SO algorithms as an offline sampling technique, and observing how the speed of convergence is affected when compared to a uniform sampling technique. The chosen algorithms include popular general SO algorithms, such as genetic algorithms (GA) and the Simultaneous Perturbation Stochastic Approximation (SPSA). We also combine it with the metamodel method (Osorio and Bierlaire, 2013) as an iterative sampling technique, where it will replace the model improvement step.

To compare the performances of the SO algorithms, we use a microscopic traffic simulation model of the city of Lausanne, Switzerland developed by Dumont and Bert (2006). This model of Lausanne contains 17 chosen signalized intersections where the signals will be controlled, leading to a total of 99 phase variables (ie. 99 dimensions in the decision vector x). This is considered a large-scale problem in the area of simulation optimization. At the same time, we also consider the performance of the sampling method under a tight computational budget, such that only a maximum of 150 simulation runs can be carried out.

As a second case study, we consider a problem for the city of Lyon, France. The model is a mesoscopic simulation instead, containing 266 signalized intersections to be controlled, leading to a total of 800 signal phases (Bert, 2016). Hence, this problem has a significantly larger number of dimensions, where we expect uniform sampling to be particularly inefficient.

References

Bert, E., 2016. Outil de prédiction et d'aide à la décision pour la gestion dynamique du trafic en temps réel pour la Ville de Lyon : Le projet Opticities de la métropole de Lyon. La revue Routes & Transports de l'AQTr 46–49.

Chen, V.C.P., Tsui, K.-L., Barton, R.R., Meckesheimer, M., 2006. A review on design, modeling and applications of computer experiments. IIE Transactions 38. doi:10.1080/07408170500232495

Dumont, A.G., Bert, E., 2006. Simulation de l'agglomération Lausannoise SIMLO. Technical report, Laboratoire des voies de circulation, ENAC, Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland.

Kleijnen, J.P.C., 2015. Design and Analysis of Simulation Experiments, International Series in Operations Research & Management Science. Springer International Publishing, Cham. doi:10.1007/978-3-319-18087-8

Little, J.D.C., 2011. Little's Law as Viewed on Its 50th Anniversary. Operations Research 59, 536–549. doi:10.1287/opre.1110.0940

Osorio, C., Bierlaire, M., 2013. A Simulation-Based Optimization Framework for Urban Transportation Problems. Operations Research 61, 1333–1345. doi:10.1287/opre.2013.1226

Osorio, C., Bierlaire, M., 2009. An analytic finite capacity queueing network model capturing the propagation of congestion and blocking. European Journal of Operational Research 196, 996–1007. doi:10.1016/j.ejor.2008.04.035

Osorio, C., Chong, L., 2015. A Computationally Efficient Simulation-Based Optimization Algorithm for Large-Scale Urban Transportation Problems. Transportation Science 49, 623–636. doi:10.1287/trsc.2014.0550

Osorio, C., Nanduri, K., 2015a. Energy-Efficient Urban Traffic Management: A Microscopic Simulation-Based Approach. Transportation Science 49, 637–651. doi:10.1287/trsc.2014.0554

Osorio, C., Nanduri, K., 2015b. Urban transportation emissions mitigation: Coupling high-resolution vehicular emissions and traffic models for traffic signal optimization. Transportation Research Part B: Methodological 81, 520–538. doi:10.1016/j.trb.2014.12.007

Extended abstract prepared for TSL Conference 2017

A Within-Day Microscopic Dynamical Model of Route Choice and Responsive Traffic Signal Control

Ronghui Liu[1*] and Mike Smith[2]

[1] Institute for Transport Studies, University of Leeds, UK

[2] Department of Mathematics, University of York, UK

ABSTRACT

In this paper, we explore the dynamical system of route choice and responsive signal control in a simulated real-life environment, where (a) the route costs are observed directly from a microsimulation model, and (b) drivers' pre-trip route choices respond to current traffic conditions and signal green-times respond to current traffic flow. Queues are explicitly modelled, as with queue spill back. We show that the dynamical systems, if well designed, can significantly improve network capacity and reduce vehicle delays, as well as lead to desirable route choices.

Keywords: Dynamics, Routing, Signal control, Within-day

1. BACKGROUND

Aside from massive road building or network re-structuring, traffic signal control, along with pricing control, is one of the few effective tools the city authorities have to manage traffic congestion and conflicts. Many traffic signal control systems have been developed and deployed around the world; see a review by Shepherd (1994) for these systems.

Most of these control systems, based on *minimising delays* for the *currently observed* traffic patterns, assume that travellers' route-choices do not change. It was first pointed out by Allsop (1974) and Gartner (1976) that traffic signals may be used to influence traffic patterns and manage travel demand. Fig. 1 shows such interactions.

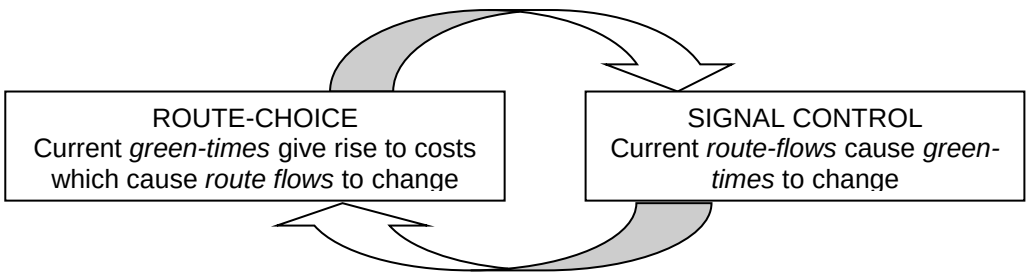


Fig 1. A dynamical system of route choice and signal control. Here, drivers' current route-choices change current green-times (according to some responsive control policy) and current green-times change current delays and hence current flows (according to drivers' route-choices). The loop is traversed indefinitely.

* Corresponding author: Tel: +44 113 343 5338, Fax: +44 113 343 5334, Email: r.liu@its.leeds.ac.uk.

The potential for signals to influence route choice becomes much more important *when there is high congestion*; in this case using signal control to deliberately discourage the most congestion-causing route-flows and to deliberately encourage congestion-reducing route or mode switches may yield substantial benefits.

Most of the previous studies in traffic routeing and signal control combined are based on macroscopic traffic flow models (e.g. Dickson, 1981; Smith, 1979, 1987; Smith et al., 2015). Macroscopic traffic models are based on an aggregate notion of traffic flows and travel times on links; they cannot represent queue spill back, which is a common feature in congested networks.

In addition, these theoretical traffic models rely on *a priori* knowledge of a link- or route-based flow-cost function. Liu and Smith (2015) show that the sensitivity of future network behaviour to the cost functions and control policies utilised. In practice, many factors affect the cost functions on a link or along a route. Liu et al (2011) demonstrate the complexity in estimating route cost functions, and show that these are affected not only by the flows on that route, but also by travel patterns over the wider network.

Furthermore, modern traffic signals encompass complex signal phases and stages[†], where a phase may contain a set of turning movements from different links and a stage may include several phases. Real-life road network includes lanes with shared traffic turning movements. Macroscopic models based on average link flows are not able to represent shared lane traffic to distinguish flows and delays by turning movements, and is thus not able to truly represent the dynamical system of Fig. 1.

All these call for dynamic disaggregate traffic modelling to account for traffic queuing and blocking back, realistic representation of road networks (e.g. for shared lane traffic) and signal plans, as well as drivers short- and long-term route choice responses to signal control. Traffic microsimulation models, based on the explicit modelling of individual vehicle trajectories and physical queues, offer the potential to overcome the difficulties mentioned above.

In this paper, we present a *large* microsimulation model to test some of the newly developed dynamical system of Fig. 1, in a *within-day dynamic* context. Queues are explicitly modelled, as with queue spill backs and shared-lane traffic. The model seeks to represent the minute to minute decisions of both travellers and signal setting "agents". Such an agent may represent an engineer who periodically resets signals on the basis of data collected over a long or short time-scale or may represent an automatic quickly responding traffic control system. In this model, trip-makers respond to the prevailing traffic conditions in the network and adjust their route choice accordingly before departure.

---

† A signal phase is defined as the sequence of signal indications given to a particular traffic movement or a set of movements. A stage is a portion of cycle during which a given combination of movements is given green.

2. A MICROSIMULATION MODEL OF THE DYNAMICAL SYSTEMS

We develop a within-day microsimulation model of the dynamical system of Fig. 1, where (a) the route costs are observed directly from a microsimulation of individual vehicle trajectories, and (b) route flows and green-times are adjusted over time (both within the micro-simulation model). Queues are explicitly modelled, as with queue spill back. The model seeks to mimic the minute to minute decisions of both travellers and traffic signal operators. We explore the dynamical interactions between route-choice and responsive equi-saturation and responsive P0 signal control policies in this simulated real-life environment.

The microsimulation platform utilized here is the DRACULA dynamic microsimulation model suite, developed at University of Leeds (Liu et al, 2006). At its most detailed level, DRACULA represents fully the individual O-D trip-makers, the individuals' decisions on whether to travel on a day, their day-to-day route choice responses to network congestion, and individual vehicles' space-time trajectories as they traverse a network (Liu, 2010).

In this study, we utilize DRACULA as a *within-day* dynamic traffic assignment model to represent the dynamical responsive signal controls and route choices within a typical peak period, with the route congestion experienced by earlier travellers being fed back to later travellers to influence their route choice. The dynamical loop of Fig. 1 is updated every few cycles of the signal timings (typically a few minutes), where the signal stage times are adjusted according to the prevailing traffic congestion on the links approaching the signal, and the route flows (and the individual travellers' route choice probabilities) are adjusted according to the prevailing congestions en-route.

From the simulation, we collect at each time stamp t: (i) for each individual trip its location (on link and lane) and speed; (ii) for each lane and link the number of vehicles, speeds, and queue length; (iii) for each individual signal stage, the signal aspect (e.g. green, red, amber), the number of vehicles exited, and their travel time over the previous link.

From these basic simulation outputs, we then define, at the end of each dynamical system response time period $\eta\tau, \eta = 1,2, \ldots$ registers the number of response iterations, the following aggregated measures:

- $x_{am}(t)$ turning flow, defined as the number of vehicles from link a making the turning movement m, during period $(t-\tau, t]$;
- $c_{am}(t)$ turning flow cost, defined as the average travel time of the vehicles $x_{am}(t)$ on link a.
- $X_{ijr}(t)$ route exit flow, as the number of vehicles from OD pair ij, route r which exited the network during time period $(t-\tau, t]$; and
- $C_{ijr}(t)$ route flow cost, as the average travel times of the vehicles $X_{ijr}(t)$.

From the above, we calculate the flows and costs for each stage at time $t = \eta\tau$:

$$x_{nk}(t) = \sum_{\{a,m\}\sim\{n,k\}} x_{am}(t) \text{ and } \qquad c_{nk}(t) = \sum_{\{a,m\}\sim\{n,k\}} c_{am}(t) \tag{1}$$

The above Equations represent the flows and costs felt by those vehicles passed through the junction n during stage k. By default, if no turning flows are made through this stage during the time period, then a zero cost would be registered and no changes will then be made to the stage timings for this response time period.

Taking the above simulated flows and costs by stage, we define for each stage a red-time cost, for time period $(t-\tau, t]$ equi-saturation policy (ES) and the P0 policy (P0):

$$RC_{nk}^{ES}(t) = x_{nk}(t) / [s_{nk} \times g_{nk}(t)] \tag{2}$$

$$RC_{nk}^{P_0}(t) = s_{nk} \times c_{nk}(t) \tag{3}$$

where s_{nk} denotes the saturation flow for stage k at signalised node n and is defined as the collection of saturation-flows of all turning movements that belong to that statge as:

$$s_{nk} = \sum_{\{a,m\}\sim\{n,k\}} s_{am} \tag{4}$$

where $\{a,m\}\sim\{n,k\}$ denotes that the turning movement m from link a belongs to signal n and stage k.

Then the new stage times and new route flows for the next time period are calculated, according to the two-commodity dynamical system proposed in Liu and Smith (2015). For vehicles which are due to depart during the next response time period $(t, t+\tau]$, their route choice will be determined probabilistically in proportion to the new route flows for that OD pair.

3. SIMULATION EXPERIMENTS AND ANALISIS

The simulation experiments were conducted on two networks shown in Fig. 2. Three systems are simulated:

S0: base case scenario with fixed route flows and fixed green times;

S1: with dynamical route swaps and responsive equi-satuation policy; and

S2: with dynamical route swaps and responsive P0 policy.

We examine the impact of the dynamical systems (S1 and S2) on network performance and compare them with those of the base system S0.

The first network consists of a single OD pairs (from origina A to destination B) and three possible routes: Routes 1 and 2 via node C along links 1-3 and 2-3 respectively, and a direct route Route 3 from A to B. Nodes B and C are signal controlled. The three routes have different capacities. Link 1 is a single carriageway, links 2 and 3 are dual carriageways with two lanes each, whilst link 4 has four lanes. Thus Route 3 has the highest capacity, followed by Route 2, while Route 1 has the least capacity. All three routes have the same free-flow travel times.

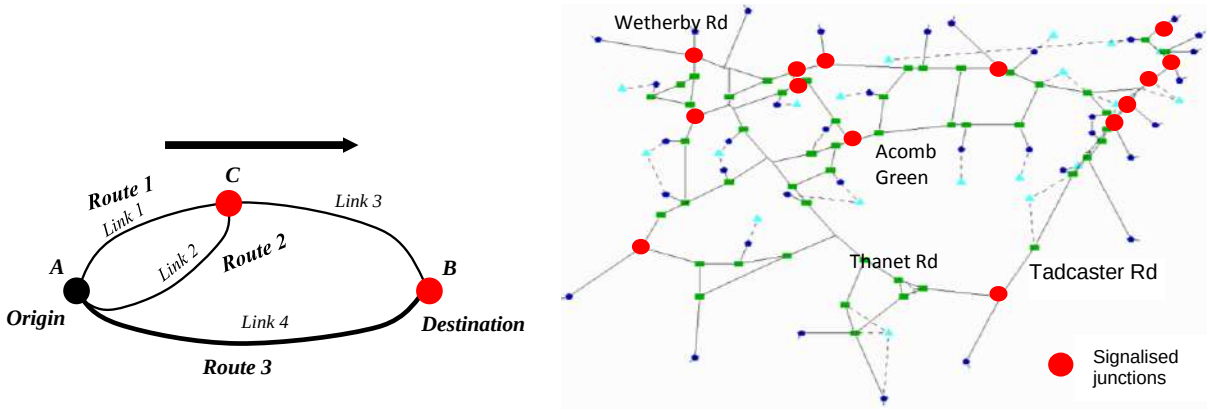


Fig. 2. Two test networks: (a) a simple schematic three-route network; and (b) a real-life network in the City of York. The signalised intersections in both networks are highlighted in red filled circles.

Fig. 3 shows the average trip journey time, under each of the three systems, as the total demand for travel increases. It can be seen that systems S1 and S2 with dynamical re-route and responsive equi-saturation and the P0 signal controls improve overall network capacity and reduce vehicle delays, especially for higher demand.

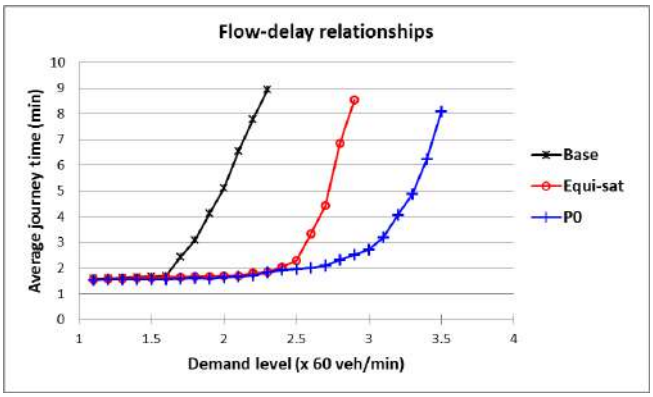


Fig. 3. Average trip journey time as a function of demand.

The second network (Fig. 2b) is based on the morning peak conditions in the western part of the City of York, England. The city centre is at the top-right hand corner of the network, reached by two major radial routes via the Tadaster Rd from the south and Wetherby Rd from the west. Thanet Rd connects the two routes through Acomb Green. The total demand for this network is 7,130 veh/hr in the morning peak, from a total of 472 OD pairs.

A 4-hour simulation is conducted for the York network, with a 4-min response period. The initial route assignment is from equlibrium assignment, which is also used as the base case. The equilibrium assignment results in 79 OD pairs having more than one route assigned. We assume that the initial route sets contain all the feasible ones; and that the dynamical re-routing is performed only among these existing route sets. Fig. 4 present the simulated results, in terms of vehicle delay times (VDT) and total trip distance travelled (D).

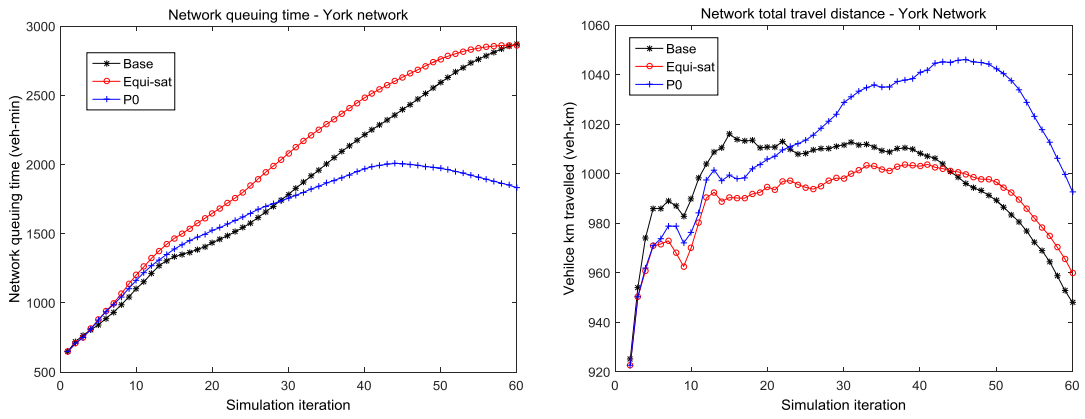


Fig. 4. The impact of dynamical systems with responsive equi-saturation and P0 signal control on: (a) vehicle delay time, and (b) vehicle km travelled for the York network, compared with the base case.

It can be seen that system S1 with equi-saturation policy has no significant impact on VDT, whilst system S2 with the P0 policy has a significantly lower VDT throughout. It is worth noting that the base network has been calibrated to the real-world traffic condition; that includes the set of optimal (fixed-plan) signals in the base case scenraio. It is likely that the signals have been designed based on the traditional delay minimisation principle. This could explain the relative small impact of the dynamical system S1 on VDT seen in Fig. 4a.

Fig. 4b displays the total trip distances completed during each iteration; and it shows that system with the P0 control has a significantly increased the trip distances. Combined with the results in Fig. 4a, this suggests that the P0 control is effective in swapping traffic from more congested shorter rotues, to less congested longer routes.

Further research will include more extensive examination of combination of signal timing and routing advice changes, and in situations when average conditions prevail as well as periods of time when there are incidents.

REFERENCES

Allsop, R. E., 1974. Some possibilities for using traffic control to influence trip distribution and route choice. 6th International Symposium on Transportation and Traffic Theory (Ed: Buckley), 345-374.

Gartner, N. H. 1976. Area Traffic Control and Network Equilibrium, Traffic Equilibrium Methods, Lecture Notes in Economics and Mathematical Systems., 118, (Editor: Florian, M.) Springer Verlag, 274 – 297.

Dickson, T. J., 1981. A note on traffic assignment and signal timings in a signal-controlled road network. Transportation Research Part B 15 (4), 267 – 271.

Liu, R. (2010) *Traffic Simulation with DRACULA*. In Fundamentals of Traffic Simulation (Ed. J Barcelo), Chapter 8. Springer. 295-322.

Liu, R, Van Vliet, D. and Watling, D. (2006) Microsimulation models incorporating both demand and supply dynamics. Transportation Research Part A, 40, 125-150.

Liu, R., May, AD and Shepherd, S (2011) On the fundamental diagram and supply cost of congested urban networks. Transportation Research Part A, 45(9), 951-965.

Shepherd, S. (1994) Traffic control in over-saturation conditions. Transport Reviews, 14(1), 13-43.

Smith, M. J. (1979). A local traffic control policy which automatically maximises the overall travel capacity of an urban road network. Proc. International Symposium on Traffic Control Systems, Volume 2A, 11-32.

Smith, M. J. (1987). Traffic control and traffic assignment in a signal-controlled network with queueing. Proc. 10th International Symposium on Transportation and Traffic Theory (Eds: Gartner and Wilson, 61 - 68.

Traffic & Mobility

SB2: Traffic Data Analysis

Saturday 11:15 – 12:45 PM

Session Chair: Serge Hoogendoorn

11:15 A Data-Driven and Integrated Evaluation of Area-Wide Impacts of Double Parking Using Macroscopic and Microscopic Models

[1]Jingqin Gao, Kaan Ozbay*, [2]Michael Marsico
[1]New York University, [2]New York City Department of Transportation

11:45 Urban Trajectory Analytics: Day-Of-Week Movement Pattern Mining Using Tensor Factorization

Jiwon Kim*, Kianoosh Soltani Naveh
The University of Queensland

12:15 Opportunities for Floating Car Data in Integrated Traffic Management: The Case of Queue Estimation

[1]Serge Hoogendoorn*, [2]Erik-Sander Smits, [2]Jaap Van Kooten
[1]Delft University of Technology, [2]Arane

INFORMS Transportation and Logistics Society, First Triennial Conference

July 26 – 29, 2017, Chicago, Illinois, USA

A Data-Driven and Integrated Evaluation of Area-wide Impacts of Double Parking Using Macroscopic and Microscopic Models

Jingqin Gao[a], Kaan Ozbay[b], Michael Marsico[c]

[a] Department of Civil and Urban Engineering, New York University
[b] Department of Civil and Urban Engineering &Center for Urban Science and Progress, New York University
[c] New York City Department of Transportation

Keywords: Double parking, Data-driven, Microscopic models, Macroscopic models, Traffic impact

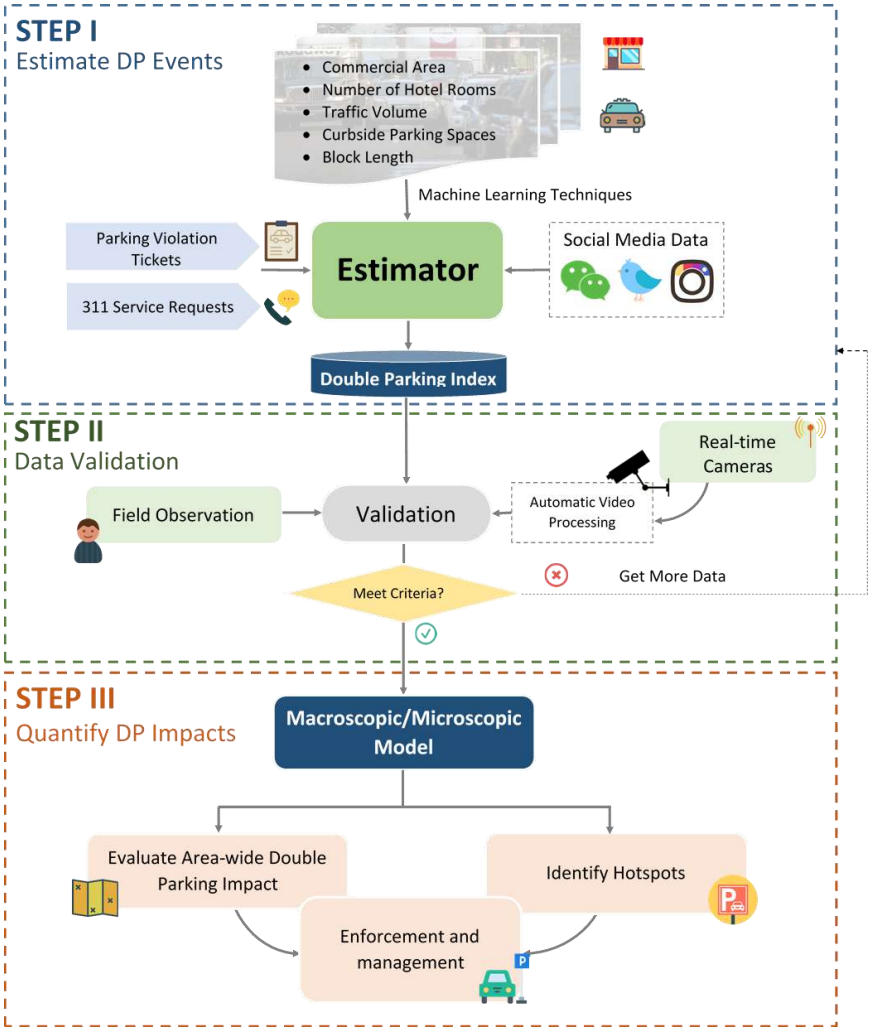
1. Introduction and Background

Double parking (DP) that often negatively affects traffic operations and safety is not a new phenomenon on urban streets. Evaluation of double parking impacts can be extremely challenging due to the difficulties in collecting data and modeling double parking behavior in a large-scale network. In New York City, clearly, there is no existing dataset that can provide an exact number of double parking activities. However, several other vast amounts of related temporal and spatial data are available citywide. In the light of all these complications, this study presents an enhanced data-driven framework that was originally proposed by the authors in a previous article *(1)* to integrate various data sources, such as parking violation tickets, 311 service requests and Twitter, to estimate the severity level of double parking in terms of its frequency and duration (Figure 1). The estimator uses Random Forests techniques and is proved to have 85% out-of-sample accuracy in estimating a combined index of double parking frequency and duration. The outputs from the estimator can then be feed into macroscopic or microscopic models as inputs with the goal of quantifying their impact on traffic.

The estimated double parking activities along with existing large scale models, such as Manhattan Traffic Model (MTM) that covers more than 800 intersections, creates opportunities to evaluate area-wide double parking impacts without other compounding factors. This allows local transportation agencies to identify not only the double-parking activity hotspots, but also the locations where transportation agencies can obtain the greatest benefits from eliminating double parking. Different parking enforcement and management strategies can be tested in the model to achieve more sustainable mobility in the study area.

On the other hand, macroscopic models should also be considered as an alternative approach to quantify double parking impacts in terms of lower complexity and computational speed compared with microscopic simulation models.

This paper first estimated the frequency and duration of double parking events using a well-known machine learning technique namely, Random Forests. Then area-wide double parking events are simulated in a large-scale microsimulation model to quantify their traffic impact. Next, we study the feasibility of using macroscopic queueing models that incorporate interruptions due to double parking instead of complex and computationally very demanding microsimulation models specially for large networks. Finally, the outputs from the both simulation and queueing models are used to identify double parking hotspots that need better enforcement or management.



*Double Parking Index (DPI) is an indicator that represents the combined severity level of DP frequency and duration.

Figure 1 Data-driven framework to quantify the impact of double parking

2. Data Needs for Estimating Network-wide Double Parking Activities

Intuitively, double parking violation tickets and 311 service requests complaining about double parking blocking the traffic are directly related to the actual double parking occurrence, yet both of the datasets have their own selection and sampling biases. Supplemental data such as street characteristics (i.e. commercial area) and traffic information (i.e. traffic volume) are needed to address this issue. Table 1 shows the summary of datasets used in this study.

Table 1 Description of Available Data

Data Name	Length	Source	Description
Parking Violations	1 year	NYC OpenData*	Double parking related violation tickets for 2015 in NYC
311 Service Requests	6 years	NYC OpenData*	Double parking related requests from 2010 to 2015 in NYC
Twitter	3 years	Twitter API	Geo-located tweets from 2013 to 2015 in Manhattan
Double parking Information	1-3 days	Field Visit	Peak hour double parking occurrence and duration from field visits and real-time cameras.
Commercial Area	-	NYC OpenData*	Total commercial area for each street in study area
Hotel Rooms	-	Hotel Websites	Total number of hotel rooms for each street in study are

Traffic Volume	Varied	Field Visit	Average peak hour traffic volume for each street in study area
Curbside Parking Space	-	Field Visit	Number of curbside parking spaces during peak hour in study area
Block Length	-	NYC OpenData*	Block length for each street in study area

*NYC OpenData *(2)* is the open data portal of New York City that offers access to government-produced data sets.

Besides traditional types of traffic datasets, the increasing use of crowdsourcing social media applications, such as Twitter *(3)*, has allowed novel approaches for obtaining extra information about human activities. This rich text information can allow us to better understand how people react to double parking activities. For example, people complaining about a double-parked vehicle blocking bike lanes or an ambulance having to double park because the curbside is occupied by trucks are both common occurrences in the twitter data. Emotional keywords can also provide indications of double parking severity and recurrent occurrence at certain locations.

3. Estimating Impacts of Double Parking

Microscopic Model

Manhattan Traffic Model *(4)* is a combined macroscopic/mesoscopic/microscopic simulation model of Manhattan with additional major arterials in the outer boroughs and New Jersey (Figure 2). While the macroscopic model disaggregates information from the regional New York Best Practices Model (NYBPM) and the mesoscopic model uses dynamic user-equilibrium following strict convergence criteria, the microscopic MTM model is a stochastic model based on dynamic traffic assignment that covers the street network from 14th Street to 66th Street, river-to-river. It is the largest, most comprehensive microsimulation model at the tri-state area.

In this study, a subarea in midtown Manhattan that contains 24 intersections and 36 roads are used. Double parking during the AM peak hour (8AM to 9AM) is coded as periodic section event that creates random incidents and is placed randomly throughout the street. During the time interval specified, the simulation creates single-lane incidents on any lane and position of the specified section following the time patterns defined *(5)*. Both the frequency and duration of the event can be either fixed or follow a selected distribution (i.e. Normal Distribution). The double-parking behavior in the model is calibrated (i.e. look ahead distance, passing speed) using the observation from multiple videos. Estimated number of Double Parking activities from our estimator are applied where no actual counts are available.

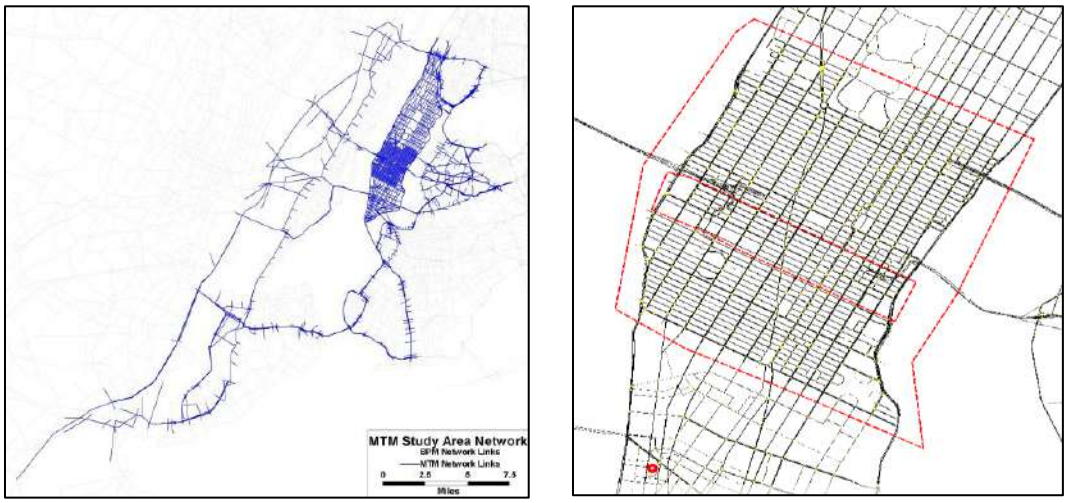


Figure 2 MTM model and its primary micro-model area *(4)*

Macroscopic Model

Although microscopic simulation models are effective for modeling this kind of extremely detailed behaviors, they are usually labor intensive to be developed and calibrated. As an alternative approach, macroscopic models can be magnitude of order inexpensive and faster to be developed and implemented. However, accuracy of such models need to be carefully validated. This study adopted Baykal-Gürsoy and Xiao's M/M/∞ model *(6-10)* with modification on vehicle type under double parking conditions. An M/M/∞ queueing model represents a system with a Markovian arrival rate, a Markov modulated service rates, and an infinite number of servers in the system *(7)*. The system in this study is designed with two server states "Failure (F)" or "Normal (N)" to illustrate the conditions with or without double parking. Denote E be the set of roadway links in the study area, *V={P,T}* to indicate the vehicle type, where *P* stands for passenger cars and *T* stands for commercial trucks. When the system experiences interruptions, the average number of vehicles on the roadway link i can be represented as follow (adopted from *(6)*):

$$N_i = \frac{D_i L_i}{v_i}\left[1 + \frac{\sum_j c_j f_{ij}\left(1 - \frac{v'_i}{v_i}\right)}{\frac{1}{d_i} + \sum_j c_j f_{ij}}\left(1 + \frac{\left(\sum_j c_j f_{ij} + \frac{v_i}{L_i}\right)(v_i - v'_i)}{\frac{v_i}{d_i} + \sum_j c_j f_{ij}\, v'_i + \frac{v_i - v'_i}{L_i}}\right)\right], \quad (\mathrm{i} \in \mathrm{E}, \mathrm{j} \in \mathrm{V}) \qquad (1)$$

Subsequently, the average link travel time is:

$$t_i = \frac{L_i}{v_i}\left[1 + \frac{\sum_j c_j f_{ij}\left(1 - \frac{v'_i}{v_i}\right)}{\frac{1}{d_i} + \sum_j c_j f_{ij}}\left(1 + \frac{\left(\sum_j c_j f_{ij} + \frac{v_i}{L_i}\right)(v_i - v'_i)}{\frac{v_i}{d_i} + \sum_j c_j f_{ij}\, v'_i + \frac{v_i - v'_i}{L_i}}\right)\right], \quad (i \in E, j \in V) \qquad (2)$$

Where *L_i=length of link i (mile); D_i=hourly traffic demand on link i (veh/h); f_{ij}=frequency of double parking for vehicle type j on link i (events/hour), d_i=average duration time of double parking events on link i (h); v=average speed without double parking on link i (mph); and v'_i= average speed with double parking on link i (mph).*

4. Preliminary Findings

The case study utilized a machine learning technique – Random Forests – to integrate different datasets from a variety of sources so that double parking events can be estimated at the street level. Instead of fitting the data into a model, machine learning techniques enable the computer to learn from the actual data. Random Forests, as one of the most popular and accurate machine learning methods that combine many decision tree predictors was used as a classifier to classify double parking events into a categorical response variable "Double Parking Index (DPI)" introduced by the authors in a previous study *(1)*. Figure 3 shows the estimation results for the study area.

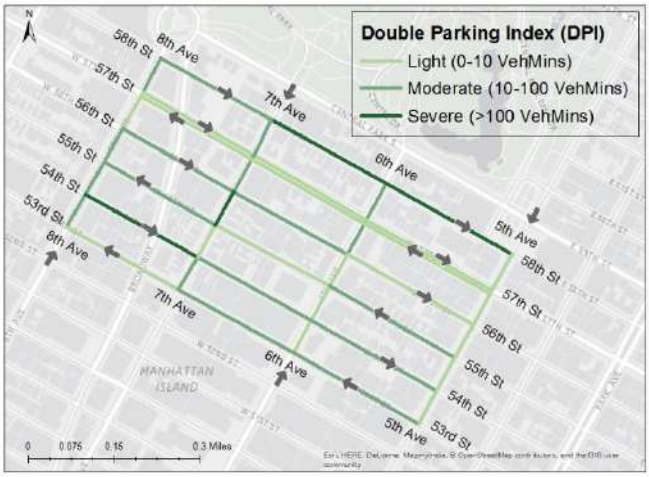


Figure 3 Double Parking Index (DPI) Estimation Results

The output generated by the random forests technique is then fed into a microsimulation model to evaluate area-wide double parking impacts. Figure 4 shows the preliminary results of the percentage difference in terms of average link travel time with or without double parking. Interestingly, although the 8th Avenue has a lower number of double parking events (3-11 events/link/hr) compared to the 6th Avenue (4-32 events/link/hr), it is more affected by double parking. One possible reason can be that majority of the double parking activities on the 6th Avenue are passenger cars or taxi drop-offs/pick-ups that have a relatively short duration less than one minute. However, the 8th Avenue has more commericial truck activities that have relatively longer durations, and all the traffic are heading to a roundabout (Columbus Circle) which can be usually a traffic bottleneck under saturated traffic conditions. Similar situations can be observed on the 5th Avenue and some crosstown streets.

The results suggest that priority should be given to the streets in dark red when considering parking enforcement as it may save more than 20% average travel time in the AM peak for these streets. Moreover, by eliminating all of the double parking activities in this subarea may reduce average travel time for each vehicle by 77.5 seconds during AM peak hour.

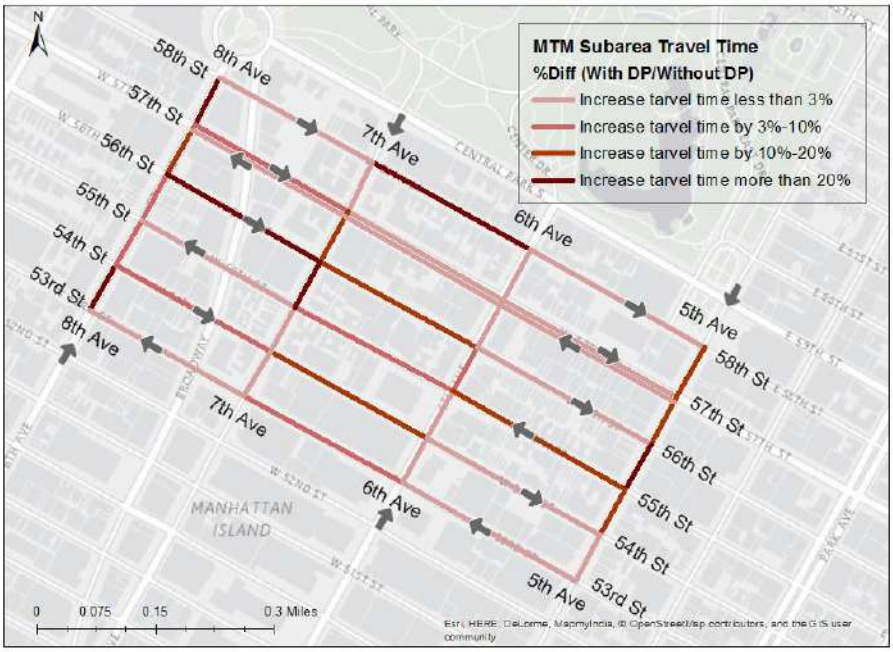


Figure 4 MTM Subarea Results

In our study, M/M/∞ queueing model is also found to be an effective approach and has the potential to be implemented instead of computationally demanding microsimulation models for large-scale network. Taking vehicle type into consideration enbles us to improve model performance. The full paper will cover more detailed computational results.

With our novel data-driven integrated framework for estimating the actual frequency of double parking, both microscopic or macroscopic models can be utilized to quantify area-wide impacts in the presence of double parking. The findings of this study can provide transportation agencies with useful insights on identifying locations that will experience the greatest benefits by removing problematic double parking. As a result, various parking enforcement and management strategies can be planned more effectively.

Reference

1. Gao, J. and K. Ozbay, A Data-driven Approach to Estimate Double Parking Events Using Machine Learning Techniques (forthcoming). In: Proceedings of the Transportation Research Board 96th Annual Meeting, 2017.
2. NYC OpenData, https://nycopendata.socrata.com/, Accessed Jan. 1st, 2017.
3. Twitter, https://about.twitter.com/company, Accessed July 18, 2016.
4. NYCDOT, Manhattan Traffic Model, https://www.aimsun.com/manhattan-traffic-model/, Accessed,
5. TSS, Aimsun 8 Dynamic Simulators Users' Manual. In *No.*, 2014.
6. Baykal-Gursoy, M., W. Xiao, Z. Duan, and K. Ozbay, Delay estimation for traffic flow interrupted by incidents. In: Proceedings of the 86th Annual Transportation Research Conf., Transportation Research Board, Washington, DC, 2006.
7. Baykal-Gursoy, M. and W. Xiao, Stochastic decomposition in M/M/∞ queues with Markov modulated service rates. *Queueing Systems 48(1-2)*, 2004, pp. 75-88.
8. Baykal-Gürsoy, M., W. Xiao, and K. Ozbay, Modeling traffic flow interrupted by incidents. *European Journal of Operational Research 195(1)*, 2009, pp. 127-138.
9. Ozbay, K.M., W. Xiao, G. Jaiswal, B. Bartin, P. Kachroo, and M. Baykal-Gursoy, Evaluation of incident management strategies and technologies using an integrated traffic/incident management simulation. *World Review of Intermodal Transportation Research 2(2-3)*, 2009, pp. 155-186.
10. Ozbay, K. and B. Bartin, Evaluation of Incident Management Systems using Simulation. *SIMULATION Journal 79(2)*, 2003, pp. 69-82.

Urban Trajectory Analytics: Day-of-week Movement Pattern Mining using Tensor Factorization

Jiwon Kim[*] *Kianoosh Soltani Naveh*[*]

1 Introduction

Urban traffic arises from population movement and interaction between different regions of a city. Understanding movement patterns of various users in a traffic network is a key to designing and operating effective transport systems. Movements of individual travellers can be captured by various sensors such as GPS devices, Bluetooth detectors or public transit smart card systems. Recently, movement trajectory data have been collected at massive scales in many cities, providing the opportunity to gain insight into network-wide traffic dynamics and understand how the urban network is used and what changes can be made to improve the overall performance of traffic in the system.

An important aspect of movement pattern mining is to discover regularity or periodicity in human mobility patterns. A particular research question addressed in this paper is whether there exist specific weekly patterns in network-wide movement data and whether we could learn and build knowledge of *typical* movement patterns for different days of the week from historical data. Such knowledge can be applied in predicting future movements and detecting anomalies, enabling better decision support systems for network planning, fleet management, and real-time traffic and incident management. To answer this question, this paper proposes a dynamic graph method that converts trajectory data for a given day into a time series of graph snapshots and represents aggregated flows between different regions using a compact graph structure. Using tensor decomposition on dynamic graphs, we obtain spatio-temporal patterns underlying population movement for any given day. By learning patterns from data that span over six months, we characterize day-of-week movement patterns and demonstrate that the resulting patterns can be distinguished by a classifier.

2 Methodology

Fig. 1 shows the overall process for day-of-week movement patterns mining proposed in this study. It involves three steps: (i) Dynamic Graph Representation, (ii) Feature Extraction and (iii) Dimensionality Reduction, and Characterization and Classification, each of which is discussed in more detail in the following subsections.

2.1 Dynamic Graph Representation

Determining how to represent movement data is the critical first step toward successful movement patterns mining. In order to get better insight into the mass movement patterns, we propose a graph-based approach that aggregates individual trajectories over space and time and represents

[*]University of Queensland

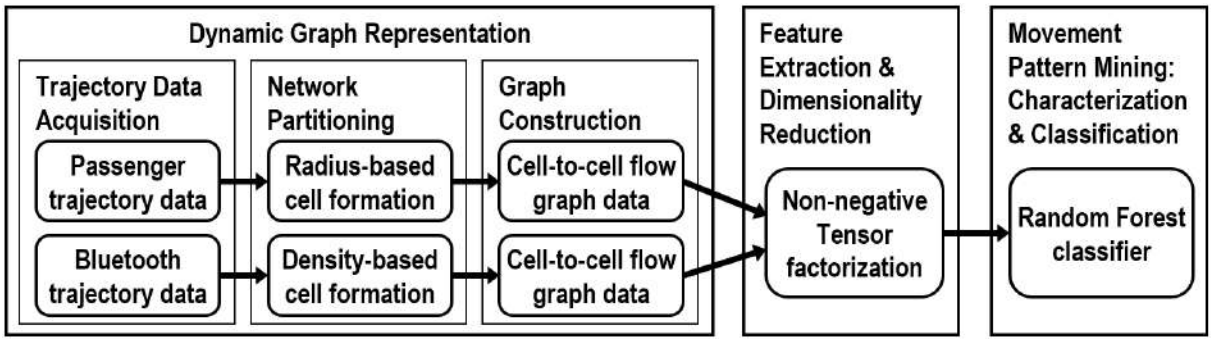


Fig. 1: An overview of data processing and preparation for machine learning.

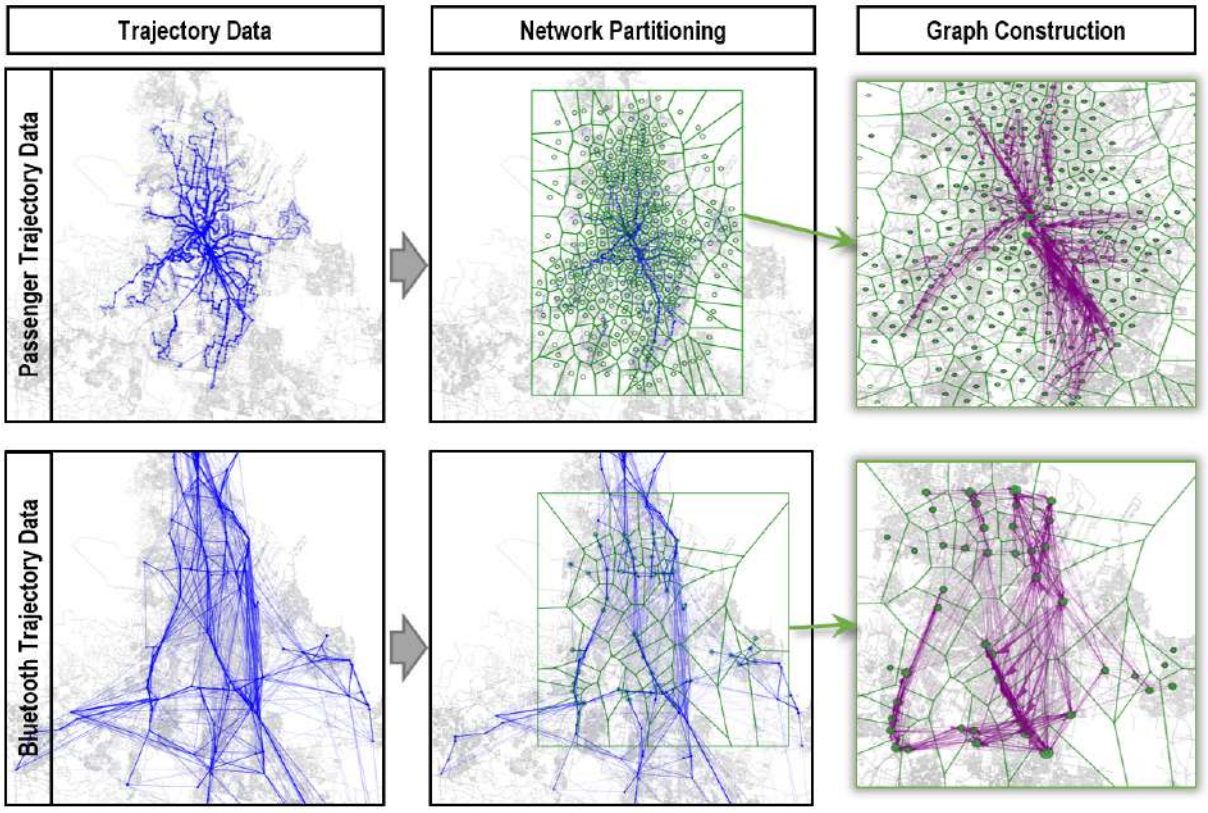


Fig. 2: Dynamic graph representation of different trajectory data.

the population movement as a dynamic directed weighted graph. Each vertex in the graph represents one *cell*, a polygonal region obtained from network partitioning, and each edge of the graph represents the *population flow* between regions. The graph is *dynamic* in that edge weight varies across different times based on the magnitude of traffic flow that evolves over time.

Representing movement data as a graph has been actively studied in visual analytics [1][5][11]. Inspired by these approaches, our previous work [7] has developed a systematic framework for dynamic graph representation that combines trajectory data acquisition, network partitioning, and graph construction. This paper extends [7] to further address different types of trajectory data, particularly to apply different network partitioning methods to different trajectory types, namely *radius-based* and *density-based* methods, as shown in Fig. 1. To partition a network,

data points in trajectories are grouped in space to obtain centroids of cells, which are then used to build a Voronoi diagram to define cell boundaries. The radius-based method aims to cluster points into equi-sized cells using approaches applied in [7][1], whereas the density-based method aims to cluster points into areas of higher density using Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm. The radius-based method is suitable for trajectories with relatively high spatial resolution (i.e., data points are distributed in space evenly and densely), while the density-based method is suitable for trajectories with low spatial resolution (i.e., trajectory sampling points are sparse). This study demonstrates this framework using two different datasets: *bus passenger trajectories* constructed from smart card records [7] and *Bluetooth vehicle trajectories* obtained from Bluetooth detectors. Fig. 2 illustrates cases of constructing graphs using these data, where the radius-based partitioning method is used for the passenger trajectories and the density-based method is used for the Bluetooth trajectories.

Once the network is partitioned into cells, these cells are used to define the vertex set of a flow graph. Given a fixed set of vertices, graph edges are defined for each time interval separately, where a directed edge is added from cell i to cell j if at least one trajectory visited from cell i to cell j, not necessarily consecutively, during the time interval. Edge weight thus represents the number of trajectories connecting two cells.

2.2 Feature Extraction and Dimensionality Reduction

The dynamic graph representation provides spatio-temporal information about population movement in a compact form when compared to the original set of trajectories. However, using a dynamic graph with tens or hundreds of vertices introduces a number of challenges when viewed from machine learning perspective:

1. The population movement graph is a dynamic directed graph. This means for N vertices representing N different regions in the city, the traffic flow matrix is N^2 dimensional. Additionally, since for any given time $t = 1, \ldots, K$ there is one graph, the overall data is a $N \times N \times K$ dimensional tensor. For a typical data sample, $N = 300, K = 24$. This leads to $300^2 \times 24 = 2,160,000$ dimensions.

2. Each vertex may only interact with a few other vertices at a time. This leads to sparse traffic flow matrix.

3. It is unknown what features should be extracted from dynamic graphs to characterize different days.

The large number of dimensions and the need for obtaining meaningful feature necessitates further processing of the data. In fact without dimensionality reduction, attempts to operate on the data using a desktop computer was slow and not practical. To overcome the dimensionality problem and to obtain meaningful features, we rely on non-negative tensor decomposition.

Non-negative Tensor Decomposition

Tensors can be considered a generalization of matrices and are useful for representing third or higher order data. Tensor decomposition which can be interpreted as a generalization of matrix factorization, is able to reveal latent features in data, while reducing data dimensionality. Advances in computational tools [3][4][2] has facilitated the increasing application of tensors in statistics, pattern recognition and machine learning. Recently, increasing number of studies have applied tensor decomposition in transportation areas. Applications include urban traffic pattern mining [8][6], missing traffic data completion [9], and short-term traffic prediction [10].

To understand how tensor decomposition works, a number of definitions must be laid out first.

Definition 1: Rank-1 tensor : An m-way tensor $\mathcal{X} \in \mathbb{R}^{N_1 \times N_2 \times \cdots \times N_m}$ is a rank-1 tensor if it can be written as an outer product of m vectors. In other words,

$$\mathcal{X} = \mathbf{a^{(1)}} \circ \mathbf{a^{(2)}} \circ \cdots \circ \mathbf{a^{(m)}}$$

Now, tensor decomposition can be defined in the following way.

Definition 2: Rank-R Tensor Decomposition is to approximate a given tensor $\mathcal{X}$ by the sum of R rank-1 tensors as shown below.

$$\mathcal{X} \approx \sum_{i=1}^{R} \mathbf{a_i^{(1)}} \circ \mathbf{a_i^{(2)}} \cdots \circ \mathbf{a_i^{(m)}}$$

For a third order tensor, we use the following notation.

$$\mathcal{X} \approx \sum_{i=1}^{R} \mathbf{a_i} \circ \mathbf{b_i} \circ \mathbf{c_i} = [\mathbf{A}, \mathbf{B}, \mathbf{C}]$$

Matrix A groups together the column vectors a_i, matrix B groups together column vector b_i, and similarly matrix C groups together the column vectors c_i.

This paper considers a third order tensor, i.e., a $N \times N \times K$ tensor, where its entry at (i, j, t) represents traffic flow from source cell i to target cell j at time interval t. Tensor decomposition is defined as an optimization problem.

$$\min ||\mathcal{X} - \hat{\mathcal{X}}||_2^2 \tag{1}$$

$$\hat{\mathcal{X}} = \sum_{i=1}^{R} \mathbf{a_i} \circ \mathbf{b_i} \circ \mathbf{c_i} \tag{2}$$

In order to make components of tensor decomposition interpretable and meaningful, non-negativity constraint must be imposed. By doing this, each component can be interpreted as an underlying spatio-temporal activity that when summed with other components, it can produce an approximation of the original data. In other words, the tensor decomposition results in an ensemble of spatio-patterns. Additionally, the non-negativity constraint can extract features that have physical meaning.

In addition to feature extraction, tensor decomposition leads to dimensionality reduction. If $\mathcal{X}$ is a $N \times N \times K$ tensor, then both A and B are $N \times R$ matrices, while C is a $K \times R$ matrix. The total number of entries in A, B, C is $(2 \times N + K) \times R$. Since R is often much smaller than N, this leads to a factor $\frac{N^2 \times K}{(2 \times N + K) \times R}$ reduction in dimensionality. For instance if $N = 300, K = 24, R = 3$, the reduction factor is $\frac{300^2 \times 24}{(2 \times 300 + 24) \times 3} \approx 1154$.

2.3 Characterization and Classification

The output of tensor decomposition is three matrices $A_{N \times R}, B_{N \times R}, C_{K \times R}$. Matrices A and B represent a measure of the interaction between different regions. More specifically, an entry a_{ir} of matrix A indicates a measure of how much cell i is acting as a source (upstream cell) in component r. Similarly an entry b_{jr} of matrix B indicates a measure of how much cell j is

acting as a target in component r. On the other hand, matrix C shows the temporal activity of a given component. For instance, entry c_{tr} indicates how active the component r is at time t. Therefore, the outer product of the vectors $a_r \circ b_r \circ c_r$ represents a spatio-temporal sub-pattern in the data. The superposition of all these sub-patterns approximates the original population movement data. For each set of A, B, C matrices that represent movement patterns for a single day, we assign the associated day-of-week label. We then design a classifier that learns patterns from data and classifies new data into different day-of-week groups. We tested various classifier models and found that Random Forest classifier performed best.

3 Case Study

3.1 Data and Study Sites

A case study was performed using six-month passenger trajectory data (Nov. 2012 - Apr. 2013) obtained from smart card data provided by TransLink, the transit agency in Brisbane and six-month Bluetooth trajectory data (July 2015 - Dec. 2015) provided by Department of Transport and Main Roads, the Queensland state government, Australia. The data covers Brisbane metropolitan area as shown in Fig. 2.

3.2 Results

To show what the output of tensor decomposition looks like, the data for one Monday and one Sunday are shown. As can be seen in Fig. 3 and Fig. 4 there are visible differences in population movement on different days. This is particularly evident in the temporal activity patterns.

To understand to what extent the ensemble spatio-temporal patterns help characterize different days, we trained a Random Forest classifier based on different types of clusters. To design the classifier, one-third of the data was put aside for test and the remaining two-thirds were used to train and cross validate. For passenger trajectory data, we found that the classification gave good results with four clusters namely, Sunday, Mid-Week (Monday, Tuesday, Wednesday, Thursday), Friday and Saturday. Using a Random Forest classifier with 200 trees, we achieved 90% accuracy on cross-validation and 93.8% on test set. The confusion matrix for classifier is shown in Fig. 7 The higher accuracy on test set may be an indicator that more samples are needed for the test set. Nevertheless, it is shown that given a daily traffic pattern, it is possible to distinguish the type of traffic pattern using ensemble spatio-temporal patterns. For Bluetooth trajectory data, however, the day-of-week patterns were less distinguishable than passenger trajectories. With a Random Forest classifier with 200 trees, the results showed that at best it is possible to distinguish Saturday, Sunday and weekday movement patterns. For this case, the classifier achieved 97.3% accuracy on cross-validation set and an accuracy of 94% on test set, as shown in Fig. 8

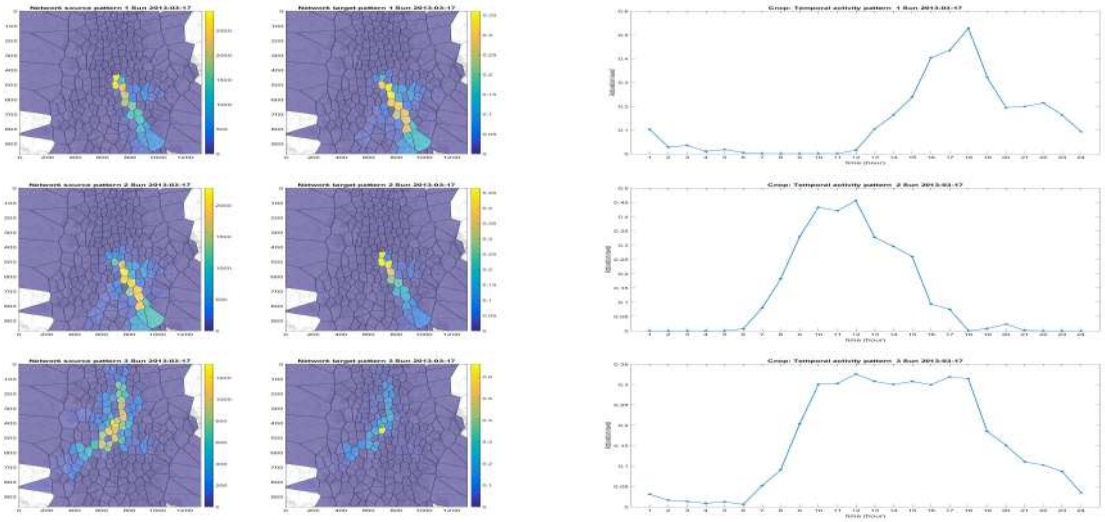


Fig. 3: Public transport spatio-temporal patterns on a Sunday. Each row shows one spatio-temporal component. Left, the source-target pattern. Right, the corresponding temporal activity pattern.

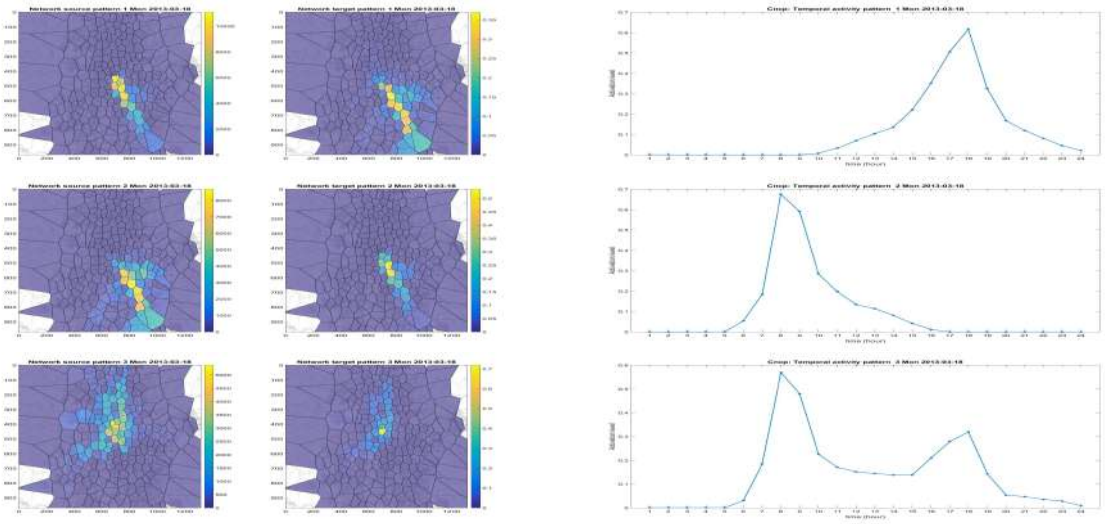


Fig. 4: Public transport spatio-temporal patterns on a Monday. Each row shows one spatio-temporal component. Left, the source-target pattern. Right, the corresponding temporal activity pattern.

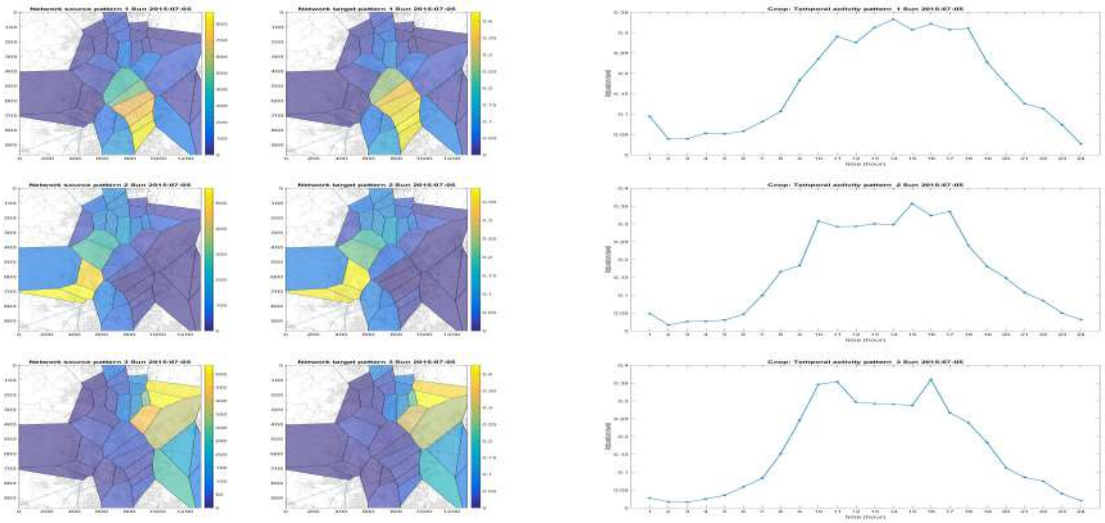


Fig. 5: Road traffic spatio-temporal patterns on a Monday. Each row shows one spatio-temporal component. Left, the source-target pattern. Right, the corresponding temporal activity pattern.

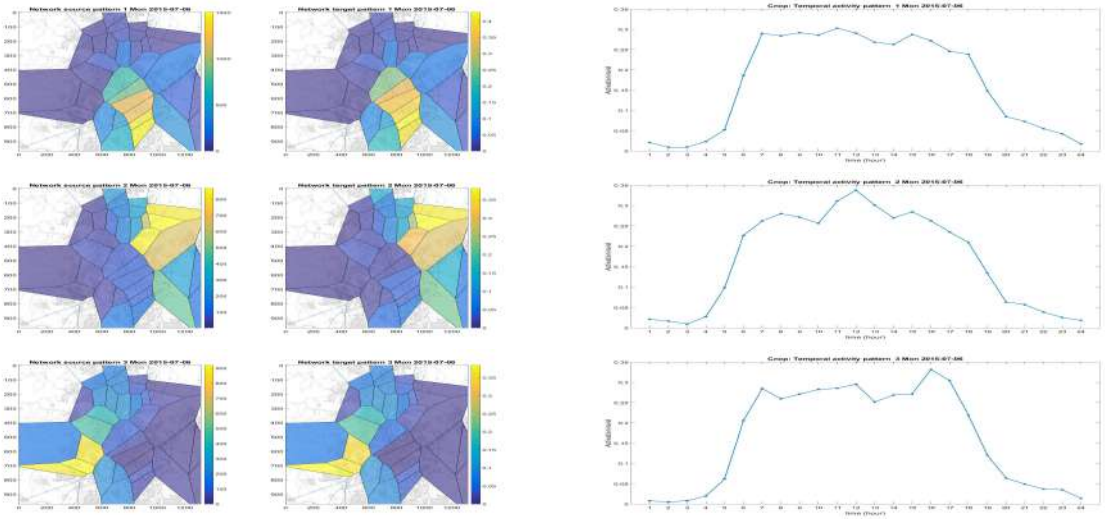


Fig. 6: Road traffic spatio-temporal patterns on a Monday. Each row shows one spatio-temporal component. Left, the source-target pattern. Right, the corresponding temporal activity pattern.

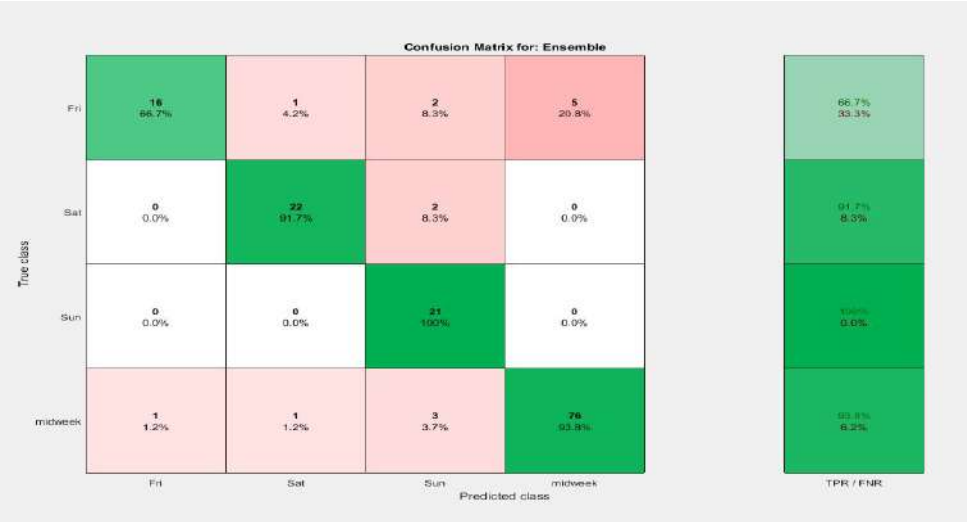


Fig. 7: Confusion matrix associated with Random Forest classifier trained with public transport data

Fig. 8: Confusion matrix associated with Random Forest classifier trained with road traffic transport data

4 Conclusion and Future Work

In this work, we demonstrated how non-negative tensor decomposition can be used to mine spatio-temporal patterns within a dynamic graph representation of population movement. Tensor decomposition of dynamic graphs provides patterns that can be easily interpreted as spatial and temporal features underlying the aggregated population movement. Additionally, tensor decomposition was shown to be very effective in achieving dimensionality reduction. We showed that using an ensemble of spatio-temporal patterns, it is possible to train classifiers that can distinguish different types of movement pattern.

In the bigger picture, this work demonstrated how machine learning can be applied to dynamic graphs. This is particularly important given the representational power of graphs. In the context of intelligent transport systems, the ability to apply machine learning techniques to dynamic graphs opens the possibility to achieve anomaly detection, incident management and data-driven transport planning.

The ability to go from a large amount of trajectory data spread across the city to visual and mathematical patterns is a significant step towards modelling of population movement in a manner that both spatial and temporal information is maintained. In the context of intelligent transport system, this allows the operators to understand where and when demands arise, understand population response to incidents and plan infrastructure developments and maintenance.

References

[1] N. Adrienko and G. Adrienko. Spatial generalization and aggregation of massive movement data. *IEEE Transactions on visualization and computer graphics*, 17(2):205–219, 2011.

[2] C. A. Andersson and R. Bro. The n-way toolbox for matlab. Available online, 2000.

[3] B. W. Bader and T. G. Kolda. Algorithm 862: MATLAB tensor classes for fast algorithm prototyping. *ACM Transactions on Mathematical Software*, 32(4):635–653, December 2006.

[4] B. W. Bader, T. G. Kolda, et al. Matlab tensor toolbox version 2.6. Available online, February 2015.

[5] D. Guo, S. Liu, and H. Jin. A graph-based approach to vehicle trajectory analysis. *Journal of Location Based Services*, 4(3-4):183–199, 2010.

[6] Y. Han and F. Moutarde. Analysis of large-scale traffic dynamics in an urban transportation network using non-negative tensor factorization. *International Journal of Intelligent Transportation Systems Research*, 14:36–49, 2016.

[7] J. Kim, K. Zheng, S. Ahn, M. Papamanolis, and P. Chao. Graph-based analysis of city-wide traffic dynamics using time-evolving graphs of trajectory data. In *Australasian Transport Research Forum 2016 Proceedings*, 2016.

[8] L. Sun and K. W. Axhausen. Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transportation Research Part B*, 91:511–524, 2016.

[9] H. Tan, G. Feng, J. Feng, W. Wang, Y.-J. Zhang, and F. Li. A tensor-based method for missing traffic data completion. *Transportation Research Part C*, 28:15–27, 2013.

[10] H. Tan, Y. Wu, B. Shen, P. J. Jin, and B. Ran. Short-term traffic prediction based on dynamic tensor completion. *IEEE Transactions on Intelligent Transportation Systems*, 17(8).

[11] T. von Landesberger, F. Brodkorb, P. Roskosch, N. Andrienko, G. Andrienko, and A. Kerren. Mobilitygraphs: Visual analysis of mass mobility dynamics via spatio-temporal graphs and clustering. *IEEE transactions on visualization and computer graphics*, 22(1):11–20, 2016.

Opportunities for floating car data in Integrated Traffic Management: the case of queue estimation

Serge Hoogendoorn, Erik-Sander Smits, and Jaap van Kooten

December 22, 2016

1 Introduction

Integrated Network Management (INM) entails the coordinated deployment of traffic management measures (traffic control, ramp metering, VMS, in-car systems) for the sake of improving the effectiveness of local traffic control and management measures. Recent pilots in The Netherlands have shown the potential of INM, but also have highlighted that for successful deployment, the following issues need to be resolved [Hoogendoorn et al. (2016)]:

1. Since the coordination concept is based on buffering traffic on other on-ramps and controller intersections than the active bottleneck, knowledge on the relation between the buffers and the active bottleneck is essential. This relation (the fraction of traffic in the buffer traveling to the bottleneck) can generally not be determined from loop detector information.

2. To determine the amount of bufferspace still left, and to determine if queues on off-ramps, on-ramps, and urban arterials are active bottlenecks, insight into the actual queue lengths is necessary. With the current loop lay out, it turned out impossible to determine a generic queue estimation technique based on loops only.

3. The INM approach depends on the timely and reliable identification of bottlenecks. Although the current algorithm using the inductive loops functions satisfactory, it remains to be seen if the use of additional or alternative data sources would improve the timely bottleneck identification.

For all of these issues, Floating Car Data (FCD) holds the promise of improving traffic monitoring and diagnostics for INM purposes. For the first point, FCD stemming from navigation devices or apps could provide the relevant information to determine the buffer effectiveness. For the third point, fusion of FCD and loop data using adequate filtering techniques has shown the potential to get more accurate state estimates when using data sources with different semantic features (e.g. [Van Lint et al. (2010)]).

This extended abstract focusses on the second issue, queue estimation, which turns out to be a difficult problem to tackle due to a number of reasons most of which related to measurement

error. None of the methods proposed in literature and tested during the INM pilot provided satisfactory and generic results.

The main research objective is to use the added value of floating car data in increasing the accuracy and reliability of queue estimates, where we will only consider current of-the-shelf FCD products (in this case from the company TomTom) that will be described in more detail in the remainder of the abstract. This in contrast to many other studies (e.g. [Kuwahara et al. (2013)]), that use vehicle trajectory data that often are collected for the sake of the study, but not being broadly available . To this end, we will develop three different queue estimation approaches to fuse the FCD with loop data. We will show that the data driven approach provides the most accurate estimates.

2 Data

Fig. 1 shows part of the data used for the analyses presented in this paper for one day (16th of June 2016) of the considered two month period (June - July 2016). On the far left side of the picture, the network link considered is shown, indicating the two lanes of the link that fork into five lanes (two left turning lanes, two through going lanes, one right turning lane). The graph shows the radar data that is used as ground truth. These data are based on individual vehicle trajectories that are used to determine the 10-second average speeds for 7 meter cells, shown by the speed contour plot. Note that an overview of the entire day is shown in the bottom of the picture (radar data), clearly showing the queue dynamics of this particular situation.

On top of the contour plot, we have plotted the (virtual) trajectories that are based on the FCD. These data provide 30-second average speeds for the link shown, dynamically sectioned into sublinks. Without going into detail, this sectioning provides more detailed information on the traffic conditions and speed heterogeneity on the link, given that sufficient FCD traces are available and there is a clear difference in traffic conditions for the different dynamic sections. Note that the circles show the locations where the sections begin and end. The FCD penetration is about 1-3%. Due to the way the raw data is processed, the FCD has a delay of 3 to 5 minutes.

Next to the floating car data, different (types of) loops are present at the considered site. These loops are located at the cross-sections at 259 meter, 37 meter, and 21 meter (next to the stop line) and provide occupancy and flow information (per lane). Note that the loop at 37 meter is a long loop.

In the remainder of the abstract, we focus on estimating the queue length on the link. This length is defined by the tail of the queue, where the queue is defined by using a speed threshold of 10 km/h. Note that other sensible definitions of the queue length exist, and that this choice is to an extent arbitrary, and in this case based on the application to INM as described in [Hoogendoorn et al. (2016)].

3 Methods

We have developed three queue estimation methods to integrate the FCD with the loop data. The first method is based on traffic flow theory, while the second and third method are data-driven

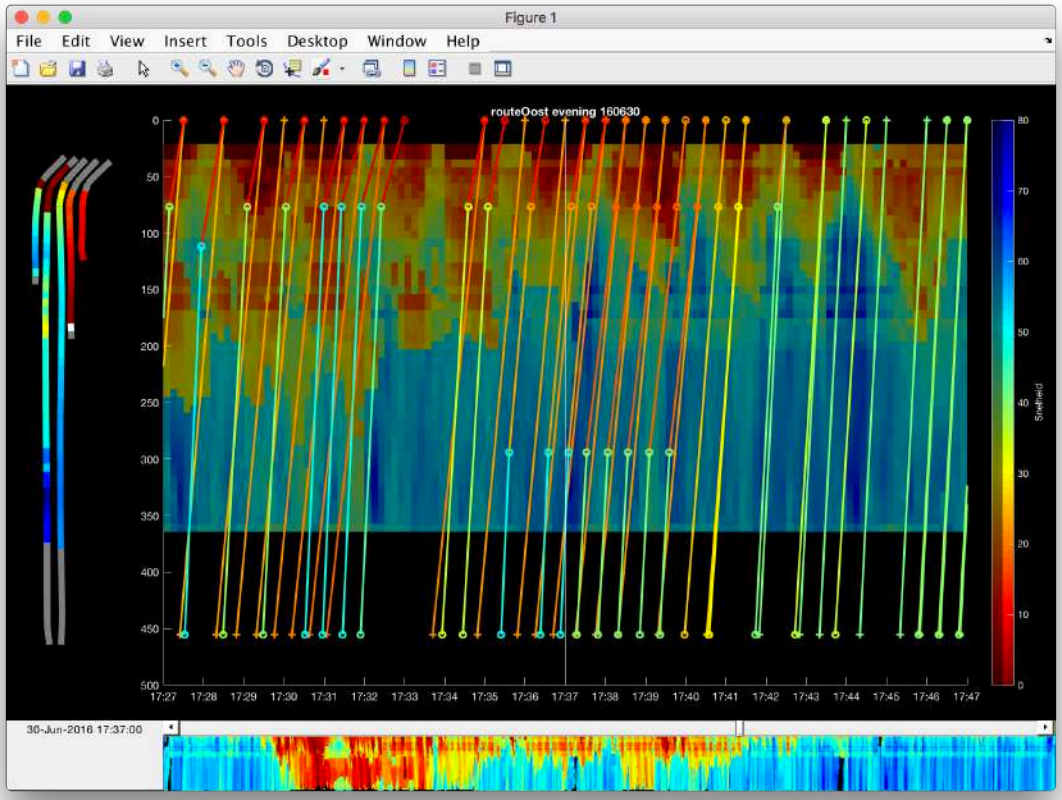


Figure 1: Screenshot of data analysis tool.

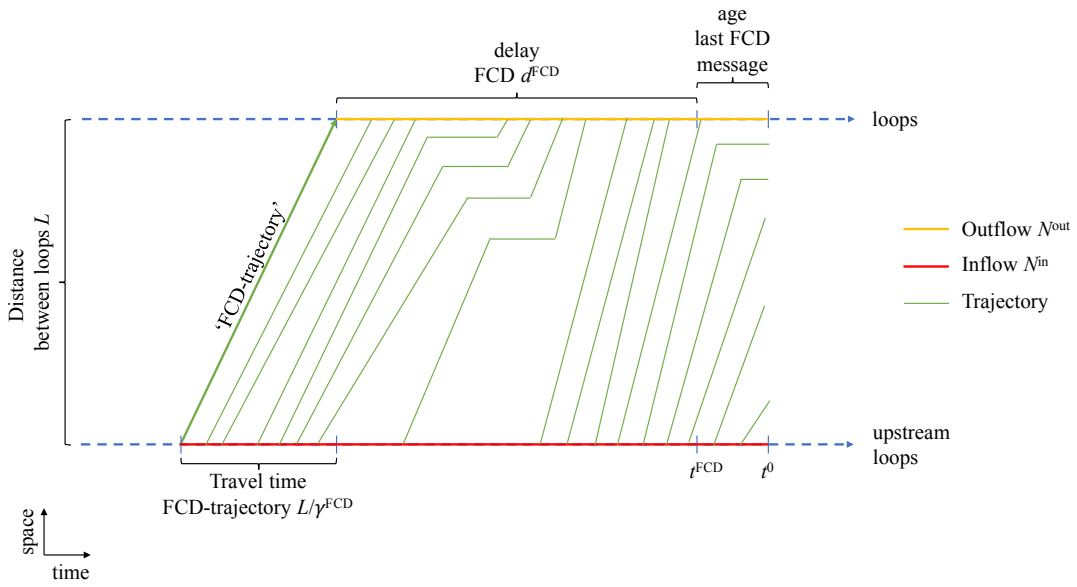


Figure 2: Schematic explanation of the data fusion queue estimator.

approaches based on simple machine learning approaches.

3.1 Traffic flow theory data fusion queue estimator

Fig. 2 graphically explains the traffic flow theory based method. The method uses the most recent (reliable) FCD to construct a vehicle trajectory. By determining the inflow and outflow since that vehicle has passed, we determine the number of vehicles in the link. The example in the figure shows how the number of vehicles in the link equals the difference between the cumulative inflow $N^{in} = 16$ and the cumulative outflow $N^{out} = 11$, meaning that there are 5 vehicles on the link. The flows are determined from the inductive loop at the entry and the exit, which poses the problem of measurement error. In fact, during the development of the method it turned out that these measurements are not reliable, in particular when traffic becomes denser. Therefore, we have used the occupancy rates to determine the flows indirectly.

3.2 Data-driven approach using radar as training data

The second approach uses what we refer to as grey box modelling, which means that we use a statistical model to correlate the input (FCD, loop data) and the output (queue lengths) that has features relevant for the process we aim to model. In this case, the objective to model queue lengths implies that we need a model that features forms of integration due to the conservative nature of the process. Therefore, we have used an autoregressive model relating the queue length $r(t)$ to the estimates from previous periods $r(t-1), r(t-2), ...$, the loop data input (flows,

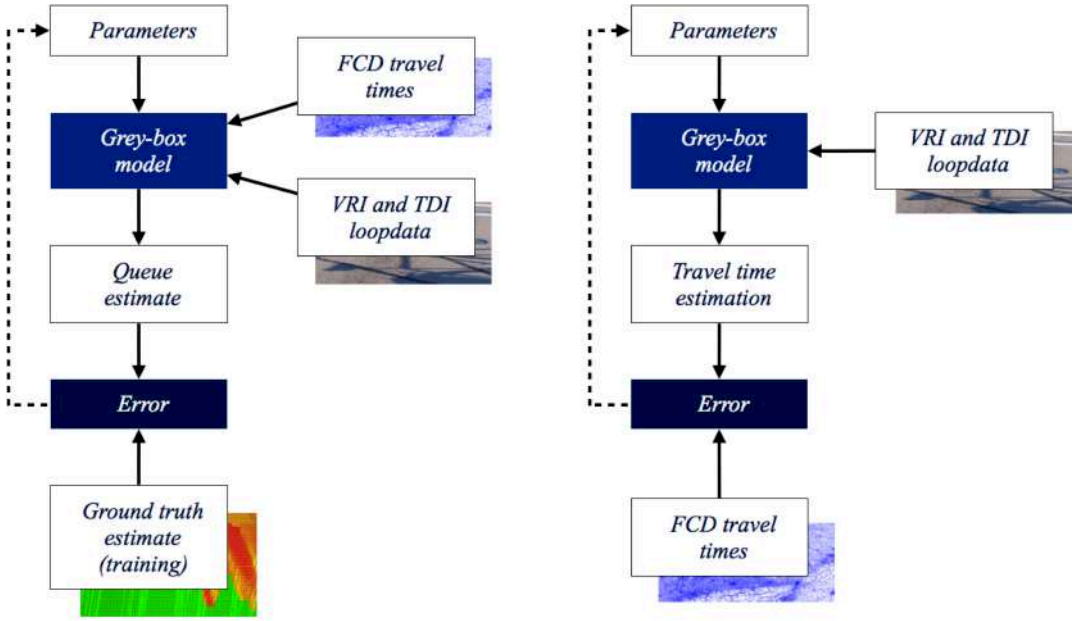


Figure 3: Visualisation of the data driven approaches (method 2 and 3).

occupancies) $u_1(t-j-n_1)$ and the FCD input $u_2(t-j-n_2)$ for $j = 0,1,...$ as follows:

$$\begin{aligned} r(t) &= a_1 r(t-1) + a_2 r(t-2) + a_3 r(t-3) + ... && (1)\\ &\quad + b_0^1 u_1(t-n_1) + b_1^1 u_1(t-n_1-1) + b_3^1 u_1(t-n_1-2) + ... && (2)\\ &\quad + b_0^2 u_2(t-n_2) + b_1^2 u_2(t-n_2-1) + b_3^2 u_2(t-n_2-2) + ... + \varepsilon(t) && (3) \end{aligned}$$

Estimation of the parameters of the model can be done using standard least-squares approaches, which will not be described in this abstract. Fig 3a shows a graphical representation of the modelling and estimation approach. Note that we assume that ground truth queue estimates are available for training (and validation) purposes. For the present case, the ground truth is available from the radar detectors. In the remainder we will use part of the data (10 days to be precise) for training of the model; the remainder of the data is used for validation purposes.

The approach allows to determine which of the input is most important to determine queues accurately. This can be achieved by specifying different model designs using specific model input (e.g. all available data, only loop data, only FCD, specific loops, etc.) to estimate the queue lengths.

Note that we will assume that the FCD are actual, i.e. the 3-5 minute delivery delay is not taking into account. Although this hampers the practical application of the method (i.e. queue estimates are in fact 3-5 minutes old, which for applications in urban traffic control is possibly too long), it will reveal what we can potentially do with the FCD.

3.3 Data-driven approach using FCD as training data

The final method developed is also based on the modelling approach described in the previous section. However, in this case we do not use the radar data for training purposes, but instead we use the FCD. Fig. 3b outlines the approach. The general idea behind the approach is that the autoregressive model would find that best parameters to relate the input (from the loops) to the erroneous output (the FCD). This to an extent resolves the problem that the FCD can be at times inaccurate due to the few number of vehicles on which the data is based at a specific moment in time.

A major benefit of this method compared to the previously described method is that it does not require radar data for training purposes. This makes the method more generically applicable. Furthermore, the FCD delivery delay of 3-5 minutes issue is resolved as well, since the FCD is used only for training.

4 Results and discussion

To assess the results, we consider different error measures reflecting both the accuracy and the reliability of the estimator. For the accuracy, we have consider the Mean Absolute Error (MAE) and the Root Mean Squared Error (RMSE). For the reliability, we consider the probability $P(p)$ of the estimate to be $p\%$ larger than then ground truth (for each 30 seconds), with $p = 10, 25$ and 50%.

Tab. 1 shows a selection of the results of the study. From the table, we can conclude that method 2 using all available data outperforms the other methods both in terms of accuracy and reliability. The performance of the method reduces when only FCD is available. In this case however, the performance is only slightly worse than the performance of method 1. In fact, very large outliers (error larger than 50%) actually occur less frequent (see value of $P(50)$), although further studies would need to provide further evidence that this is indeed a general feature.

In particular the results of method 3 are interesting: the use of FCD as a means to train the autoregressive model seems to result in an estimator that performs comparable to the other methods.

Method	Description	MAE	RMSE	$P(10)$	$P(25)$	$P(50)$
Method 1	Using occupancy during green	19.46	34.89	52.8%	21.6%	8.1%
Method 2	All data	15.53	24.93	49.9%	17.3%	5.1%
Method 2	Only FCD	19.79	30.66	58.8%	23.7%	6.5%
Method 3	FCD as training	18.96	32.67	61.0%	21.6%	7.5%

Table 1: Overview of queue estimator methods performances. For method 2, the performance pertains to the validation set only.

Although providing a detailed account of the implications of these results for INM purposes is beyond the scope of this abstract, we can conclude that the methods proposed seem to provide more accurate and reliable queue estimates than the previously used loop-based ap-

proaches [Hoogendoorn et al. (2016)]. In fact, we - somewhat hesitantly - conclude that the performance is adequate for current INM applications (in particular buffering), given that we explicitly consider the unreliability of the estimates in the INM design. That is, understanding that we underestimate the queue with say 10% in 20% of the cases allows us to determine the required margins in the buffer lengths.

Looking at the specific methods, we can conclude that each of the approaches has if pros and cons. Method 1 has the benefit that no training (and hence no training data) is necessary. Its downside lies in the configuration of the system and the tuning to gather for the specific characteristics of the considered location. Method 2 has the benefit that it outperforms the other methods. Furthermore, it does not require complex configuration of the model. However, it does require the dedicated collection of ground truth data which may be costly and cumbersome for large scale field applications.

Method 3 underperforms compared to the other methods, although the differences are relatively small. Next to limited configuration effort, no dedicated data collection is needed. In fact, we could cater for auto tuning of the model (e.g. each moment new FCD becomes available) making the approach adaptive to for instance changes location conditions (e.g. in case of an incident blocking a lane on the arterial).

Further research will be geared towards investigating the relation between the quality of the queue estimates and the effectiveness of the INM approach, but at first glance, all methods seem promising angles for developing a robust queue estimation module. Next to this, investigating the impact of the FCD quality (related to the penetration rate) is of importance to understand when loop data are no longer required for accurate queue estimation. Furthermore, we could consider the use of the raw FCD trajectory data as a data source to further improve the queue estimates. However, since this is currently not an 'of the shelf' product, it needs to be seen if the improvement in queue estimation quality would outweigh the investment cost.

References

[Hoogendoorn et al. (2016)] S. P. Hoogendoorn, J. Van Kooten, R. Adams, Lessons Learned from Field Operational Test of Integrated Network Management in Amsterdam. Transportation Research Record: Journal of the Transportation Research Board Jan 2016, Vol. 2554, pp. 111-119

[Van Lint et al. (2010)] J.W.C. Van Lint, S.P. Hoogendoorn, A robust and efficient method for fusing heterogeneous data from traffic sensors on freeways. Computer?Aided Civil and Infrastructure Engineering 25 (8), 2010, 596-612

[Kuwahara et al. (2013)] M. Kuwahara, T. Ohata, T. Takigawa, Estimating Vehicle Trajectories on a Signalized Urban Arterial and a Motorway by Data Fusion of Probe and Detector Data. In Optimum 2013 - international symposium on recent advances in transport modelling 2013 (pp. 1-11).