
Northwestern | TRANSPORTATION CENTER
DTA 2023
Beyond DTA
9TH INTERNATIONAL SYMPOSIUM ON
DYNAMIC TRAFFIC ASSIGNMENT
CONFERENCE
JULY
10-12
2023
Northwestern University | Evanston Campus
ABSTRACTS
WWW.SITES.NORTHWESTERN.EDU/DTA2023/

# DTA 2023

## *BEYOND DTA*

# ABSTRACTS

# Session 1

Day-to-day Traffic Dynamics with Strategic Commuters
*Minghui Wu, Yafeng Yin and Jerome Lynch*

Adjustment process of adaptive signal control strategies with route choices: a case study with Policy P0
*Koki Satsukawa, Takamasa Iryo, Naoki Yoshizawa, Michael J. Smith and David Watling*

# Extended Abstract: Multiday Traffic Dynamics with Strategic Commuters

Minghui Wu
University of Michigan

Yafeng Yin
University of Michigan

Jerome P. Lynch
Duke University

## 1. Introduction

With rapidly-advancing communications and vehicular technologies, commuters are increasingly connected. Connectivity allows drivers to access various decision-support technologies such as a navigation app that assist them in making travel decisions. Such decision support is expected to become even stronger in the coming decades with the development and deployment of driving automation when drivers feel comfortable with rendering more driving and travel agency to machines. In the connected and automated mobility era, commuters (connected drivers or automated vehicles) will possess strong computation capability, enabling them to make more foresighted and strategic travel decisions. Instead of being myopic, commuters can optimize their decision sequences for a longer time range, which results in a lower total cost because there is inertia or cost associated with switching route choices. It is thus intriguing to investigate how the behavior of, and interaction among strategic commuters would dictate the traffic pattern evolution.

In this paper, we model the multiday travel choice problem as a mean field game, where the decision-making process of individual commuters is explicitly modeled; each commuter makes optimized decisions instead of following a heuristics or behavioral rule. On the ground of the proposed model, we investigate the multiday traffic pattern by examining the mean field equilibrium where a commuter cannot reduce their overall cost by changing their policy sequence. We then conduct a thorough analysis of the properties of the MFE such as its existence and uniqueness. We also prove that under mild conditions, the MFE will converge to stationarity regardless of the initial flow pattern. Connection with traditional Wardropian equilibrium is established by examining the

physical interpretation of the stationary solution. To the best of our knowledge, this paper is the first to model and analyze multiday commute choices with strategic and farsighted commuters, accounting for the multi-commodity flow and heterogeneity in cost preferences.

## 2. Selected results of computational studies

In this section, the proposed model is applied to route choices in the Nguyen-Dupuis network [1] as shown in Figure 1. We consider a planning horizon of 30 days. There are four OD pairs in the network, i.e., OD 1 (Node 1→ Node 2), OD 2 (Node 1→ Node 3), OD 3 (Node 4→ Node 2), and OD 4 (Node 4→ Node 3). The OD demands are 4,130 vehicles per hour for OD 1 and OD 4, and 1,870 vehicles per hour for the other two OD pairs. In total, the network has 19 links and 25 paths, as elaborated in Table 1. The link performance function is $t(x) = 3\left[1 + 4\left(\frac{x}{2200}\right)^4\right]$ minutes for all links. The inertia cost is $d(s, s') = \epsilon \cdot \mathbf{1}_{s=s'}$, which indicates that commuters will receive $\epsilon$ cost whenever they switch to another route. The inertia weight $\epsilon$ equals 0.5 for the demands in the first three OD pairs and equals 2 for the fourth OD pair. The dispersion parameter $\theta$ of the perception error equals 3 for all demands. Fictitious play [2, 3] is used as the algorithm to solve for the MFE.

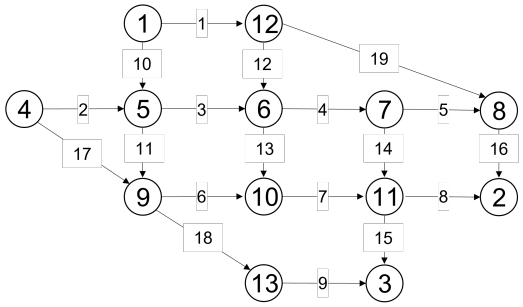



**Figure 1 Nguyen and Dupuis Network**

The MFE, represented in path flow evolution, is shown in Figure 2. Each dot represents the flow of the path. Here we use curves connecting dots on the same day to better demonstrate the evolution. To make the figure clear, we only label the curve for the first three days and the last day. The system is initialized in the blue curves. Except for the first OD pair which reaches the final flow very quickly, all other OD pairs demonstrate a clear convergence pattern. For a better illustration of the evolution pattern, Route 9,11, and 12 from OD2 are plotted in Figure 3. The flow on the first two paths first goes up and then drops down. On contrary, the path flow on Route 12 monotonically increases.

## Acknowledgments

| OD pair | Path No. | Link No. |
|---|---|---|
| 1-2 | 1 | 3,4,5,10,16 |
| | 2 | 3,4,8,10,14 |
| | 3 | 3,7,8,10,13 |
| | 4 | 6,7,8,10,11 |
| | 5 | 1,4,5,12,16 |
| | 6 | 1,4,8,12,14 |
| | 7 | 1,7,8,12,13 |
| | 8 | 1,16,19 |
| 1-3 | 9 | 3,4,10,14,15 |
| | 10 | 3,7,10,13,15 |
| | 11 | 6,7,10,11,15 |
| | 12 | 9,10,11,18 |
| | 13 | 1,4,12,14,15 |
| | 14 | 1,7,12,13,15 |
| 4-2 | 15 | 2,3,4,5,16 |
| | 16 | 2,3,4,8,14 |
| | 17 | 2,3,7,8,13 |
| | 18 | 2,6,7,8,11 |
| | 19 | 6,7,8,17 |
| 4-3 | 20 | 2,3,4,14,15 |
| | 21 | 2,3,7,13,15 |
| | 22 | 2,6,7,11,15 |
| | 23 | 2,9,11,18 |
| | 24 | 6,7,15,17 |
| | 25 | 9,17,18 |

**Table 1 Path-link relationship**

## References

[1] Sang Nguyen and Clermont Dupuis. "An efficient method for computing traffic equilibria in networks with asymmetric transportation costs". In: *Transportation Science* 18.2 (1984), pp. 185–202.

[2] Sarah Perrin et al. "Fictitious play for mean field games: Continuous time analysis and applications". In: *Advances in Neural Information Processing Systems* 33 (2020), pp. 13199–13213.

[3] Julien Perolat et al. "Scaling up mean field games with online mirror descent". In: *arXiv preprint arXiv:2103.00623* (2021).

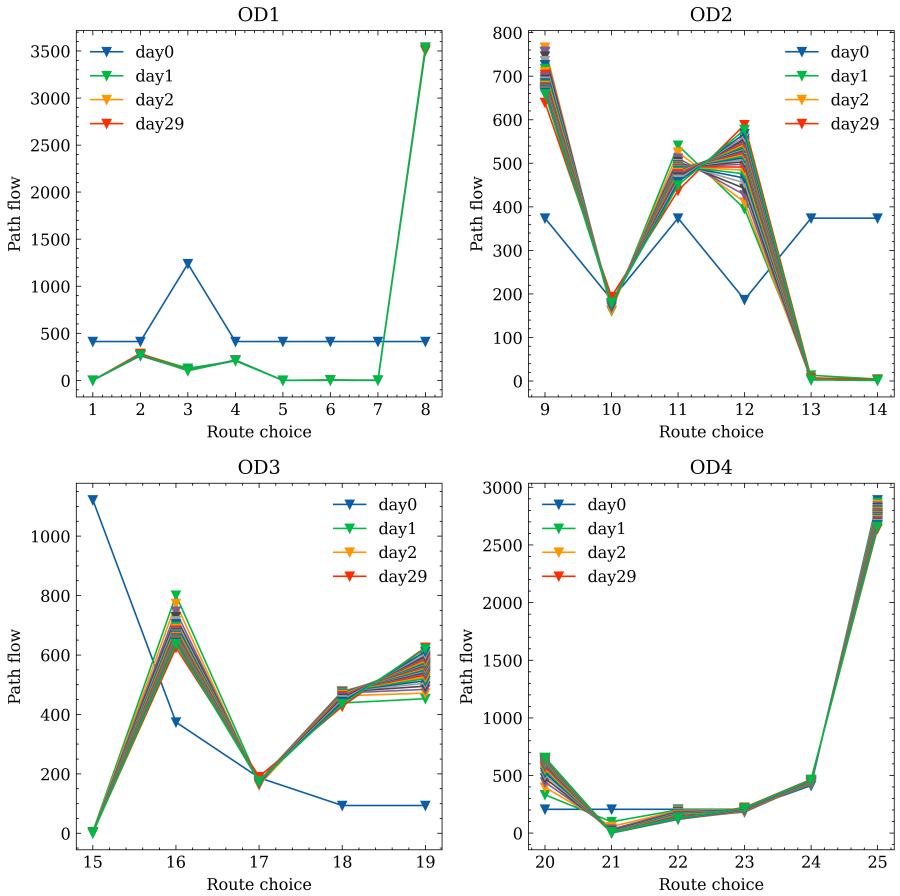



**Figure 2** **Path flow evolution**

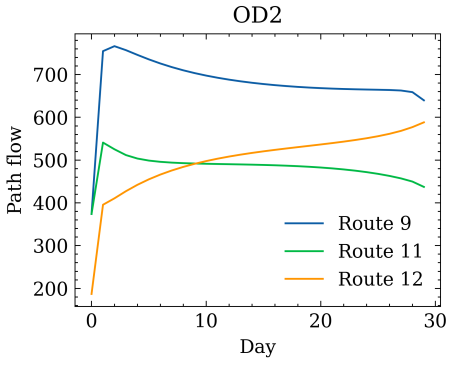



**Figure 3** **Path flow evolution of OD2**

# Adjustment process of adaptive signal control strategies with route choices: a case study with Policy $P_0$*

Koki Satsukawa[a,*], Takamasa Iryo[a], Naoki Yoshizawa[a],
Michael J. Smith[b], and David Watling[c]

[a]*Graduate School of Information Sciences, Tohoku University, Japan*
**Corresponding author: satsukawa@tohoku.ac.jp.*
[b]*Department of Mathematics, University of York, United Kingdom of Great Britain and Northern Ireland*
[c]*Institute for Transport Studies, University of Leeds, United Kingdom of Great Britain and Northern Ireland*

## 1 Introduction

A number of studies have been conducted for adaptive signal control strategies considering drivers' responses (e.g. Xie et al., 2014; Thunig and Nagel, 2019; Xie and Wang, 2019; Yongqiang, 2020; Safadi and Haddad, 2021). The interaction between adaptive signal controls and drivers' route choices constitutes a dynamical system, either day-to-day or within-day, that may or may not converge to a stationary state in which both signal controls and route choices are equilibrated. It raises concerns regarding properties such as the uniqueness and stability of stationary state(s), which have been extensively studied in dynamic traffic assignment problems (e.g. Iryo (2008), Iryo (2011), Iryo (2013), Guo et al. (2018), Iryo (2019)). Investigating these properties is crucial for effective implementations of adaptive control systems. For example, if multiple stable equilibria exist, we have to identify which equilibrium will be realised to assess the performance of the adaptive signal control.

In the present study, we particularly consider the signal control strategy called *Policy* $P_0$ (Smith (1979)). It aims to adjust the green splits of a signalised intersection to prevent the situation in which drivers' route choices and the green split concentrate on a phase whose saturation flow rate is lower than others. As a result, the (static) user equilibrium (UE) flow and consistent green split patterns can be stably realised for a relatively large traffic demand: this property is called 'capacity maximisation'. It implies a desirable characteristic of Policy $P_0$ in near-saturated conditions.

Compared to the near-saturated conditions, few investigations have been conducted for unsaturated conditions with Policy $P_0$. It would be important to investigate the properties of Policy $P_0$, such as the uniqueness and stability of stationary states, for various traffic demand levels to clarify its desirable and undesirable characteristics, which might be in common with other adaptive signal control policies.

This study aims to numerically investigate the dynamical properties of Policy $P_0$ against the traffic demand change to a network. We first define an equilibrium state as a state in which the UE condition and an equilibrium condition for the signal control of Policy $P_0$. We next compute the equilibrium states by changing the traffic demand from unsaturated to near-saturated conditions and examine the theoretical properties such as uniqueness and stability for every traffic demand.

## 2 Traffic assignment and adaptive signal control models

We consider a signal-controlled network with one-to-one origin-destination (OD) demand (Figure 1). Users select link 1 or link 2, which is corresponding to route 1 or 2, to travel from the origin to the

*Keywords:** adaptive signal control, policy $P_0$, user equilibrium, route choice, stability

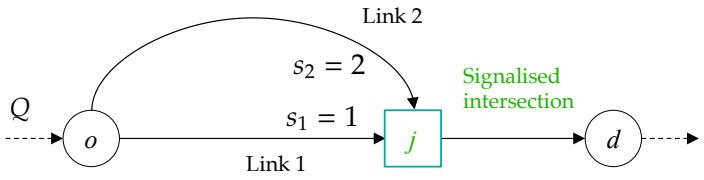



Figure 1: One-to-one origin-destination network

destination. The total travel demand is denoted by $Q$. The saturation flow rate of a link $i \in \{1, 2\}$ is denoted by $s_i$. The free-flow travel time of route 2 is larger than that of route 1 and the difference is denoted by $\Delta\tau$ ($> 0$). The node $j$ is a signalised intersection where the green splits of the two links incoming to the node, $\lambda_1$ and $\lambda_2$, are controlled. The sum of the green splits is assumed to be one. The minimum and maximum green splits for each link are denoted by $\lambda_{\min}$ and $\lambda_{\max}$ ($= 1 - \lambda_{\min}$).

For calculating the delay on the incoming links at the signalised intersection, we utilise Webster's delay formula (Webster, 1958) with slight modifications. Specifically, the expected intersection delay on link $i \in \{1, 2\}$ is defined as follows referring to Smith (1979):

$$d_i(f_i, \lambda_i) = \frac{9}{20}\left[\frac{c(1-\lambda_i)^2}{1-y_i} + r\frac{y_i^2}{f_i\lambda_i(\lambda_i - y_i)}\right], \tag{1}$$

where $f_i$ is the flow on link (i.e. route) $i$, $s_i$ is the saturation flow rate and $c$ is the cycle time. The notation $y_i$ is $f_i/s_i$. We assume that the under-saturation condition is satisfied so that the delay formula is valid: $f_i < s_i\lambda_i$. $r$ is a weight coefficient for the second term. We introduce this to observe the behaviour of equilibrium states under a variety of delay functions with changing $r$. Note that we do not define the travel time of the link 3 since it does not affect users' route choices and signal control.

Under the above settings, we define an equilibrium state as a state in which (i) the route flow pattern satisfies Wardrop's first principle (Wardrop, 1952), i.e. at user equilibrium (UE), and (ii) the green split pattern satisfies a signal control condition of policy $P_0$ proposed by Smith (1979). Equilibrium condition (i) means that the routes selected by users have the same travel time and it is not larger than the travel time of unused routes. This is formulated as follows in our problem:

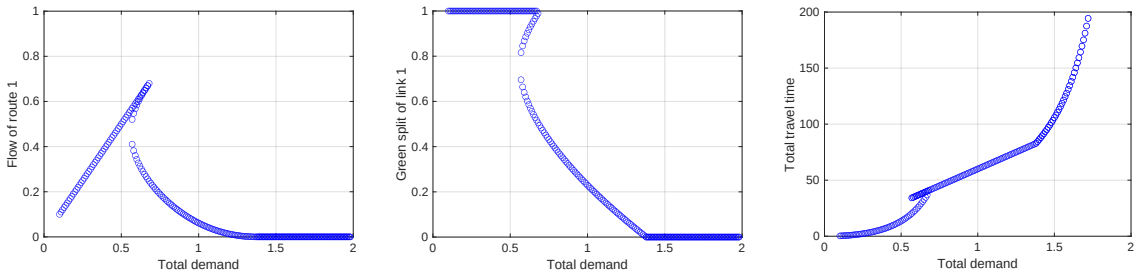
$$\begin{cases} \text{(A) } d_1(f_1, \lambda_1) = d_2(f_2, \lambda_2) + \Delta\tau & \text{if } f_1 > 0 \cap f_2 > 0, \\ \text{(B) } d_1(f_1, \lambda_1) \le d_2(f_2, \lambda_2) + \Delta\tau & \text{if } f_2 = 0, \\ \text{(C) } d_1(f_1, \lambda_1) \ge d_2(f_2, \lambda_2) + \Delta\tau & \text{if } f_1 = 0. \end{cases} \tag{2}$$

It should be noted that there is a possibility that case (C), in which the travel time of route 1 is equal to or larger than that of route 2, will occur even though the free-flow travel time of route 2 is larger than the other. This is because Webster's delay function produces a positive delay for the zero link flow, i.e. $d_i(0, \lambda_i) > 0$. Hence, the travel time of route 1 may become larger than the other even if $f_1 = 0$ depending on the parameter settings of Eq (1).

We next explain Equilibrium condition (ii), policy $P_0$. Under the signal control policy, the following 'pressure' of each incoming link to the signalised intersection is first calculated:

$$p_i(f_i, \lambda_i) = \begin{cases} s_i d_i(f_i, \lambda_i), & \text{if } f_i > 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall i \in \{1, 2\}. \tag{3}$$

Then, the green split of the link whose pressure is larger than the other is increased so as to reduce the delay for equalising the pressures of the links. As the result, the equilibrium condition of the signal




(a) Flow of link 1 (b) Green split of link 1 (c) Total travel time

Figure 2: Flow and green split of link 1 and the total travel time at the equilibrium states

control is formulated as follows:

$$\begin{cases} (1)\ p_1(f_1, \lambda_1) = p_2(f_2, \lambda_2) & \text{if } \lambda_{\min} < \lambda_1 < \lambda_{\max}, \\ (2)\ p_1(f_1, \lambda_1) \leq p_2(f_2, \lambda_2) & \text{if } \lambda_1 = \lambda_{\min}, \\ (3)\ p_1(f_1, \lambda_1) \geq p_2(f_2, \lambda_2) & \text{if } \lambda_1 = \lambda_{\max}. \end{cases} \tag{4}$$

Case (1) represents a situation where the pressures of the two links are equal. The other cases represent the situations where the pressure of either link is still larger than the other even when the split is maximised, i.e. the traffic state is at a corner solution.

It should be noted that the second case in Eq. (3) (i.e. the pressure of the zero-flow link) is added to the original definition of Policy $P_0$ in this study. This prevents the occurrence of a positive pressure for a link with zero flow, which causes unnecessary assignment of the green split to such a route.

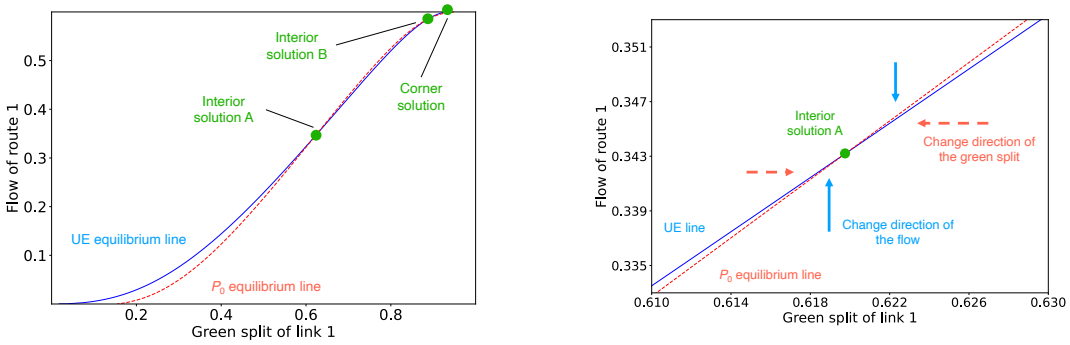
## 3 Numerical experiments

Numerical experiments are conducted to investigate how equilibrium flow, split and cost patterns change in response to the change of $Q$. We change $Q$ from 0.1 to 2.0, which is the saturation flow rate of route 2, by 0.01, and calculate equilibrium traffic states for each $Q$. We here set $\Delta\tau$ to 30 and $r$ to 60. Equilibrium states are numerically derived using the `vpasolve` function of MATLAB.

Figure 2 shows the results. Interestingly, we observe the existence of multiple equilibria for values of $Q$ in the middle of the tested range ($0.57 \leq Q < 0.68$). There exists one corner solution satisfying the route choice condition (B) and signal control condition (1). There also exist two interior solutions satisfying the conditions (A) and (1). In the latter case, both of the two routes are utilised. Figure 2(c) shows that the total travel times are also different between these multiple equilibria. Specifically, the total travel time of the corner solution is smaller than the total travel times of the interior solutions. Note that the interior solutions have the same total travel time; this can be confirmed from mathematical analysis of the equilibrium conditions.

For a larger $Q$ ($0.68 \leq Q$), the equilibrium again becomes unique. At the equilibrium, $\lambda_1$ becomes smaller as $Q$ increases by allocating the green split to link 2 with a higher saturation flow rate to handle the large traffic demand. $\lambda_2$ is eventually maximised in near-saturated conditions; this confirms that Policy $P_0$ achieves the capacity maximisation.

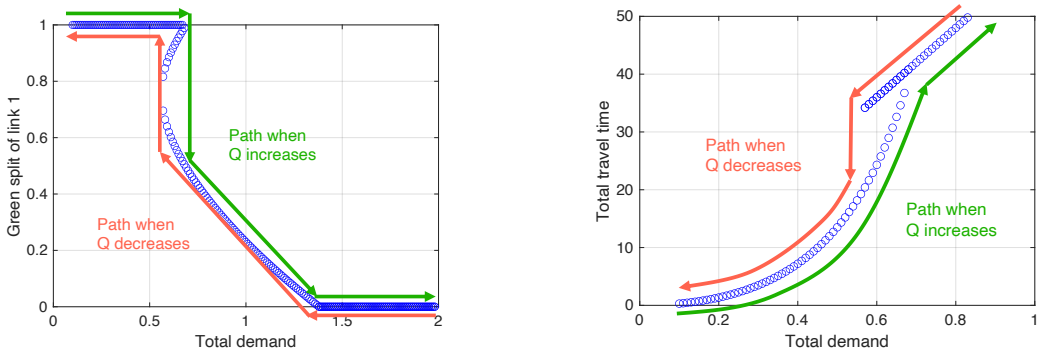
It is important to examine the stability of the equilibria for investigating whether equilibrium is a plausible state to be actually realised (Beckmann et al., 1956). To this end, we employ an approach utilising a 'phase diagram' referring to the work of Iryo and Watling (2019) (see Figure 3(a)). In the phase diagram, there are two lines, each of which represents the patterns of $f_1$ and $\lambda_1$ satisfying the




(a) Phase diagram and equilibria

(b) Around the interior solution A

Figure 3: Phase diagrams when $Q = 0.60$




(a) Green split of link 1

(b) Total travel time

Figure 4: Hysteresis loops

equilibrium condition (2) or (4). These lines enable us to understand the directions of their changes of at a given non-equilibrium state under natural evolutionary dynamics consistent with the equilibrium conditions. Specifically, under such dynamics leading a traffic state towards equilibrium (i.e. satisfying the positive correlation property (Sandholm, 2010)), $f_1$ and $\lambda_1$ change towards the lines. Therefore, we can graphically examine the stability of the equilibria from the change direction around the equilibria, i.e. intersection points (interior solutions) and the boundary point (corner solution).

Figure 3(b) shows the result of the analysis around the interior solution A. This figure suggests the stability of the interior solution because the change directions (i.e. arrows in the figure) point towards the solution. We can also confirm that the corner solution is stable, but the interior solution B is unstable. Consequently, there exist two stable equilibria and one unstable equilibrium.

# 4 Discussion

The existence of the multiple stable equilibria could cause a *hysteresis loop* of traffic states associated with the change of the total OD demand $Q$. Figure 4 graphically explains about this phenomenon. This figure shows that the corner solution is realised when $Q$ increases to the values within the range of the existence of the multiple equilibria from a lower value. Then, the (stable) interior solution becomes realised when $Q$ exceeds the range, and the green split of route 1 decreases as $Q$ increases after the sudden decrease at $Q = 0.68$. Meanwhile, the stable interior solution is realised when $Q$ decreases to the range, like after the peak hour. The green split suddenly increases and the corner solution becomes

realised when $Q$ decreases to $0.57$. As a result, a clockwise hysteresis loop exhibits higher green splits of route 1 as $Q$ increases than as it decreases. Note that the same applies to the flow of route 1.

Moreover, these hysteresis loops cause the hysteresis loop of network performance, as shown in Figure 4(b). This suggests an anti-clockwise loop of the total travel time within the multiple equilibria range: the hysteresis loop exhibits lower total travel times as $Q$ increases than as it decreases.

These results suggest the importance of investigating the behaviour of unsaturated traffic states under Policy $P_0$. Policy $P_0$ is originally proposed to achieve capacity maximisation, i.e. maximise the network performance in *near-saturated conditions*. Meanwhile, this study observed the existence of multiple equilibria and hysteresis loops in *unsaturated conditions*, which affects the network performance. Therefore, it is desirable to incorporate another signal control policy with Policy $P_0$ complementarily for improving the network performance in the unsaturated conditions.

In addition, we will show the theoretical and numerical analysis for various delay functions and adaptive signal control policies. We will also extend our analysis to problems with within-day dynamics, for example, by employing quasi-dynamic traffic assignment models.

## Acknowledgements

This study was financially supported by JSPS KAKENHI Grant Numbers 20K14843 and 20H00265.

## References

Beckmann, M., McGuire, C.B., Winsten, C.B., 1956. Studies in the Economics of Transportation. Yale University Press, New Haven.

Guo, R.Y., Yang, H., Huang, H.J., 2018. Are we really solving the dynamic traffic equilibrium problem with a departure time choice? Transportation Science 52, 603–620.

Iryo, T., 2008. An analysis of instability in a departure time choice problem. Journal of Advanced Transportation 42, 333–356.

Iryo, T., 2011. Multiple equilibria in a dynamic traffic network. Transportation Research Part B 45, 867–879.

Iryo, T., 2013. Properties of dynamic user equilibrium solution: existence, uniqueness, stability, and robust solution methodology. Transportmetrica B: Transport Dynamics 1, 52–67.

Iryo, T., 2019. Instability of departure time choice problem: A case with replicator dynamics. Transportation Research Part B 126, 353–364.

Iryo, T., Watling, D., 2019. Properties of equilibria in transport problems with complex interactions between users. Transportation Research Part B 126, 87–114.

Safadi, Y., Haddad, J., 2021. Optimal combined traffic routing and signal control in simple road networks: an analytical solution. Transportmetrica A Transport Science 17, 308–339.

Sandholm, W.H., 2010. Population Games and Evolutionary Dynamics. The MIT Press.

Smith, M.J., 1979. Traffic control and route-choice; a simple example. Transportation Research Part B 13, 289–294.

Thunig, T., Nagel, K., 2019. Effects of user adaption on traffic-responsive signal control in agent-based transport simulations, in: 2019 6th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS), pp. 1–7.

Wardrop, J., 1952. Some theoretical aspects of road traffic research. Proceedings of the Institution of Civil Engineers 1, 325–362.

Webster, F.V., 1958. Traffic signal settings. Technical Report No 39. Road Research Laboratory.

Xie, X.F., Feng, Y., Smith, S.F., Head, K.L., 2014. Unified route choice framework: Specification and application to urban traffic control. Transportation research record 2466, 105–113.

Xie, X.F., Wang, Z.J., 2019. Combined traffic control and route choice optimization for traffic networks with disruptive changes. Transportmetrica B: Transport Dynamics 7, 814–833.

Yongqiang, F., 2020. Research on the development of a traffic signal control model based on route travel time equilibrium. IOP Conference Series: Earth and Environmental Science 587, 012045.

# Session 2

Estimation of time-dependent traffic volume using data-oriented dynamic traffic assignment
*Masaaki Ishihara, Takamasa Iryo, Kenji Ikeda and Shohei Yasuda*

Scalable Regional DTA with Meso-to-Macro Level Analysis and An Analytical and Simulation-Based Approach: Case Study of Northern Virginia
*Mohammad Abbasi, Feng Liu and Xuesong Zhou*

Bi-level Optimization model for DTA Flow and Speed Calibration
*Maryam Samaei, Mostafa Ameli, Jon Davis, Sean McQuade, Jonathan Lee, Benedetto Piccoli and Alexandre Bayen*

# Estimation of time-dependent traffic volume using data-oriented dynamic traffic assignment*

Masaaki Ishihara[a], Takamasa Iryo[a,*], Kenji Ikeda[b], and Shohei Yasuda[c]

[a]*Graduate School of Information Sciences, Tohoku University, Japan *Corresponding author: iryo@tohoku.ac.jp.*
[b]*West Nippon Expressway Company Limited*
[c]*Graduate School of Engineering, The University of Tokyo*

## 1 Introduction

The planning and operation of road networks require the analysis of traffic flows in congested road networks. As is well known, road congestion is a time-dependent phenomenon requiring an assessment using a dynamic traffic assignment problem.

The dynamic traffic assignment problem consists of a traffic flow theory and a drivers' behaviour model. In general, the dynamic user equilibrium (DUE) assignment problem is considered, which assumes that all drivers choose the route that minimises the generalised traffic cost, including the error term.

However, finding an equilibrium solution in a real-world network is not always easy. There are two main issues. First, the uniqueness and stability of DUE solutions are not always guaranteed, and no solution algorithm is yet known that can always find an equilibrium solution (Iryo, 2013). This problem can be avoided by extending the concept of equilibrium, e.g. by formulating the changes in drivers' behaviour and the state of the traffic flow in a Markov chain and obtaining their stationary distribution (e.g. Hazelton and Watling (2004) and Watling and Hazelton (2018)), but this requires an enormous amount of computation. Another major obstacle is that it is currently quite challenging to accurately estimate the parameters in traffic flow theory for all links. The difficulty in parameter estimation is partly owing to the availability of data. It is becoming possible to obtain travel times accurately from sample data such as probe-vehicle data. On the other hand, it is not easy to obtain traffic volume data and saturation flow rates accurately and comprehensively from such data at the moment.

This study proposes a methodology to calculate the time-dependent sectional traffic volume on major roads in the network with high accuracy using the data-oriented dynamic traffic assignment based on observed travel times to avoid the aforementioned problems. The main point of this methodology is that, instead of using a traffic flow model to calculate travel times, observed travel times obtained from probe-vehicle data are directly used. As the observed travel times are fixed in the observed values, there is no need to calculate the equilibrated travel times resulting from the interaction between the route choice model and the traffic flow model. This approach eliminates the need to estimate parameters such as traffic capacity in traffic flow theory models and avoids issues related to the properties of solutions and algorithms for calculating equilibrium solutions. A similar approach has also been proposed by Tsanakas et al. (2021).

The main issue of the proposed method is, of course, its inapplicability to networks for which no observed travel times can be obtained. It is also not suitable for forecasting situations where observed travel times are expected to change significantly, for example, due to major road constructions or significant changes in toll rates. On the other hand, such large-scale changes are infrequent in the real world. In practice, smaller changes in network structure and minor policy changes are more frequent. For such small-scale changes, the change in observed travel time is limited, and the approximation that *the observed travel time after the change is equal to that before the change* may be adopted. At

*Keywords:** travel time data, route choice model, dynamic traffic assignment

the very least, the error introduced by this approximation would be expected to be sufficiently smaller than the error associated with the estimation of traffic flow model parameters such as traffic capacity.

In order to improve the accuracy of the proposed method, it is necessary to introduce more accurate route choice models and calibrate the OD traffic volumes more precisely. In particular, regarding the route choice model, an all-or-nothing assignment cannot be expected to give accurate results, as the observed travel times are fixed. For this reason, stochastic route choice models must always be used. Time-dependent OD traffic volumes need to be calibrated, even if they are given by surveys such as trip surveys. To achieve highly accurate dynamic traffic assignment, the time-dependent OD traffic volumes must be calibrated dynamically, i.e. the traffic volume of each time period calculated by the dynamic traffic assignment must be adjusted with the observed time-dependent traffic volume.

We propose a data-oriented dynamic traffic assignment using observed travel times based on the stochastic route choice model and the calibration algorithm of the time-dependent OD traffic volumes, assuming the availability of the following data: (1) time-dependent link travel times acquired by the probe vehicles, (2) generated/attracted traffic volume by geographic grid acquired by the probe vehicles, (3) time-dependent sectional traffic volumes on selected links in the network, and (4) OD traffic volumes from trip surveys. The performance of the proposed methodology is assessed using real data in the Kansai region of Japan.

## 2 Overview of methodology and data

**Route choice model:** A probabilistic route choice model that enumerates routes is used. In order to limit the number of routes, an upper limit on the generalised travel costs of the routes is set. Furthermore, the major roads in the network to be analysed are extracted as the upper-layer network (e.g. Figure 1A), and only routes that satisfy one of the following conditions are enumerated:

1. routes not passing through the upper-layer network at least once (e.g. Figure 1B).
2. routes passing through the upper-layer network only once from the origin to the destination (e.g. Figure 1C).

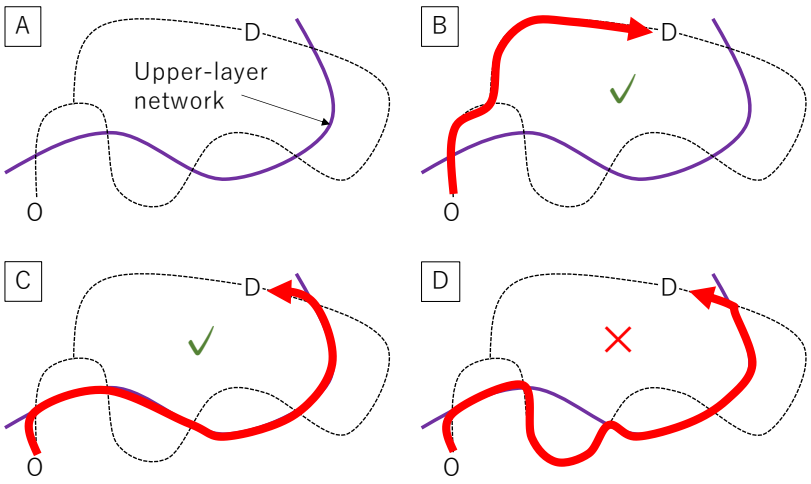



Figure 1: Examples of permitted / prohibited routes. **A**: example including the upper-layer network **B**: route not passing through the upper-layer network (permitted) **C**: route passing through the upper-layer network only once (permitted) **D**: Route not permitted

Condition 2 eliminates routes that include links in the upper-layer network and those not in the upper-layer network alternately (e.g. Figure 1D), thus significantly reducing the number of routes enumerated. The proposed method does not estimate the traffic volume on links that are not included in the upper-layer network. Therefore, route enumeration is performed only in the upper-layer network. Only the shortest paths are considered for the routes from the origin to the entry node of the upper-layer network and from the exit node of the upper-layer network to the destination node.

**OD traffic calibration algorithm:** The method of Bell (1991) is extended to be applicable to time-dependent traffic volumes. As the travel times are fixed in this study, the entries of the matrix for calculating the link traffic volume from the OD traffic volume can be made constant. Only the traffic volume of the links in the upper-layer network is used for the calibration.

**Observed travel times by probe vehicles:** Observed travel times on major roads obtained from ETC 2.0, the Japanese probe-vehicle system, are used. The sample rate depends on the road section and is several per cent. Not a sufficient amount of probe-vehicle data is available on all links. In this study, links with a certain amount of probe-vehicle data are aggregated using the method of Yasuda et al. (2019). The probe-vehicle data of the links are aggregated to obtain time-of-day travel times for the links.

**OD traffic data:** The statistics obtained from road traffic censuses conducted by the Japanese government are used. This OD traffic volume is estimated using data such as trip survey data.

**Generated and attracted traffic volume in geographic grids obtained by probe vehicles:** The geographical resolution of OD traffic volume data from road traffic censuses is low. Therefore, it is improved by using the generated and attracted traffic volume of probe vehicles aggregated in geographic grids of approximately 500x500 m in size. OD traffic volumes from probe vehicles are not used because the penetration rate of probe vehicles is low and not spatially homogeneous.

**Time-dependent sectional traffic volume data:** Time-dependent sectional traffic volume data from detectors available in the upper-layer network is used.

## 3 Case study

A case study is carried out for the major road network in the Kansai region of Japan. Figure 2 shows the overview of the network and the upper-layer network.

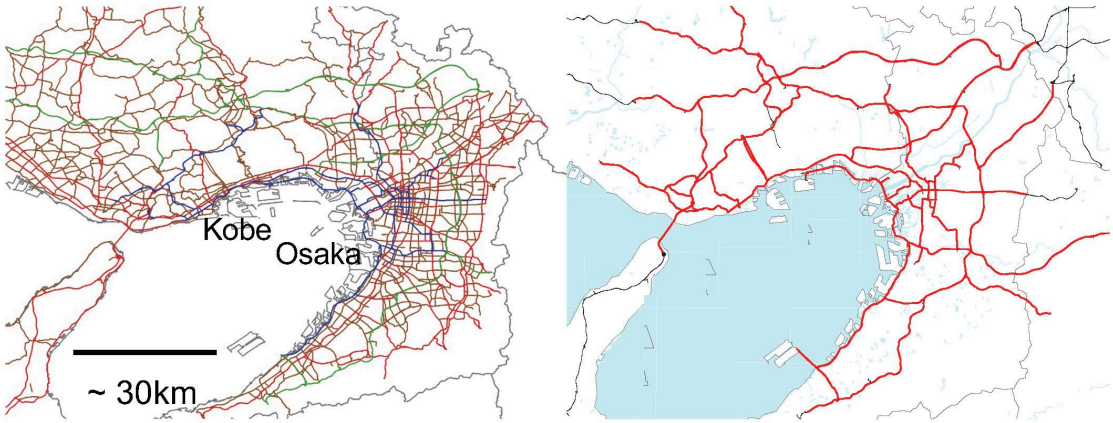



Figure 2: Case study network. Left: All links, Right: Upper-layer network (red links)

The time-dependent sectional traffic volume data in the case study is evaluated on the Hanshin Expressway, one of the motorways comprising the upper-layer network shown in Figure 2. The following two evaluations are carried out:

1. evaluation of the reproducibility of the current time-dependent sectional traffic volumes
2. evaluation of forecasted time-dependent sectional traffic volumes for the newly opened section.

The results of the case study will be shown at the presentation.

## Acknowledgements

This study was financially supported by Hanshin Expressway Company Limited. A part of the proposed methodology has been developed by the research project 'Proposal for the basic design of the next-generation ETC through the use and evaluation of ETC 2.0 data', which is a technical research and development project to improve the quality of road policy by the Ministry of Land, Infrastructure, Transport and Tourism of Japan.

## References

Bell, M.G.H., 1991. The estimation of origin-destination matrices by constrained generalised least squares. Transportation Research Part B 25, 13–22.

Hazelton, M.L., Watling, D.P., 2004. Computation of equilibrium distributions of markov traffic-assignment models. Transportation Science 38, 331–342.

Iryo, T., 2013. Properties of dynamic user equilibrium solution: existence, uniqueness, stability, and robust solution methodology. Transportmetrica B: Transport Dynamics 1, 52–67.

Tsanakas, N., Ekström, J., Gundlegrd, D., Olstam, J., Rydergren, C., 2021. Data-driven network loading. Transportmetrica B 9, 237–265.

Watling, D.P., Hazelton, M.L., 2018. Asymptotic approximations of transient behaviour for day-to-day traffic models. Transportation Research Part B 118, 90–105.

Yasuda, S., Iryo, T., Sakai, K., Fukushima, K., 2019. Data-oriented network aggregation for large-scale network analysis using probe-vehicle trajectories, 2019 IEEE Intelligent Transportation Systems Conference (ITSC). pp. 1677–1682.

# Scalable Regional DTA with Meso-to-Macro Level Analysis and An Analytical and Simulation-Based Approach: Case Study of Northern Virginia

**Mohammad Abbasi**
School of Sustainable Engineering and the Built Environment
Arizona State University, Tempe, AZ 85281, United States
Email: mohammad_abbasi@asu.edu

**Feng Liu**
Principal and Regional Manager
Cambridge Systematics, Inc.
Bethesda, MD, United States
Email: fliu@camsys.com

**Xuesong (Simon) Zhou**
School of Sustainable Engineering and the Built Environment
Arizona State University, Tempe, AZ 85281, United States
Email: xzhou74@asu.edu

**ABSTRACT**
This paper presents a scalable analytical and simulation-based Dynamic Traffic Assignment (DTA) model for the dynamic user equilibrium (DUE) problem. The model integrates a meso-to-macro framework and a column generation-based procedure, using a gap function as a nonlinear minimization problem (NMP). The meso-to-macro framework connects a macroscopic average travel delay function with a mesoscopic queueing-based vehicular flow model, while the column generation-based procedure uses a simulation-based dynamic network loading model to capture traffic dynamics and determine experienced path travel costs. The proposed DTA model is computationally efficient and analytically tractable, making it particularly useful for large-scale DTA applications.

**Keywords:** Dynamic Traffic Assignment (DTA), Scalable Regional DTA, Analytical DTA, Traffic simulation, Column generation algorithm, Dynamic user equilibrium

## 1. INTRODUCTION

Operational models that are used for large-scale or regional applications require more resources and data compared to travel demand models. In recent years, mesoscopic models have been used more frequently in research projects to bridge the gap between travel demand and microscopic simulation models. Mesoscopic models are useful for large-scale operational analysis because they can simplify the modeling of demand or supply, or the way they interact, depending on the approach used [1], [2]. Mesoscopic models and dynamic traffic assignment (DTA) models are often used interchangeably, but it is generally accepted that DTA models are a type of mesoscopic model that uses the dynamic user equilibrium (DUE) principle for route and/or departure time choice. In recent research projects, multi-resolution models have been used to provide analysis at different levels of resolution. Depending on the specific needs of the assignment, multi-resolution models can be two-tiered (e.g. meso/micro or macro/micro) or three-tiered (e.g. macro/meso/micro or macro/sub-area macro/micro). [3],[4]. When developing and using a DTA model, it is important to consider factors such as run time, data requirements, geographic scale, and temporal scale. There are two main types of DTA models: analytical-based DTA (A-DTA) and simulation-based DTA (S-DTA). [6].

In simulation-based DTA (S-DTA) models, the traffic assignment model uses the output from a simulation engine to generate and evaluate multiple paths in order to find a near-optimal solution to the equilibrium problem. Despite being widely studied in recent decades, applying S-DTA models to large-scale networks is still difficult and has therefore been limited. The use of S-DTA models typically requires careful calibration of many inputs and parameters, and if these inputs are not reliable, the results of the model may be less accurate than those from static models. [6], [7]. In analytical dynamic user equilibrium (DUE) models, link/node exit constraints are often used to determine traffic flow and continuous, strictly monotonic link performance functions are used to calculate path trip costs. While these models can produce solutions that satisfy DUE requirements, they may not accurately represent real traffic flow patterns because they need to sacrifice behavioral realism for theoretical elegance. For example, using the exit constraints necessary to achieve a first-in, first-out model would make the model less analytically tractable. [6].

This paper discusses two key components of regional dynamic traffic assignment (DTA) modeling: (1) a mesoscopic analytical DTA procedure that includes queue dynamics, and (2) a lightweight dynamic network loading simulator that uses a column path generation procedure and Newell's simplified kinematic wave model to adjust path flow volume. The proposed models were tested in real-world scenarios to assess their effectiveness and performance under different network and data conditions. The DTALite tool was also developed as an open-source, lightweight DTA package that is based on the General Modeling Network Specification (GMNS) format and allows for the efficient use of advanced dynamic traffic analysis capabilities. [8].

This paper is structured as follows: Section 2.1 discusses the requirements for implementing regional DTA, Section 2.2 provides network structure and user equilibrium of regional dta problem, and Section 2.3 presents the the macro to meso connection and analysis. In Section 2.4, calibration and validation of the modeling approach in a case study is presented. The paper concludes in Section 3.

## 2. SCALABLE REGIONAL DTA WITH MESO-TO-MACRO LEVEL ANALYSIS

### 2.1. Modeling Framework

In recent research, simulation-based approaches have been used to solve the user equilibrium dynamic traffic assignment (UEDTA) optimization problem. However, accurately modeling the time-dependent queue discharge rates and inflow patterns at the mesoscopic link level with a simulation-based DTA tool can be complex. The hourly capacity from a macroscopic model may be overestimated or underestimated if it is used as the default queue discharge rate. Calibrating

the queue discharge rate at the downstream location of each bottleneck and obtaining precise time-dependent inflow patterns at various incoming links upstream of a bottleneck (which is influenced by route choice behaviors and a time-varying origin-destination matrix) are important for accurately modeling the queue discharge rate internally through spatial queue propagation. [10].

Many analytical methods used to simulate UEDTA problems use traditional link performance functions to depict traffic congestion, which do not take into account the time-dependent nature of costs incurred in seemingly equivalent circumstances. These functions may not accurately reproduce dynamic traffic phenomena, may be violated under dynamic conditions of peak period flows, and do not address the First-In-First-Out (FIFO) requirement. Additionally, the path integral formulations used in these methods can be computationally intensive and require solutions to a system of simultaneous integral equations. [11].

The proposed DTA solution combines analytical and simulation-based approaches. The analytical-based (ADTA) portion involves establishing a meso-to-macro framework to connect a macroscopic average travel delay function with a mesoscopic queueing-based vehicular flow model. This cross-resolution approach can accurately capture congested bottlenecks at both the macro and meso scales. The meso-to-macro framework includes implementing a mesoscopic queue-based analysis, developing a queue-based volume delay function (QVDF), and calibrating the final model. The QVDF model is simple and computationally efficient, which makes it easier to model and analyze dynamic traffic systems more effectively. It can also measure both long-term average performance over multiple days/years and time-dependent performance measures within a single oversaturated period.

The simulation component of the proposed DTA solution is based on a lightweight dynamic network loading simulator that uses Newell's simplified kinematic wave model. This traffic simulation only requires a few traffic flow model parameters, such as outflow, inflow capacity, and storage capacity constraints. The model can accurately represent congestion propagation and capture shockwaves resulting from bottleneck capacities by explicitly using cumulative arrival and departure curves to track kinematic waves. It can also capture within-day traffic stream elements and queuing phenomena using improved traffic stream models with capacity constraints. To ensure internally consistent model calibration results, the DTA package should produce zone-by-zone time-dependent travel time skim and facilitate a streamlined feedback process between DTA and a 4-step process or agent-based model (ABM).

1 **TABLE 1 Symbols and definitions used in this study**

| Symbols | Definitions |
|---|---|
| $O$ | subset of origin nodes; $O \subseteq N$ |
| $D$ | subset of destination nodes; $D \subseteq N$ |
| $T$ | set of current ($t$) and ($\tau$) departure time intervals |
| $o$ | subscript for an origin node, $o \in O$ |
| $d$ | subscript for a destination node, $d \in D$ |
| $\tau$ | superscript for a departure time interval, $\tau \in T$ |
| $P(o, d, \tau)$ | set of all feasible paths for a given triplet ($o$, $d$, $\tau$) |
| $p$ | subscript for a path $p \in P(o, d, \tau)$ |
| $q_{od}^{\tau}$ | number of trips departing from node $o$ to node $d$ in time interval $\tau$ |
| $r_{odp}^{\tau}$ | number of trips departing from $o$ to $d$ in interval $\tau$ and assigned to path $p \in P(o, d, \tau)$ |
| $r$ | time-varying path flow vector, $\boldsymbol{r} = \{r_{odp}^{\tau}, \forall o \in O, d \in D, \tau \in T$, and $p \in P(o, d, \tau)\}$ |
| $\delta_{odp}^{\tau,ta}$ | time-dependent link-path incidence indicator, equal to 1 if vehicles departing from $o$ to $d$ in interval $\tau$ assigned to path $p$ are on link $a$ in interval $t$. |
| $tt_a^{\tau}$ | link travel time for the travelers traversing link a in interval $\tau$ |
| $c_{odp}^{\tau}(r)$ | path travel time for the travelers departing from $o$ to $d$ in interval $\tau$ and assigned to path $p \in P(o, d, \tau)$; $c_{odp}^{\tau}(r) = \sum_{a \in A} \sum_{t \in S} tt_a^{\tau} \delta_{odp}^{\tau,ta}$, and is a function of the time-varying path flow vector $\boldsymbol{r}$. |
| $\boldsymbol{c}(\boldsymbol{r})$ | vector of path travel times; $\boldsymbol{c}(\boldsymbol{r}) = \{c_{odp}^{\tau}(r), \forall o \in O, d \in D, \tau \in T$, and $p \in P(o, d, \tau)\}$ |
| $\pi_{od}^{\tau}$ | the least travel time from $o$ to $d$ in departure time interval $\tau$ |
| $\boldsymbol{\pi}$ | vector of the least travel times; $\boldsymbol{\pi} = \{\pi_{od}^{\tau}, \forall o \in O, d \in D$, and $\tau \in T\}$ |
| $t_0$ | start time of congestion period |
| $t_1$ | time index with maximum inflow rate |
| $t_2$ | time index with maximum queue length |
| $t_3$ | end time of congestion period |
| $\mu(a)$ | capacity (or discharge rate), assumed to be a constant value |
| $D(a)$ | total in-flow demand during the whole peak period |
| $C(a)$ | lane-based ultimate hourly capacity |
| $V(a)$ | total lane-based volume loaded on a road link during an analysis period (i.e., AM, MD, PM, or NT) |
| $v_{co}$ | cut-off speed |
| $v_{t2}$ | lowest speed on a link |
| $\gamma$ | inflow curvature parameter used in polynomial form |
| $\lambda(t, a)$ | inflow rate function at time $t$ |
| $Q(t, a)$ | queue length at time $t$ |
| $w(t, a)$ | traffic delay departing at time $t$ |
| $\alpha, \beta$ | parameters in the BPR-form link performance function |
| $f_d$ | constant in elasticity function for mapping D/C ratio to congestion duration |
| $n$ | elasticity coefficient of congestion duration in response to D/C changes, i.e., oversaturation-to-duration elasticity |

### 2.2. Network Structure and User Equilibrium DTA Regional Problem

In this study, it is assumed that OD demand and road capacity are deterministic variables. There are still two different kinds of travel time variability in this context: the first, which affects travel decisions in a predictable way across departure time intervals, and the second, which affects choices inside departure time intervals. In order to identify and bridge a gap between a fine resolution for describing precise traffic dynamics and a relatively coarse resolution in modeling traveler departure time choice behavior, this study takes into account travel time variability within a departure time interval (i.e. the second type). In particular, many discrete choice models quantify passenger behavior responses using 5-min or 15-min departure time intervals since a finer temporal resolution could lead to complex autocorrelation between alternative choices. In contrast, relatively shorter time intervals (such as 6 s in this work) are employed in traffic simulation to simulate timing variations in traffic signals and, more crucially, to capture dynamic traffic phenomena, including the development and dissipation of congestion

Consider a network G = (N, A), where N is a set of nodes indexed by i, and A is a set of directed links indexed by a = (i, i+1), i∈N and i+1∈N. The time period of interest (planning horizon) is discretized into a set of small-time intervals, T, indexed by t. Associated with each link is the time-varying link travel time $tt_a^{\tau}$ required to traverse link a when departing from node i in time interval τ ∈T. The following notation and variables are used in the paper.

The UEDTA problem of interest aims to find the time-varying traffic flow pattern that gives the minimal total system time in a transportation network. In this problem, it is assumed that travelers will behave cooperatively when choosing their routes to minimize the system's total travel time. The time-varying OD demands for the entire planning horizon (i.e.,$q_{od}^{\tau}$, ∀o, d, and τ) are assumed to be known a priori, and hence departure time choices are not addressed. It is also assumed that, for each (o, d, τ), all the trips departing at time τ from o to d have complete and accurate information about all the available paths connecting this OD pair and their characteristics. No en-route path-switching is allowed after departure from origins.

This study adopts the path based UEDTA formulation ([5],[6],[13]) which can be mathematically stated as follows.

$$\underset{r\in\Omega,\pi}{Min} TT(r,\pi) = min \sum_{o\in O}\sum_{d\in D}\sum_{\tau\in T}\sum_{p\in P(o,d,\tau)} r_{odp}^{\tau}[c_{odp}^{\tau}(r) - \pi_{od}^{\tau}] \quad (1a)$$

$$\text{Subject to } \sum_{p\in P(o,d,\tau)} r_{odp}^{\tau} = q_{od}^{\tau}, \forall o, d, \text{ and } \tau \quad (1b)$$

$$c_{odp}^{\tau}(r) - \pi_{od}^{\tau} \geq 0, \forall\ o, d, \tau, \text{ and } p\in P(o, d, \tau) \quad (1c)$$

$$r_{odp}^{\tau} \geq 0, \forall\ o, d, \tau, \text{ and } p\in P(o, d, \tau) \quad (1d)$$

Where Ω ≡ {r} is the set of feasible path flow vectors satisfying constraints 1(b) and 1(c). In this formulation, the decision variables are time-varying path flows, r. A measure of the violation of the DUE conditions is represented by *TT(r)* in terms of the difference between the total actual experienced path travel time and the total shortest path travel time calculated at any given time-dependent path flow pattern r ∈ Ω. When the time-varying path flow vector r∗ satisfies the DUE conditions, the difference vanishes; that is, there exists path flow vector r∗ ∈ Ω and π∗ such that TT(r∗, π∗) = 0.

### 2.3. Meso-To-Macro Level Analysis

In most existing UEDTA models, link flows are the decision variables. However, path-based routing policies are ideal because the controller in TMC needs to provide system optimal route

guidance to tripmakers. The problem with obtaining path flows from link-based formulations using link-path incidence relationships is that uniqueness is not guaranteed. Moreover, while link-path incidence relationships are relatively straightforward in the static case, these relationships are far from trivial in the time-dependent case. In the latter, vehicles assigned to a path at a given time are not simultaneously present on all links forming that particular path, and hence these relationships must recognize the time at which vehicles are actually present on a link. In the simulation-based solution to UEDTA, for a given path assignment r, a dynamic network loading (DNL) model is most commonly used to generate the resulting traffic flow pattern from which the average link travel times, average experienced path costs c(r), and thus time-dependent, or dynamic, link-path incidence relationships, $\delta_{odp}^{\tau,ta}$, can be extracted. However, link flows and link travel times are assumed to remain constant over the planning horizon of interest, typically the peak period in static traffic assignment models. Thereby, a link flow pattern that is intended to replicate the peak period flow results from a matrix of steady-state origin-destination trip rates which are assigned to the network links. Therefore, the link-path incidence, $\delta_{odp}^{a}$, statically relates path flows to link flows.

Our ADTA solution follows a two-phase process, where first, a route and a departure time interval, τ, $(i.e. 15\ minutes)$ from each origin are statically assigned to a traveler to generate mean link- and path-based travel times. Then, for each departure time interval τ, the time dependent vehicle inflow rate $\lambda(t,$ a), for each link, within period congestion is approximated using the fluid-based point queue model, which links the temporal queue evolution with the traffic performance measures (2a-2f). The results from the QVDF model feed the traveler path choice to regenerate the path flow assignment iteratively. We consider this link-path incidence relationship as the third category that relates path flows to link flows in a semi-dynamic manner, which is noted as $\delta_{odp}^{a,\tau}$.

Enumerating the entire set of feasible paths for every pair of ODs in a large scale road network is typically very difficult, if not impossible. In addition, only a (small) portion of the paths in the DUE solution would have positive flows, in which case the path travel times should be equal to the shortest travel times of each corresponding triplet (o,d, τ), and only the constraints in (1c) that relate to these used paths are binding. This study uses a column generation-based approach to construct a representative selection of paths with competitive trip durations, avoiding the explicit enumeration of all potential paths.

Considering a single bottleneck on some link, the time-dependent arrival rate $\lambda(t,\mathrm{a})$ can be approximated by a cubic polynomial function and the discharge rate $\mu(a)$ is assumed to remain constant across the relevant period. Similar to Newell's model, the difference between a given arrival rate $\lambda(t)$ and a discharge rate $\mu$ can be expressed in factored form as $\lambda(t) - \mu = \gamma(t - t_0)(t - t_2)(t - \bar{t})$, where $\bar{t}$ represent another root in addition to $t_0$ and $t_2$. The time period when speed is slower than $v_{co}$ is defined as congestion duration P. In PAQ models, the peak time from $t_0$ to $t_3$ where the arrival rate is higher than the average discharge rate, is also referred to as the congestion duration P (i.e., $P = t3 - t0$). According to the definition, $\mu$ should be less than the maximum capacity C. The total volume occurring during the congested period P with a speed slower than $v_{co}$ is considered as inflow demand, or queued demand, D. The ultimate hourly capacity in D/C is considered in this analysis as the hourly maximum flow rate per lane when the level of service is below E. The QDF is defined as the reciprocal to peak hour factor PHF used to convert the ultimate hourly capacity C to the period-based capacity $Cp$ for the entire assignment period (e.g. AM or PM): $\frac{D}{C} = \frac{V \cdot QDF}{C} = \frac{V}{C_P}$ [10].

The cross-resolution travel time performance model, which is intended to establish a connection between the continuous-time fluid-based polynomial arrival queue (PAQ) model within period congestion, and the system-wide VDFs over a wide range of traffic intensity conditions adopted by this study as follows:

$$V(a) = \sum_{o \in O} \sum_{d \in D} \sum_{\tau \in T} \sum_{p \in P(o,d,\tau)} r_{odp}^{\tau} \delta_{odp}^{a,\tau} \qquad (2a)$$

$$D(a) = V(a) \cdot QDF;\ D(a) \leq V(a) \text{ and } 0 \leq QDF \leq 1. \quad (2b)$$

$$P(a) = max\left[f_d\left(\frac{D(a)}{C(a)}\right)^n, \frac{D(a)}{C(a)}\right];\ P(a) = t3 - t0 \quad (2c)$$

$$\mu(a) = max\left[\frac{C(a)}{f_d P(a)^{(n-1)/n}}, \frac{D(a)}{C(a)}\right]. \quad (2d)$$

If the cubic PAQ model is adopted with $\frac{t_2-t_0}{t_3-t_0} = 0.5$, we will have the following time-dependent queue and delay:

$$w(t,a) = \frac{\gamma}{4\mu} \cdot (t-t_0)^2(t-t_3)^2. \quad (2e)$$

$$tt_a^t = w(t,a) + FFTT(a). \quad (2f)$$

The D/C ratio offers a queue-theoretic measure consistent with the congestion dynamics in comparison to the V/C ratio in the BPR function. Discharge rate and D/C can be connected, as illustrated in Fig. 1, such that D = P, where P is thought to represent an exogenous component in the QVDF model.

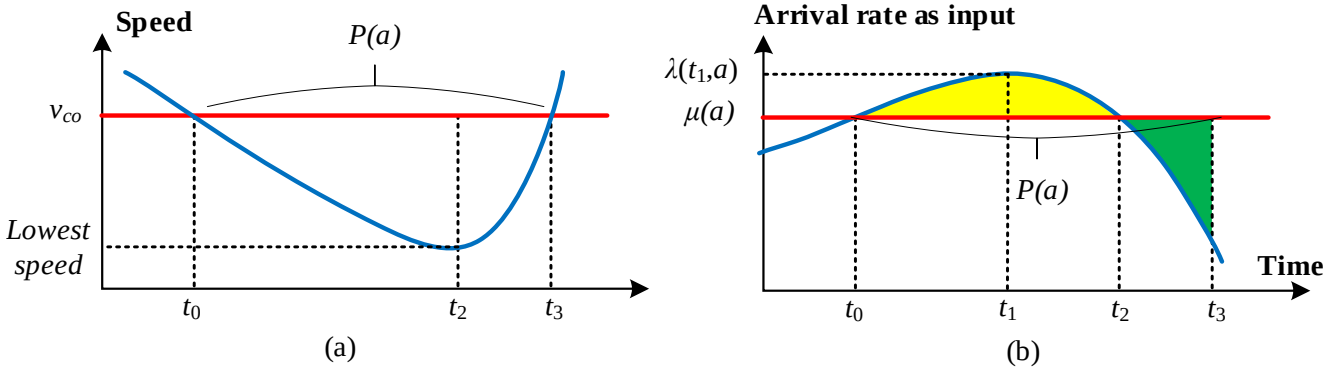



**Figure 1 The relationship among time-dependent speed, arrival rates, cut-off speed, and average discharge rate (adopted from [10].)**

Our Dynamic Traffic Assignment (DTA) modeling approach is a comprehensive method for analyzing traffic patterns and congestion in metropolitan areas and state agencies. It combines analytical and simulation-based techniques to provide a detailed and realistic representation of traffic flow and behavior. The approach is able to capture various aspects of traffic dynamics, including speed-density relationships, capacity constraints, wave propagation, and turning movements. It is also capable of modeling the impact of various transportation systems and strategies, such as Intelligent Transportation Systems (ITS) and traffic management measures. The DTA approach is flexible, allowing for analysis at various levels of resolution, from regional network analysis to more detailed simulation of specific corridors or freeways. It is also able to produce a range of performance measures and can be calibrated using observed data. Overall, the DTA approach is a valuable tool for supporting decision-making in transportation planning and strategy development.

### 2.4. Model Calibration and Validation

The DTALite model calibration and validation has been conducted in Northern Virginia. The main steps of the calibration and validation procedure is described as follows (**Figure 2**):

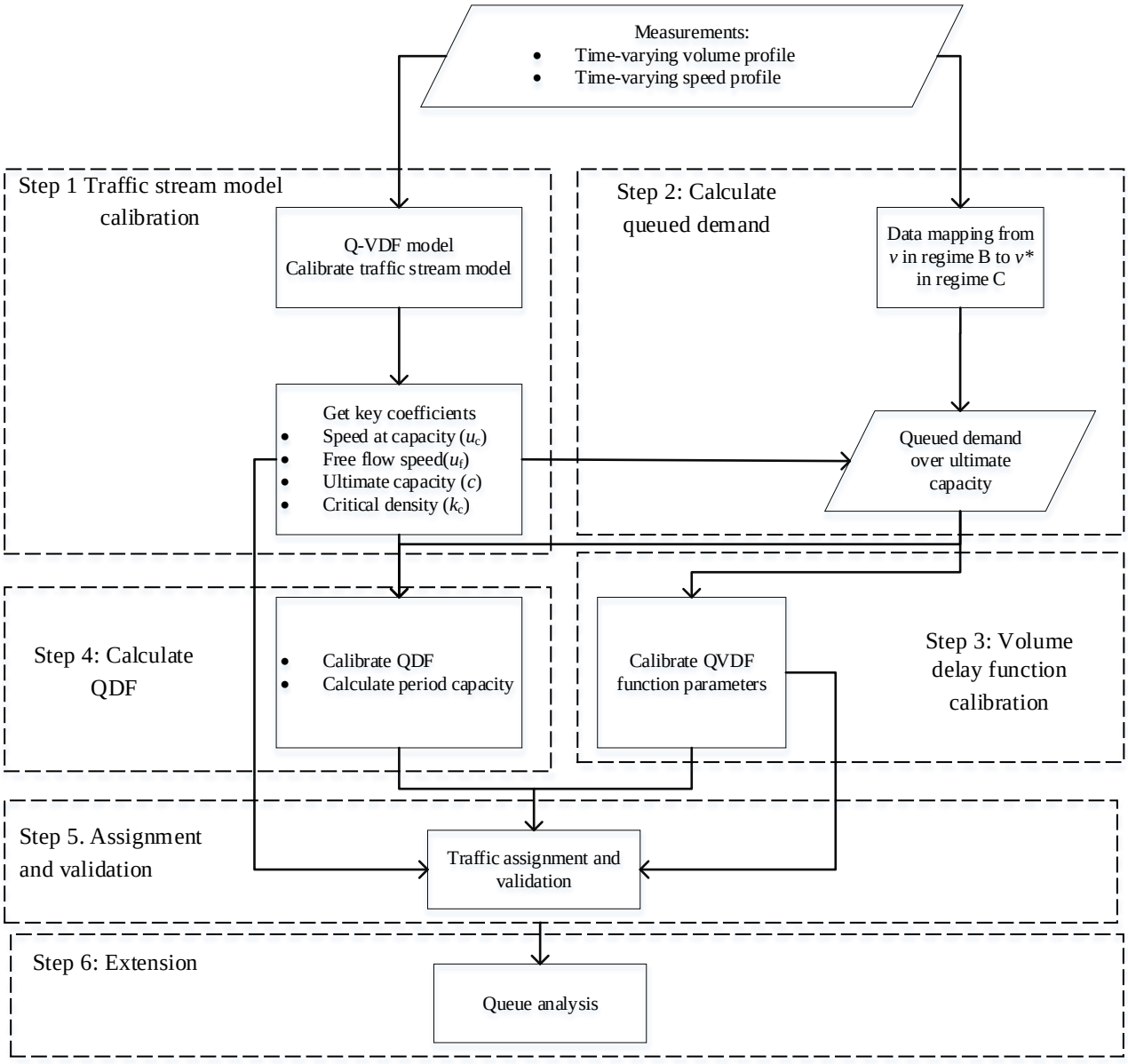



**Figure 2 model calibration framework (adopted from [14].)**

The DTALite model is a tool for analyzing traffic patterns and congestion in a regional area (Figure 1a). It uses a 3,722-zone system to divide the modeled area, which has a population of approximately 7.5 million people and approximately 4.3 million jobs. The model consists of 17,660 nodes and 49,230 links. The simulation process typically takes between 2 and 3 hours, with each iteration taking 5-6 minutes and requiring 20-30 iterations to converge. The model is run on a computer with 48 CPU cores and 80-90 GB of RAM for the agent-based simulation model and 20 GB for the analytical DTA model. The model is part of a case study that illustrates the use of the Generalized Macroscopic, Mesoscopic, and Microscopic Network Structure (GMNS) to represent traffic patterns at various levels of resolution and to enhance the calibration of traffic bottlenecks and signal timing using probe data. The GMNS format also allows for more consistent definition of bottleneck locations and intersection turning movements across different modeling resolutions (Figure 2a). The DTALite model is able to apply a cross-resolution approach, using the OSM2GMNS package to automatically build mesoscopic and lane-by-lane cell-based microscopic transportation networks as needed. The model's results are validated using traffic message channel (TMC)

data from RITIS/INRIX and by visually checking the results and comparing TMC distances to network distances (Figure 2b and Figure 3).

**Step 1-4: Meso-to-Macro Model Calibration**

There are several steps involved in calibrating a traffic flow model for large-scale DTA at the regional level. The first step is to use time-dependent speed and volume data to calibrate the critical parameters of a traffic fundamental diagram (FD) model. In this case, an S3 FD model was used, which has three parameters: free-flow speed, critical density, and a flatness-of-curve parameter[15]. The DTALite model can make use of large data sources, such as those provided by RITIS, to identify locations and patterns of recurring congestion in great detail. To accurately capture queue dynamics and local traffic conditions, it is necessary to identify and compare traffic congestions, which is done in the Congestions and Bottlenecks Identification (CBI) stage. Using data-driven approaches, time-dependent speed profiles are processed to calibrate the parameters of a QVDF model, including average speed and congestion duration, for the AM and PM periods. These recurring congestions are used to validate the DTA portion of the model on 11 priority corridors in Northern Virginia, which are further divided into segments for validation purposes (Table 2).

**Step 5: Link-level and Corridor-level DTA Model Validation**

To confirm the accuracy and reliability of the DTALite model, travel time and speed validation processes are carried out. These processes include comparing the corridor speed profiles of the model to observed data (Figure 4) and generating congestion heat maps based on spatio-temporal speed data (Figure 5). The accuracy of the model is determined using visualization tools and descriptive statistics, including Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Root-Mean Squared Error (RMSE) (Figure 6). To generate a congestion heat map, the average speed on each road segment is summarized for each minute, and the resulting map is compared to observed data to ensure that the model correctly captures bottleneck duration and location. The calibration and validation processes are iteratively repeated until the error rates and visualized results are satisfactory compared to the ground truth data. It should be noted that the use of heat maps for model validation requires a more complex procedure than the typical calibration and validation process, resulting in more detailed and accurate model performance. As shown in the figures, most corridor segments have absolute errors smaller than 5 mph, with few exceptions. The modeled speeds generally follow the patterns of observed data, as shown in the speed profiles for three selected locations along I-395 (Figure 7) and the speed heat maps for the northbound traffic along the I-395 corridor (Figure 8).

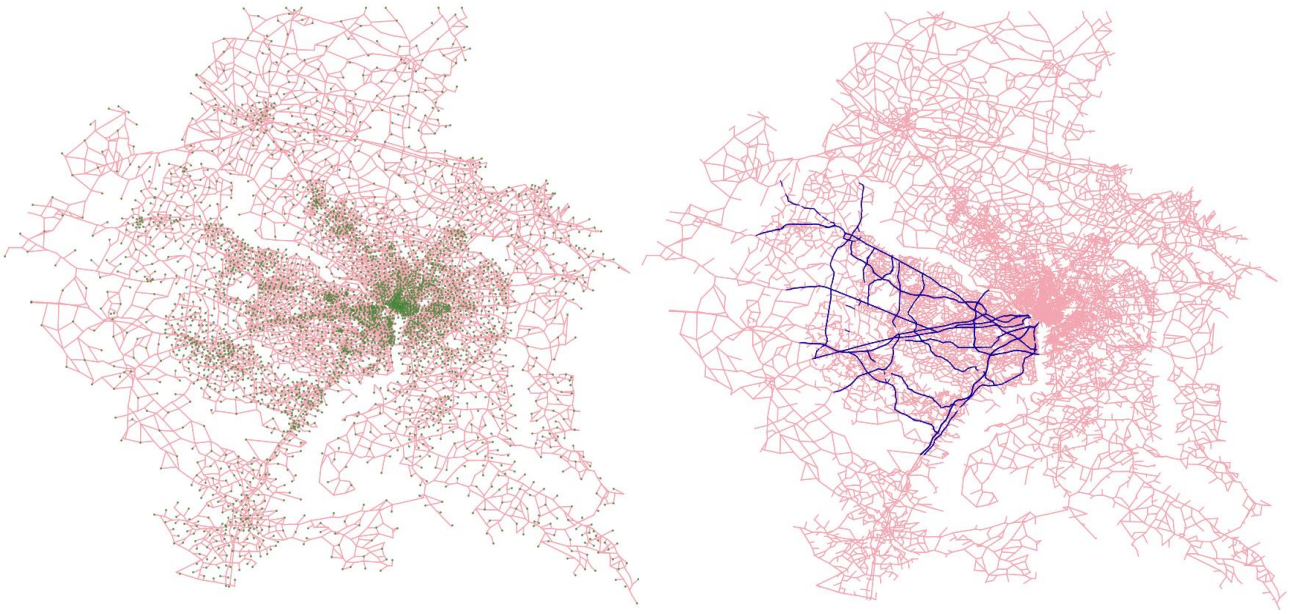


**Figure 2 a) Base Year Network; b) Links with Inrix speed data**

| B | C | G | H | K | L | O | P | Q | R | S |
|---|---|---|---|---|---|---|---|---|---|---|
| tmc | tmc_corridor_name | tmc_road | tmc_direct | from_node_id | to_node_id | FT | AT | nlanes | free_speed | tmc_reference_speed |
| 1310222695 | I-395 N-E-VA | I-395 N | E | 34588 | 36623 | 2 | 1 | 4 | 40 | 33.727146 |
| 1310222695 | I-395 N-E-VA | I-395 N | E | 36016 | 34711 | 2 | 1 | 4 | 40 | 33.727146 |
| 1310222695 | I-395 N-E-VA | I-395 N | E | 36623 | 36016 | 2 | 1 | 4 | 40 | 33.727146 |
| 388852440 | I-395 N-E-VA | I-395 N | E | 38341 | 38143 | 2 | 3 | 3 | 52 | 46.17429 |
| 1310272762 | I-395 N-E-VA | I-395 N | E | 36623 | 34588 | 2 | 1 | 4 | 40 | 39.468571 |
| 1310437210 | I-395 N-E-VA | I-395 N | E | 38143 | 38737 | 2 | 3 | 3 | 52 | 58.344997 |
| 1310437210 | I-395 N-E-VA | I-395 N | E | 38737 | 38172 | 2 | 3 | 3 | 52 | 58.344997 |
| 1310488504 | I-395 N-E-VA | I-395 N | E | 34588 | 36008 | 2 | 1 | 3 | 40 | 33.66357 |

**(a)** **(b)** **(c)**

**Figure 3 a) Features from INRIX TMC data; b) Features from planning model; c) Speed comparison**

**TABLE 2 the congestions and bottlenecks identification (CBI) stage**

| **Corridor Name** | **Speed (mph)** | | | **Congestion Duration (hour)** | | |
|---|---|---|---|---|---|---|
| | AM | PM | Mean | AM | PM | Mean |
| I-395 N-E-VA | 32.24 | 45.46 | 38.85 | 2.80 | 1.32 | 2.06 |
| ARLINGTON BLVD-E-VA | 31.15 | 32.91 | 32.03 | 2.24 | 1.61 | 1.92 |
| LEE JACKSON MEMORIAL HWY-E-VA | 38.31 | 35.78 | 37.05 | 1.06 | 2.10 | 1.58 |
| LEESBURG PIKE-E-VA | 33.50 | 30.14 | 31.82 | 0.71 | 2.30 | 1.51 |
| I-395 S-W-VA | 58.32 | 32.05 | 45.19 | | 2.70 | 1.35 |
| I-95 S-S-VA | 66.83 | 44.87 | 55.85 | | 2.37 | 1.19 |
| I-66 E-E-VA | 48.59 | 54.45 | 51.52 | 1.31 | 0.84 | 1.07 |
| LEESBURG PIKE-W-VA | 37.84 | 29.60 | 33.72 | 0.35 | 1.66 | 1.00 |
| I-66 W-W-VA | 62.24 | 48.80 | 55.52 | 0.04 | 1.61 | 0.83 |
| LEE HWY-E-VA | 34.70 | 32.47 | 33.58 | 0.69 | 0.75 | 0.72 |
| I-95 N-N-VA | 47.38 | 64.38 | 55.88 | 1.40 | 0.00 | 0.70 |
| RICHMOND HWY-N-VA | 36.88 | 36.43 | 36.65 | 0.38 | 0.96 | 0.67 |
| ARLINGTON BLVD-W-VA | 41.35 | 33.89 | 37.62 | 0.10 | 1.23 | 0.67 |
| VA-267 E-E-VA | 51.60 | 61.65 | 56.62 | 1.07 | 0.19 | 0.63 |

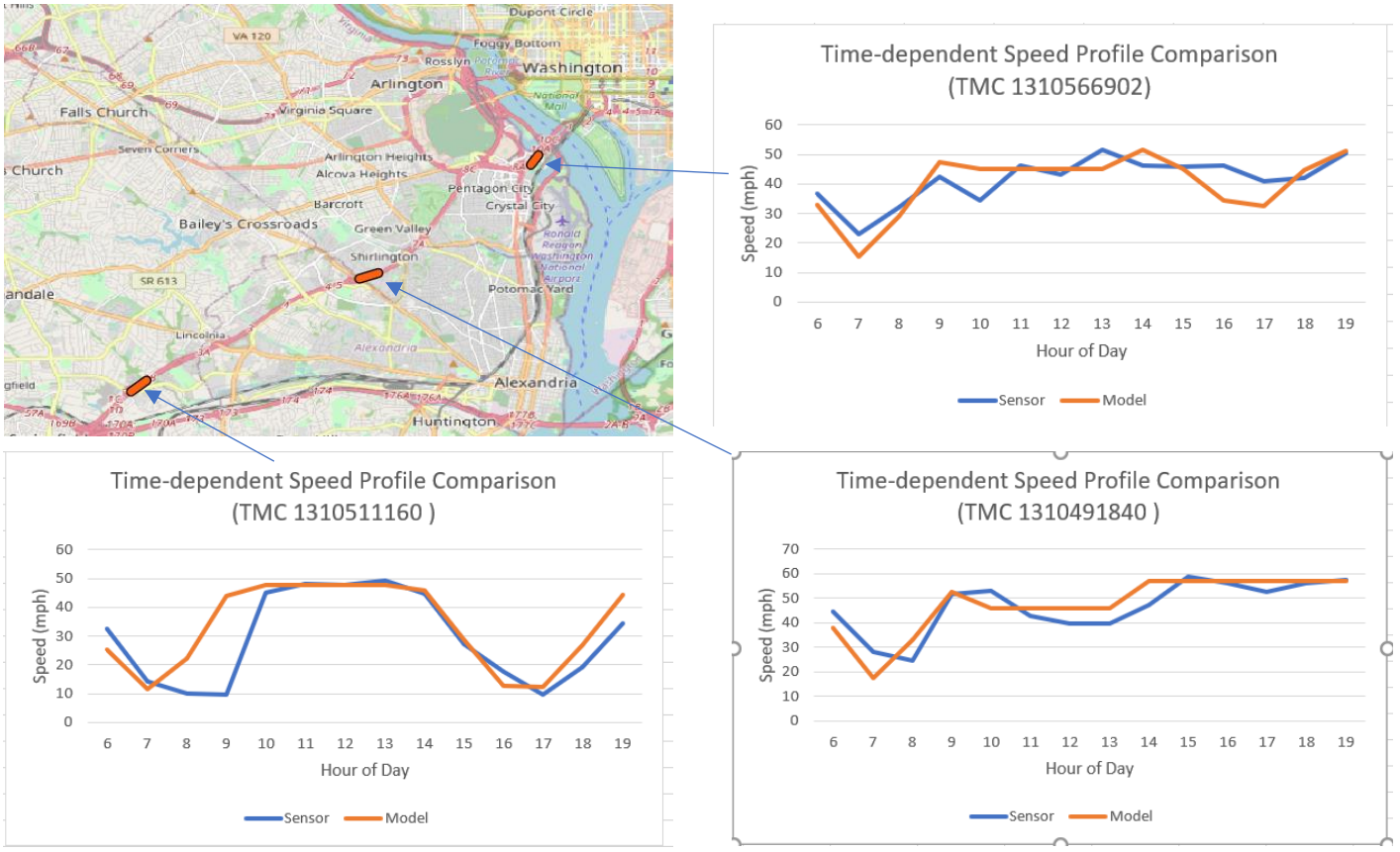



**Figure 4 Speed Profiles for Three Locations along I-395**

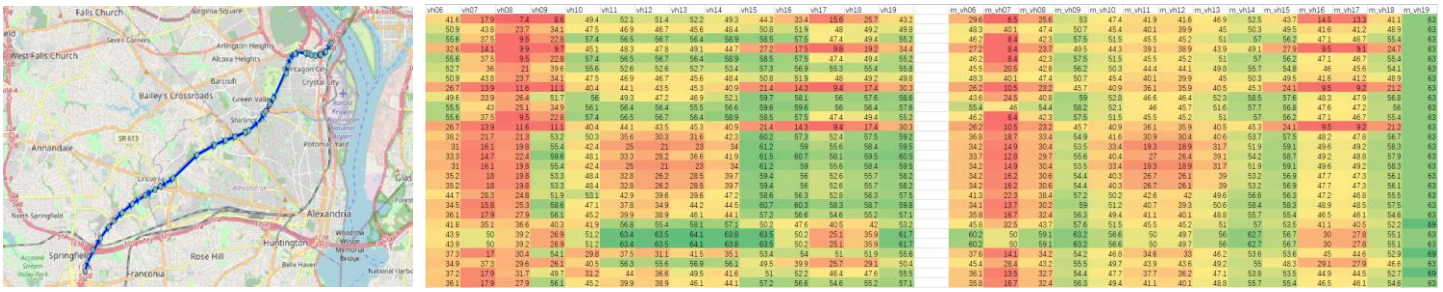


**Figure 5 Speed Heatmaps for Northbound Traffic along I-395**

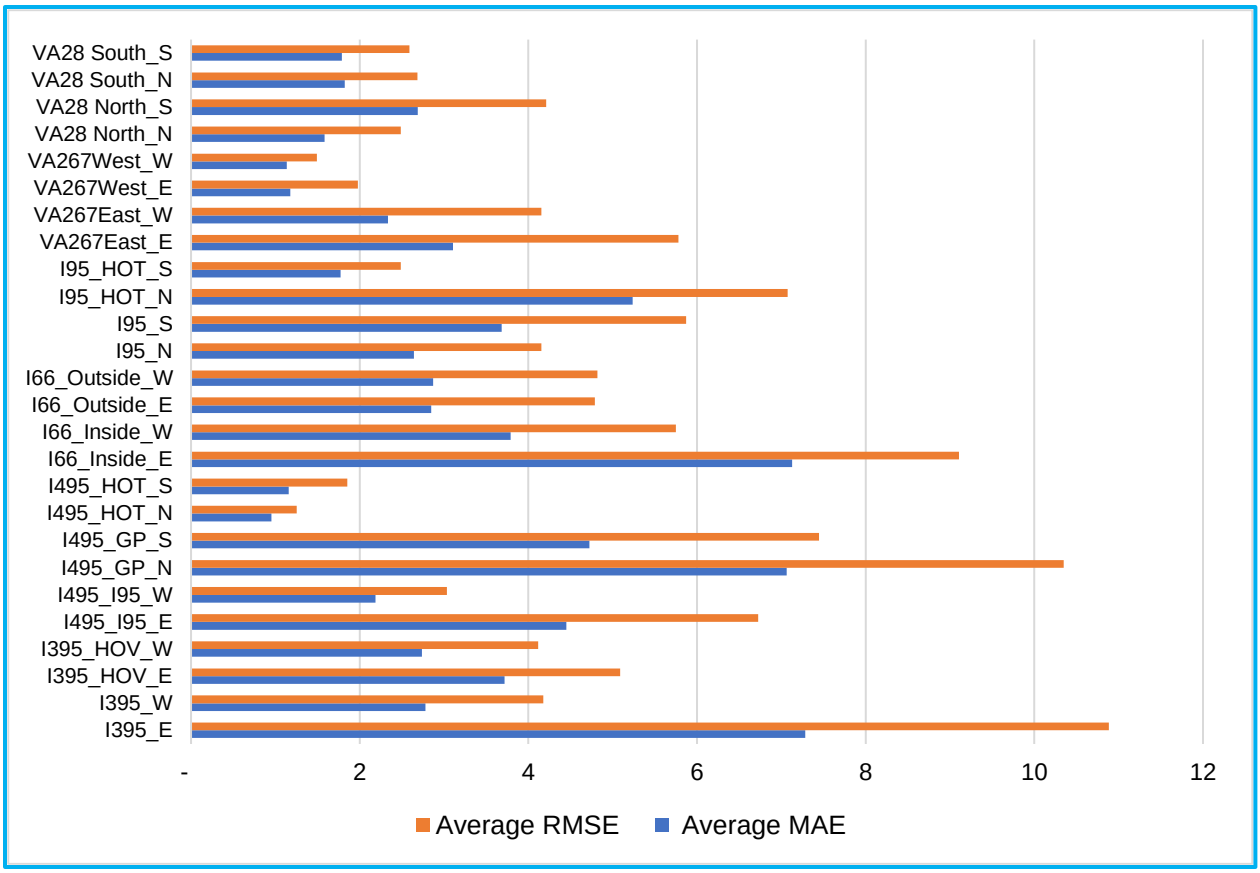



**Figure 6 Mean Absolute Error (MAE) and Root-Mean Squared Error (RMSE): Freeways**

## 3. CONCLUDING REMARKS

User equilibrium DTA models are used increasingly to describe and predict time-varying traffic network flow patterns, as well as to generate anticipatory and coordinated control and information supply strategies for intelligent traffic network management. The simulation-based approach has been successful at tackling many practical aspects that are essential in the application of DTA models in real networks, while the analytical approach contributes to theoretical insights about the problem and its solution. In an effort to improve the theoretical basis for the simulation based DTA approach, this paper addresses a series of critical and challenging issues such as scalability and data uncertainty in modeling and solving the large-scale UEDTA problem. This study proposed an integrated analytical-based and simulation-based DTA to approximate gap function as a nonlinear minimization problem. An efficient column generation-based framework was integrated into dynamic network loading to avoid explicitly enumerating all feasible paths. The adoption and integration of the above two methods, coupled with the embedded simulation-based dynamic traffic model, could enhance the development and deployment of large-scale simulation-based DTA models. Computational results on real large networks demonstrate that the proposed DUE algorithm is efficient and effective in obtaining close-to-DUE solutions.

The presented data preparation, calibration, and validation procedure is generic and thus can be applied to other large-scale networks. However, there are a number of limitations that needs to be addressed to enhance the DTA outputs further. One critical extension is that the inflow demand model could be extended to generate $P$ using departure time choices, and $\mu$ can be extended to incorporate dynamic patterns, i.e., $\mu(t)$ depending on the loading and unloading process of traffic.

**ACKNOWLEDGMENTS**

**AUTHOR CONTRIBUTIONS**

The authors confirm their contribution to the paper as follows: All authors contributed equally in all parts.

**REFERENCES**


1. Mahmassani, H. S. (2001). Dynamic network traffic assignment and simulation methodology for advanced system management applications. Networks and spatial economics, 1(3), 267-292.

2. Ziliaskopoulos, A. K., Waller, S. T., Li, Y., & Byram, M. (2004). Large-scale dynamic traffic assignment: Implementation issues and computational analysis. Journal of Transportation Engineering, 130(5), 585-593.

3. Zhou, X., Hadi, M., & Hale, D. K. (2021). Multiresolution Modeling for Traffic Analysis: State-of-Practice and Gap Analysis Report (No. FHWA-HRT-21-082). United States. Federal Highway Administration.

4. Hadi, M., Zhou, X., & Hale, D. (2022). Multiresolution Modeling for Traffic Analysis: Guidebook (No. FHWA-HRT-22-055). United States. Federal Highway Administration.

5. Zhou, X., Mahmassani, H. S., & Zhang, K. (2008). Dynamic micro-assignment modeling approach for integrated multimodal urban corridor management. Transportation Research Part C: Emerging Technologies, 16(2), 167-186.

6. Lu, C. C., Mahmassani, H. S., & Zhou, X. (2009). Equivalent gap function-based reformulation and solution algorithm for the dynamic user equilibrium problem. Transportation Research Part B: Methodological, 43(3), 345-364.

7. Shafiei, S., Gu, Z., & Saberi, M. (2018). Calibration and validation of a simulation-based dynamic traffic assignment model for a large-scale congested network. Simulation Modelling Practice and Theory, 86, 169-186.

8. Zhou, X., & Taylor, J. (2014). DTALite: A queue-based mesoscopic traffic simulator for fast model evaluation and calibration.

9. Smith, S., Berg, I., & Yang, C. (2020). General Modeling Network Specification: documentation, software, and data.

10. Zhou, X. S., Cheng, Q., Wu, X., Li, P., Belezamo, B., Lu, J., & Abbasi, M. A Meso-to-Macro Cross-Resolution Approach for Connecting Polynomial Arrival Queue Model to Volume-Delay Function with Inflow demand-to-Capacity Ratio.

11. Peeta, S. (1994). System optimal dynamic traffic assignment in congested networks with advanced information systems. The University of Texas at Austin.

12. Lu, C. C., Mahmassani, H. S., & Zhou, X. (2008). A bi-criterion dynamic user equilibrium traffic assignment model and solution algorithm for evaluating dynamic road pricing strategies. Transportation Research Part C: Emerging Technologies, 16(4), 371-389.

13. Peeta, S., & Mahmassani, H. S. (1995). System optimal and user equilibrium time-dependent traffic assignment in congested networks. Annals of Operations Research, 60(1), 81-113..

14. Wu, X. B., Dutta, A., Zhang, W., Zhu, H., Livshits, V., & Zhou, X. S. (2021). Characterization and Calibration of Volume-to-Capacity Ratio in Volume-Delay Functions on 2 Freeways Based on a Queue Analysis Approach. Transportation Research Record: Journal of the Transportation Research Board, No. TRBAM-21-04304.

15. Cheng, Q., Liu, Z., Lin, Y., & Zhou, X. S. (2021). An s-shaped three-parameter (S3) traffic stream model with consistent car following relationship. Transportation Research Part B: Methodological, 153, 246-271.

**Extended ABSTRACT for 9th International Symposium on Dynamic Traffic Assignment (DTA2023)**

# BI-LEVEL OPTIMIZATION MODEL FOR DTA FLOW AND SPEED CALIBRATION

Maryam Samaei[a], Mostafa Ameli[a, b, *], Jon F. Davis[b], Sean T. McQuade[c], Jonathan Lee[b], Benedetto Piccoli[c] and Alexandre Bayen[b]

[a] *University Gustave Eiffel, COSYS, GRETTIA, Paris, France*

[b] *Department of Civil and Environmental Engineering University of California, Berkeley, USA*

[c] *Department of Mathematical Sciences, Rutgers University–Camden, USA*

[*] *Corresponding author, Email: mostafa.ameli@univ-eiffel.fr*

**Keywords:** Agent-based simulation; dynamic traffic assignment; speed calibration; SUMO.

## 1 INTRODUCTION

Dynamic traffic assignment (DTA) models are capable of capturing traffic dynamics and are well-known as a critical tool in controlling and predicting the traffic situation. DTA simulation allows us to measure the results of deploying different technologies and applying different policies along with real experiments. One of the crucial steps to achieve realistic results from simulation tools is calibration. It aims to determine the DTA model's input such that the output represents traffic scenarios with a reliable level of accuracy (Antoniou, 2004). The inputs can be divided into two groups: demand and supply. The supply parameters define the environment of the simulation and the field constraints, e.g., traffic network topology and capacity, traffic signals, speed limitation, etc. In contrast, the demand inputs represent the travelers and their behavior in the system, e.g., time-dependent origin-destination matrix, routing, lane changing, etc. Besides, the simulation output is evaluated and validated based on the data collected from real operations of the transportation system. Many technologies, such as loop detectors, radars, and cameras, have been developed to measure and record traffic data. The main key indicators to evaluate the simulation output are flow and speed of the road segment or the whole network, which are collected at certain locations and times (Hou *et al.*, 2013). The mentioned technologies usually provide us with spatiotemporal data. Indeed, the rich time-varying traffic data can reflect field conditions for online real-time applications. However, the availability of such data for traffic modelers is restricted due to operational limitations and security reasons. This study aims to address the calibration problem for simulation-based DTA with multiple sources of aggregated data that are more likely to be available for road networks.

In this context, the calibration of agent-based DTA simulations is more complex than continuous (flow-based) models (Patwary *et al.*, 2021). Because the inputs of the model, in addition to the network supply, are the trip profile of each agent (particle), including the departure time, route, and driving model, which result in the agent's trajectory. In the state of the art of agent-based simulation, the calibration problem is defined mainly around flow or speed calibration. However, few studies address both problems. Motivated by a field experiment, called *MegaVanderTest* as a part of CIRCLES project [1] (Bayen, 2020), this study is an effort to create a high-fidelity simulation scenario using aggregated available data of the road segment flow and speed. A new methodology is proposed for flow and speed calibration of agent-based DTA simulation, wherein the mathematical model does not depend on the type of simulator. The data needed to feed the model is time-dependent road segment flow and speed (i.e., aggregated data, which is not hard to achieve). The next section presents a literature review on calibration problems for agent-based dynamic traffic simulation to clarify the problem and highlight the research question.

### Brief Literature review

For the literature review, we focus on the studies that have proposed a methodology to solve the DTA calibration problem for both flow and speed in an agent-based road simulation setting. Antoniou *et al.* (2011) has proposed a framework for offline and online flow and speed calibration. However, it has used a mono-objective function

[1] `https://circles-consortium.github.io/`

which may reduce the accuracy by adding scaling parameters for each data type. Djukic *et al.* (2017) has used a bi-level approach to solve the OD estimation problem, but it calibrates the network only according to the flow counts. (Hu *et al.*, 2017) consider a two-stage model to solve the calibration process according to the flow and speed data, while there is no iteration between the two stages to adjust the parameters accordingly.

The dynamic simulator determines the number of parameters to be calibrated, mainly for the supply side. Multiple dynamic simulators are developed for DTA models, e.g., SUMO, MATSIM, DYNASMART, and DYNAMIT. We use SUMO as an open-source simulator motivated by the project, and Langer *et al.* (2021). We deploy a simulation-based iterative approach to solve the DTA problem. Note that exact solution methods are computationally expensive or even impossible to apply due to the non-linear (non-analytical) and stochastic nature of the DTA calibration problem for spatiotemporal data (Lu *et al.*, 2015). Using meta-models is also recommended in the literature but mostly for large-scale multimodal networks (Osorio, 2019; Patwary *et al.*, 2021).

Regarding the data for calibrating flow and speed, we consider two data sources: loop detectors and Probe vehicles. The loop detectors count the number of agents in road segments, and Probe vehicles measure the average speed of the road segment for each time interval. The characteristics of the two data sets are not synced, meaning that the time interval, the data collectors' positions, and the devices' accuracy are not the same for the two data sets. This leads us to formulate a new bi-level optimization framework to iterate between two levels to calibrate the simulation scenario with respect to both data sets while considering the correlation between speed and flow results from the agents' route and departure time.

# 2 METHODOLOGY

Figure 1 presents the bi-level calibration framework proposed by this study. The process is started by importing the data of link flow and building the network graph. The initial link flow distribution is denoted by $\hat{X}$, and it is the set of time-dependent link flows collected by the loop detectors. $\hat{x}_i^k$ represents the reference flow of link $i$ ($i \in E$, set of all links) at time interval $k$, $\hat{x}_i^k \in \hat{X}$. $K$ is the set of time intervals for the flow data, $k \in K$. Note that for the speed data, we consider the same time horizon. However, the set of time intervals is different and denoted by $R$ and indexed by $r \in R$. The reason is that the duration of time intervals in $R$ and $K$ is given by the data set and is not necessarily equal for both sets. In this study, the flow data is collected every hour, and speed data is collected every minute. Thus the duration of $r$ is less than $k$. In the first (upper) level, the flow calibration problem is formulated as a mixed integer linear model and solved to determine the path flow distribution, $\Pi$. The flow of link $i$ at time $k$ resulting from $\Pi$ is denoted by $x_i^k$. The objective function of the upper level is to minimize the Squared Error (SE) between $x_i^k$ and $\hat{x}_i^k$. The output of this step is the Origin-Destination (OD) matrix and the number of agents traveling on paths between each OD pair, $\pi_{OD}, \forall \pi_{OD} \in \Pi$. We fixed the path flow distribution for the second (lower) level, wherein we determined the departure time distribution of agents.

To find the optimal departure time, we propose a modified version of the simultaneous perturbation stochastic approximation (SPSA) algorithm inspired from Lu *et al.* (2015). For the initial solution in Step 2, we consider the uniform distribution for the departure times. In every iteration of the SPSA algorithm, we run a simulation (Step

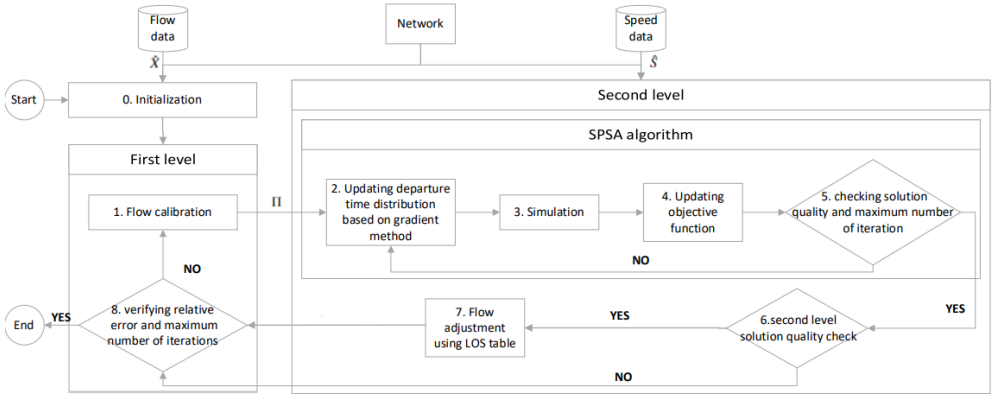


Figure 1: Methodology process

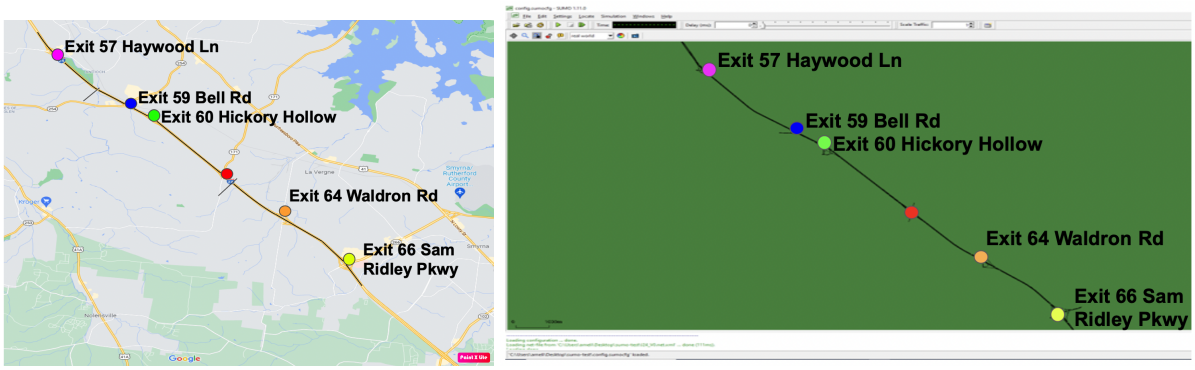
3) with the demand profile from the previous step. In this study, we run SUMO simulation with speed variable sign (Lopez *et al.*, 2018) come from the data set. We consider a variance around the speed sign for the speed choice of the agents; therefore, they can violate this sign if they have the opportunity to gain speed. Note that the variable speed sign is not lane-dedicated and does not impact the lane-changing behavior of agents. In Step 3, We also measure $s_l^r$, the average speed of road segment $l$ at time intervals $r$ and compare it to the corresponding value $\hat{s}_l^r$ from data set $\hat{S}$ collected by the Probe vehicles. Then in step 4, the objective function of the lower level is updated based on the simulation results.

The next step is checking the convergence conditions of the SPSA algorithm based on the maximum number of iterations and comparing the solution quality (SE of the speed values) with a threshold. If the convergence is not achieved, we go to Step 2. Otherwise, we perform the second level solution quality check in Step 6 by considering the speed absolute error: $|x_i^k - \hat{x}_i^k|$. If there is any link with more than 10% absolute error, we go to the next step; otherwise, we go to Step 8. In Step 7, we aim to address the temporal correlations between two data sets. We modify the values in $\hat{X}$ w.r.t the level of service (LOS) of the targeted network. LOS gives us a level of flow for the targeted speed considering the characteristic of the road, e.g., number of lanes and road type (Prassas *et al.*, 2020). For example, if the speed measured by simulation is higher than the data, we increase the value of the target flow. As a result, the density will increase at $k$ then we can expect that the speed will reduce at $r$. This modification results in additional errors for the upper level. Therefore, in Step 8, we check the relative error of flow and speed in addition to the convergence of the model by the maximum number of upper-level iterations. In other words, if the modification of the flow is minor and the speed error is acceptable, we converge; otherwise, we go to Step 1 to update the path flow distribution.

We deploy the Intelligent driver model (IDM) in this study. The calibration of IDM parameters is carried out based on the characteristics of the test case, e.g., driving laws and culture in the location of the test case. In our study, we address this issue by conducting multiple sensitivity analyses on the parameter of the simulator driving model. Then we set the parameters based on the available data and expert comments. The detailed mathematical model of the proposed framework (Figure 1) and further information about the sensitivity analysis will be presented at the conference.

# 3 NUMERICAL EXPERIMENT

The proposed methodology is applied to a road network of a portion of I24 highway in Nashville, Tennessee, the U.S., presented in Figure 2. The road network consists of 154 nodes, 452 links, and 16 traffic signals. The flow data set, LOS data and the cycle time of signals are provided by the Tennessee Department of Transportation (TDOT), and the speed data is provided by INRIX company. TDOT flow data corresponds to the network links, while the INRIX data provide the speed for every road segment. One link can consist of more than one road segment, and there is no overlap between road segments and links. The data sets represent the morning peak hour (6:30 am - 8:30 am). 20 min simulation warm-up is considered to insert the agents with their optimal departure times.

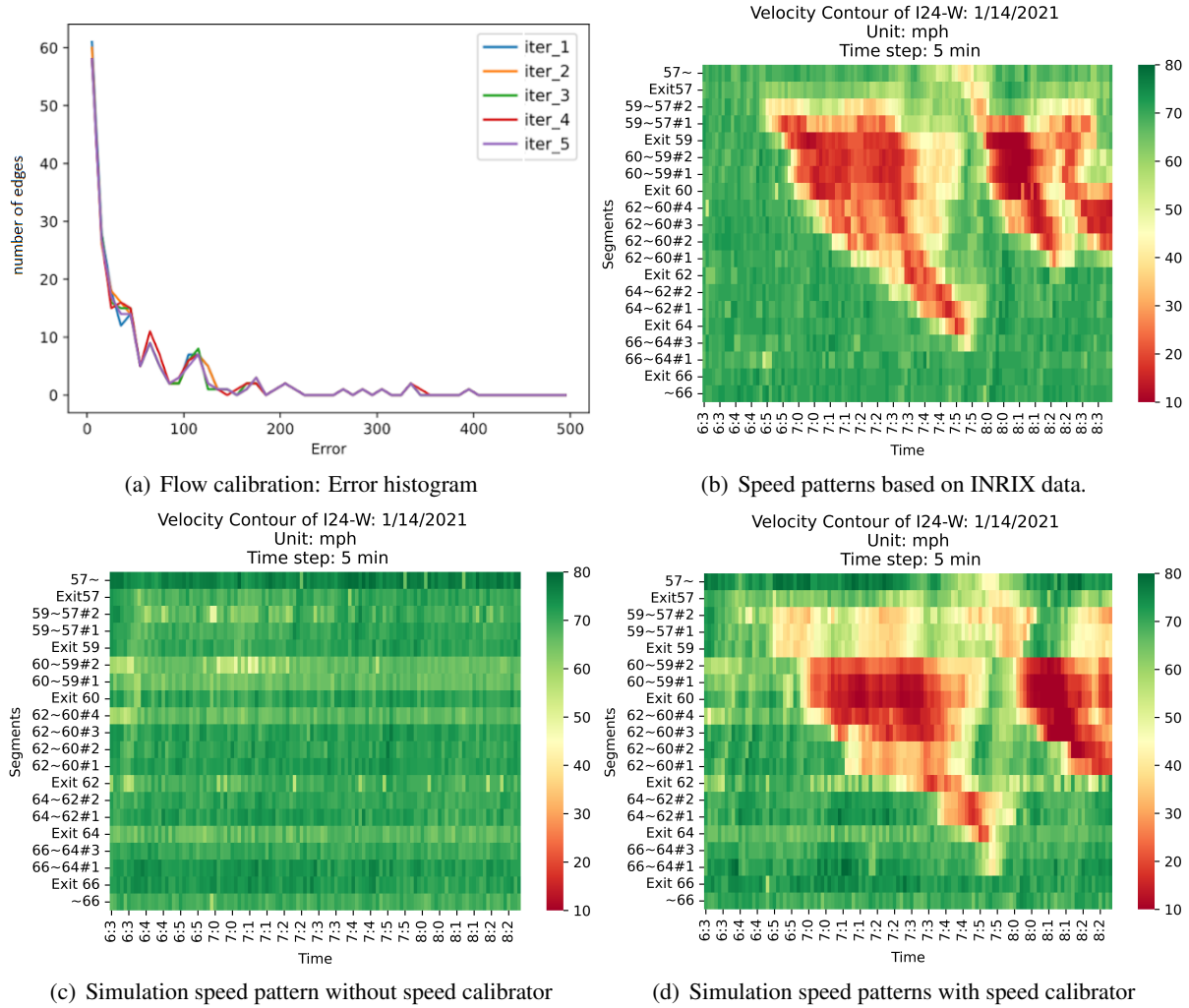



(a) I24 mapping data ©Google 2022 (b) SUMO network.

Figure 2: I24 road network.

# 4 PRIMARY RESULTS AND DISCUSSION

We first execute the bi-level process sequentially to evaluate the performance of each level. Figure 3 presents the primary results of flow and speed calibration. For the first level, we calculate the solution of the MILP model with the Scipy optimization solver. Figure 3(a) presents the error histogram of the objective function for the flow calibration when we run the optimizer five times. The results show that the optimizer performs consistently. Next, we use calibrated demand profile with uniform departure time distribution to run the simulation-based speed calibration.




(a) Flow calibration: Error histogram

(b) Speed patterns based on INRIX data.

(c) Simulation speed pattern without speed calibrator

(d) Simulation speed patterns with speed calibrator

Figure 3: Primary results

Figure 3(b)-(d) illustrate the spatiotemporal speed pattern of the road network. Figure 3(b) corresponds to the INRIX data for a real scenario, which shows a traffic jam starting from Exit 59 at 7:00 am. Figure 3(c) presents the execution of the simulation with uniform departure time distribution and no speed calibration. We can observe that this scenario does not reflect any traffic jams during the peak hour. Figure 3(d) illustrates that using speed calibrator with simulation shows a similar pattern to the INRIX data. Stop-and-go waves are observed in the simulation. However, in some segments, the maximum absolute error is high in Figure 3(d) compared to the real data (Figure 3(b)), particularly for Exit 59 at the end of the simulation. Recall that the two optimization levels are conducted sequentially for these results. We are currently coupling the two levels in order to execute the iterative bi-level framework, which seems promising according to the primary results.

## References


Antoniou, Constantinos. 2004. *On-line calibration for dynamic traffic assignment*. Ph.D. thesis, Massachusetts Institute of Technology.

Antoniou, Constantinos, Balakrishna, Ramachandran, Koutsopoulos, Haris N, & Ben-Akiva, Moshe. 2011. Calibration methods for simulation-based dynamic traffic assignment systems. *International Journal of Modelling and Simulation*, **31**(3), 227–233.

Bayen, Alexandre M. 2020. CIRCLES: Congestion Impacts Reduction via CAV-in-the-loop Lagrangian Energy. *University of California, Berkeley*.

Djukic, Tamara, Masip, David, Breen, Martijn, & Casas, Jordi. 2017. Modified bi-level optimization framework for dynamic OD demand estimation in the congested networks. *Pages 27–29 of: Proceedings of the Australasian Transport Research Forum, Auckland, New Zealand*.

Hou, Tian, Mahmassani, Hani S, Alfelor, Roemer M, Kim, Jiwon, & Saberi, Meead. 2013. Calibration of traffic flow models under adverse weather and application in mesoscopic network simulation. *Transportation research record*, **2391**(1), 92–104.

Hu, Xianbiao, Chiu, Yi-Chang, Villalobos, Jorge A, & Nava, Eric. 2017. A sequential decomposition framework and method for calibrating dynamic origin—destination demand in a congested network. *IEEE Transactions on intelligent transportation systems*, **18**(10), 2790–2797.

Langer, Marcel, Harth, Michael, Preitschaft, Lena, Kates, Ronald, & Bogenberger, Klaus. 2021. Calibration and Assessment of Urban Microscopic Traffic Simulation as an Environment for Testing of Automated Driving. *Pages 3210–3216 of: 2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*. IEEE.

Lopez, Pablo Alvarez, Behrisch, Michael, Bieker-Walz, Laura, Erdmann, Jakob, Flötteröd, Yun-Pang, Hilbrich, Robert, Lücken, Leonhard, Rummel, Johannes, Wagner, Peter, & Wießner, Evamarie. 2018. Microscopic traffic simulation using sumo. *Pages 2575–2582 of: 2018 21st international conference on intelligent transportation systems (ITSC)*. IEEE.

Lu, Lu, Xu, Yan, Antoniou, Constantinos, & Ben-Akiva, Moshe. 2015. An enhanced SPSA algorithm for the calibration of Dynamic Traffic Assignment models. *Transportation Research Part C: Emerging Technologies*, **51**, 149–166.

Osorio, Carolina. 2019. Dynamic origin-destination matrix calibration for large-scale network simulators. *Transportation Research Part C: Emerging Technologies*, **98**, 186–206.

Patwary, Ashraf Uz Zaman, Huang, Wei, & Lo, Hong K. 2021. Metamodel-based calibration of large-scale multimodal microscopic traffic simulation. *Transportation Research Part C: Emerging Technologies*, **124**, 102859.

Prassas, Elena S, P. Roess, Roger, Prassas, Elena S, & P. Roess, Roger. 2020. Concepts of Capacity and Level of Service for Interrupted Flow. *The Highway Capacity Manual: A Conceptual and Research History Volume 2: Signalized and Unsignalized Intersections*, 37–49.

# Session 3

Departure time models with separable supply
*Stephen Boyles and Hillel Bar-Gera*

Simultaneous-swap Algorithms for Departure Time Choice
*Jeroen Verstraete and Chris Tampere*

Influence of Simulation Timestep Size and Pathfinding Interval Length on Path Patterns and Computational Cost in Dynamic Traffic Assignment
*Siwei Hu, Pengyuan Sun, Daisik Nam, R. Jayakrishnan and Michael Hyland*

# Departure time models with separable supply

Stephen D. Boyles* Hillel Bar-Gera†

January 16, 2023

## 1 Introduction

Two of the most influential models in transportation planning are Vickrey's (1969) single-bottleneck model, and the static traffic assignment model formulated by Beckmann et al. (1956). The single-bottleneck model focuses on the temporal nature of transportation systems, representing queue dynamics and the departure time choice decision. Static traffic assignment, by contrast, focuses on these systems' spatial nature, representing congestion in different locations and the route choice decision. Subsequent research has explored these themes in deeper and more sophisticated ways. For instance, the field of dynamic traffic assignment aims to bring both the spatial and temporal aspects of transportation systems under a single modeling umbrella. Nevertheless, the single-bottleneck and static traffic assignment models remain notable for their simplicity and clarity, respectively highlighting the key temporal and spatial features of transportation systems and supply/demand interactions.

One key difference between the two models is that congestion is *separable* in the classical static traffic assignment problem: the travel time on a link depends only on its own flow, and not the flow on any other link. By contrast, the single-bottleneck model is not, and in fact exhibits asymmetric interactions: travelers departing at a given time affect the travel time of those departing later, but not those departing earlier, as causality demands. This simple distinction has several significant implications. First, the "price of anarchy" is higher in the single-bottleneck model — a traveler affects many others who cannot affect them in return, whereas in separable traffic assignment models travelers can only cause congestion on the links they are traveling on (and therefore experiencing that congestion themselves). Second, the asymmetry of interactions complicates the process of numerically computing equilibria in the bottleneck model — the resulting cost mapping is not monotone, and some natural dynamic processes do not converge to equilibrium (although the equilibrium can often be constructed analytically).

The natural question is whether the distinction of "separable vs. non-separable" is fundamentally linked to the "temporal vs. spatial" distinction between the bottleneck and traffic assignment models, or whether there is room for meaningful models which are *temporal, but separable*, or *spatial, but non-separable*. There has been quite a bit of research investigating the latter category, formulating traffic assignment models with link interactions of various sorts. However, to our knowledge, the former category has not received much attention. There are certain transportation systems which can be well-represented by a separable temporal model, such as a traffic signal operating in undersaturated conditions. Demand is less than capacity, so all vehicles arriving during a red indication can depart during the next green, and queues do not accumulate from one cycle to the next. Nevertheless, the delay increases with demand within each cycle as the queue length on red grows. Imposing schedule delays as in the single-bottleneck model, one can naturally ask what an equilibrium departure time profile would look like in such a system. The single-bottleneck model would

*Civil, Architectural, and Environmental Engineering, The University of Texas at Austin
†Industrial Engineering and Management, Ben-Gurion University of the Negev

represent such an equilibrium in *oversaturated* conditions (where a queue persists between cycles), but would fail to capture the proper traffic behavior in the undersaturated case, perhaps around the "shoulder" of the peak period.

The goal of this paper, therefore, is to formulate and analyze a departure time choice model with a separable cost function. Like the single-bottleneck model, our focus is solely on the temporal aspects of the system; the models we propose in the current paper do not have any spatial dimension. We have several aims in studying such a model. First, as described above, there are certain traffic systems that can be well-represented by a separable temporal model. Second, the resulting model is simple, comparable in complexity with the single-bottleneck model. This clearly highlights the effect of the "separable vs. non-separable" distinction by comparison with the single-bottleneck model, and the "temporal vs. spatial" effect by comparison with the static traffic assignment model. Third, we also provide a "hybrid" model yielding the classical single-bottleneck model and our purely separable model as special cases. This latter model is intended to represent the same system as it transitions between undersaturated and oversaturated conditions (that is, between transient and persistent queues), highlighting the differences in the equilibrium departure time profiles during these periods. Obviously, the simple model presented here also offers options for further investigations.

## 2 Formulation

Consider a peak period within a given day. We assume that delays are separable, in that the delay $d$ experienced by travelers who arrive at a particular point in time $t$ is entirely determined by the rate of travelers also arriving at that point in time, but none other. That is,

$$d(t) = D(f(t)) \,, \tag{1}$$

for some function $D$. The analysis in this abstract will use a linear form for $D$ for convenience. The full paper will also explore some nonlinear functions numerically. As in Vickrey's model, we assume that there is also a scheduling cost $s(t)$ for arriving at time $t$, consisting of penalties for arriving earlier or later than the desired arrival time $t^*$. Here we will use the standard piecewise-linear schedule delay function, leading to a generalized cost function

$$g(t) = D(f(t)) + \beta[-t]^+ + \gamma[t]^+ \,. \tag{2}$$

(choosing units so that the value of time is 1 and the desired arrival time is 0). In total there are $N$ travelers, and the user equilibrium occurs when they are distributed among the time intervals $t$ such that every arrival time $t$ with positive $f(t)$ has equal and minimum generalized cost $g(t)$. The system optimum occurs when the $N$ travelers are distributed over time intervals such that total generalized cost $\int_{\mathbb{R}} f(t)g(t)\ dt$ is minimized.

A discretized version of the model can be formulated and solved as an instance of the traffic assignment problem with a single origin and destination connected with a set of parallel links indexed by time interval; the link performance functions are given by (2), and the total demand is $N$ vehicles. This is useful for nonlinear versions of the model, and will be discussed more in the full paper.

## 3 Equilibrium and analysis with linear delay

Assume a linear delay function of the form

$$D(f) = d_0 + af \,, \tag{3}$$

where $d_0$ is the free-flow delay, $f$ is the rate at which travelers are arriving at a specific instant in time, and $a$ is a non-negative constant reflecting congestibility. In this case, the user equilibrium and system optimum flow profiles can be derived analytically, and the full paper includes such a derivation. The key feature is

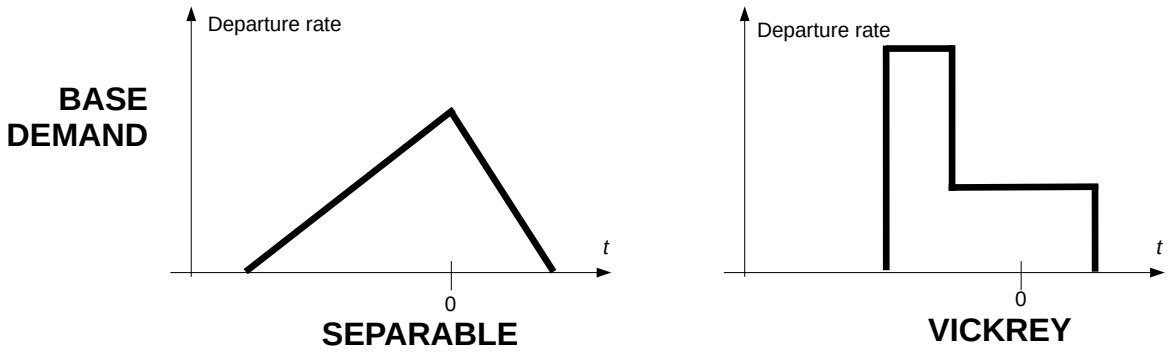



Figure 1: Equilibrium departure rates in the separable model and in Vickrey's model.

Table 1: Comparing nonseparable and separable departure time choice equilibria.

| | Separable (with linear $D$) | Nonseparable (Vickrey) |
|---|---|---|
| First arrival $t_s$ | $-\sqrt{\frac{2aN}{\beta+\gamma}\frac{\gamma}{\beta}}$ | $-\frac{\gamma}{\beta+\gamma}\frac{N}{u}$ |
| Last arrival $t_e$ | $\sqrt{\frac{2aN}{\beta+\gamma}\frac{\beta}{\gamma}}$ | $\frac{\beta}{\beta+\gamma}\frac{N}{u}$ |
| Equilibrium cost $c$ | $\sqrt{2aN\frac{\beta\gamma}{\beta+\gamma}}$ | $\frac{N}{u}\frac{\beta\gamma}{\beta+\gamma}$ |
| System optimal arrival window | Longer by $\sqrt{2}$ | Same as UE |
| Equilibrium total cost $TC_{UE}$ | $\sqrt{2aN^3\frac{\beta\gamma}{\beta+\gamma}}$ | $\frac{N^2}{u}\frac{\beta\gamma}{\beta+\gamma}$ |
| Optimum total cost $TC_{SO}$ | $\frac{4}{3}\sqrt{aN^3\frac{\beta\gamma}{\beta+\gamma}}$ | $\frac{N^2}{2u}\frac{\beta\gamma}{\beta+\gamma}$ |
| Price of anarchy | $\frac{3\sqrt{2}}{4}\approx 1.061$ | 2 |

that both of these flow profiles are *triangular* (Figure 1), with departure rates linearly increasing from zero at an initial time $t_s$, reaching a peak at time 0, and then linearly decreasing to zero at a final time $t_e$. This contrasts sharply with Vickrey's model, which predicts piecewise constant departure rates: one rate greater than capacity for travelers arriving early, and a second rate below capacity for travelers arriving late. In our model, departure rates and the length of the departure window are both proportional to $\sqrt{N}$, whereas in Vickrey's model the departure rates are unaffected by the total demand, and the length of the departure window is directly proportional to $N$. The main consequence of this is a much lower price of anarchy, which we compute analytically in the case of linear costs, and evaluate numerically (in the full paper) in the case of nonlinear costs. For linear costs, the ratio of user equilibrium and system optimum costs is $3\sqrt{2}/4 \approx 1.061$ (independent of $N$, $a$, $\beta$, and $\gamma$), much smaller than the ratio of 2 in Vickrey's model, and even smaller than the 4/3 ratio in static traffic assignment with linear costs. Table 1 compares key features of the equilibria in in our separable model an the Vickrey model, with $u$ the capacity in the latter.

# 4 Extensions

The full paper also includes discussion of two variants of this basic model:

**A discrete version of the above continuous model:** Discretizing the above model has two potential advantages. The first is computational, since our model reduces to a special case of the static traffic assignment problem. This formulation can accommodate nonlinear cost functions, and may provide a technique for extending the model in other ways (such as elastic demand, or stochasticity in demand or supply). The second is from a modeling perspective. The main physical system we intend to represent in our model is a traffic signal operating in undersaturated conditions, where queues clear fully in each cycle of the light (and thus delays in one cycle are independent of those in any other), but where expected delay within each cycle increases with volume. The continuous-time model is an approximation to what is in reality a discrete system. Furthermore, the time discretization is given directly by the cycle length and cannot be set arbitrarily by the modeler. We discuss the properties of such a discrete-time model, showing that the continuous and discrete versions yield similar results even with a relatively small number of discrete time intervals, and exploring the effects of nonlinear delay functions on the price of anarchy.

**A hybrid model representing both over- and undersaturated conditions:** Our model presumes separable delays, and no queue accumulating over an extended period of time. This is a reasonable description of a traffic signal where inflows are below capacity. If inflows exceed capacity, the accumulating queue in the Vickrey model is a better representation. We formulate a "hybrid" version of the model encompassing both of these, formulating a system with separable delays when departure rates are small, and with a (nonseparable) accumulating queue when departure rates are large. We show that in such systems, there are at least two equilibria. The classic Vickrey equilibrium is one of them; the other has separable conditions for at least some of the departure window. We show that the second equilibrium always has lower cost than the first.

# References

Beckmann, M. J., C. B. McGuire, and C. B. Winston (1956). *Studies in the Economics of Transportation.* New Haven, CT: Yale University Press.

Vickrey, W. S. (1969). Congestion theory and transport investment. *American Economic Review 59*(2), 251–261.

# Simultaneous-swap Algorithms for Departure Time Choice

Jeroen Verstraete, Chris Tampère
KU Leuven

Abstract

In this small paper, a new family of simultaneous-swap algorithms for the departure time choice problem is introduced. In this conference paper, we focus mainly on the concept of this family for a single class case. In the full paper, exact formulas and/or pseudo code for each submodule will be given and the family of algorithm will be extended to multi-class cases, which further increases the potential of this new family of algorithms. The uniqueness of this family comes from considering swaps simultaneously where the inflicted cost changes induced by swaps are taken into account for each alternative instead of only the alternatives involved in the swap. A first numerical example will illustrate the potential of this new method.

# 1. Introduction to Departure Time Choice

Congestion, and therefore travel time, is time dependent. Starting from a low level, the congestion level increases in the first part of the rush hours, reaches a peak or plateau and decreases in the last part. This whole process repeats, in general, in morning and evening peaks. Just like travellers try to optimize their route to their destination to minimize their perceived cost, so will travellers optimize the time at which they depart from their origin. The cost a traveller perceives for departing at a certain time depends on the travel time (this includes a variable queueing time) but also on their cost for arriving at their destination too early or too late compared to their preferred arrival time; the latter is called the schedule delay cost. Knowing that travellers will take these costs into account in their decisions, traffic models should be able to predict this behaviour when analysing future scenarios; especially those scenarios where changing departure time is among the first reactions of travellers, e.g., a study about road pricing with a smart toll that depends on the congestion level (Ubbels and Verhoef, 2005). The module in a Dynamic Traffic Assignment (DTA) model that is responsible for predicting this behaviour is called the departure time choice model.

The main contribution of this paper is the presentation of a new family of algorithms that solves the single bottleneck problem efficiently. The unique feature that characterizes this family is that in the optimization for searching the new departure rate, not only are all the selected time intervals calculated together, the influence of a change from the departure rate of one time interval to all other time intervals is taken into account. An algorithm of this new family consists of different submodules, each having their own specific tasks and properties. After the presentation of the generic new family of algorithms, one algorithm of this family is discussed and used on a case studies to illustrate the gains and potential of the new generic family. Knowing that these algorithms will need to scale to a whole network to be practical, for all our implementations no analytical information nor unknown network parameters will be used. All the presented implementations, with the exception of the network loading, are fully generic, meaning that they do not rely on detailed insights into delay dynamics or the cost structure (Daly and Bar-Gera, 2021).

## 1.1. Vickrey Model

Vickrey (Vickrey, 1969) was the first to model the departure time choice equilibrium problem for a single bottleneck. To understand the problem and the numerical method developed in this paper, the Vickrey model is briefly summarized here. Consider a network of one road with a single bottleneck. The model considers a fixed number of travellers $(N)$ that need to choose their departure time given the fixed bottleneck capacity $(Q_{max})$ and their preferred arrival time $(t^*)$. The cost travellers are assumed to experience at time $t$ can be written as:

$$c_t = \begin{cases} (t^* - t - TT_t) * \beta + TT_t * \alpha & \forall\, t + TT_t < t^* \\ (t + TT_t - t^*) * \gamma + TT_t * \alpha & \forall\, t + TT_t \geq t^* \end{cases} \quad (1)$$

With:

$$TT_t = T_0 + W_t \quad (2)$$

With $TT_t$ the travel time departing at time $t$, $T_0$ the free flow travel time and $W_t$ the waiting time at the bottleneck at time $t$. The parameters $\alpha, \beta, \gamma$ represents the weighting factor for the cost of traveling, being too early and being too late, respectively. This problem has an unique departure flow solution where each traveller has the same cost (Daganzo, 1985; Smith, 1984).

## 1.2. State-Of-The-Art Solution Algorithms

Since Vickrey first introduced this problem, much additional research has been published on this problem. Only the most relevant literature will be discussed here, for a more complete overview, we refer to existing reviews, e.g., (Small, 2015).

The problem as introduced by Vickrey is in continuous time. (Han et al., 2016) proves that an equilibrium exists considering both the departure time choice and the route choice in a dynamic traffic assignment in continuous time. The generalized Vickrey model (Han et al., 2013) allows working both in continuous and discrete time. A numeric method is given in (Han et al., 2013), formulated as a mixed integer linear program. Other methods, e.g., (Ramadurai et al., 2010), discretize time in time intervals.

(Ramadurai et al., 2010) show that the deterministic formulation can be transformed into a variational inequality. One could solve the problem with generic variational inequality methods. The analogy to route choice is large (see also next section), this is probably the reason why many methods for the departure time choice are inspired by route choice algorithms. An example is the proportional swap algorithm (Guo et al., 2017; Smith, 1984), which swaps travellers from higher-cost alternatives to lower-cost alternatives (the alternatives being, depending on the problem considered, either routes, departure times, or combinations thereof). Based on the same principle and evolution theory, some methods use replicator dynamics (Iryo, 2019; Schuster and Sigmund, 1983). It has been shown that these algorithms not always converge to a stable point (Daly and Bar-Gera, 2021; Guo et al., 2018; Iryo, 2008), meaning that no fixed point can be found. A possible reason for this is explained in the next section. Due to the bad convergence of these algorithms, some authors (Guo et al., 2017) redefine the equilibrium conditions of the model. Instead of requiring the exact same cost over all departure times, the bounded rational user equilibrium is satisfied when all costs are within a predefined indifference bound. In this context, there is no longer a single equilibrium departure flow solution, as many departure rates satisfy this condition. More recently, (Daly and Bar-Gera, 2021) claim that the maximum-to-minimum shift with an adaptive shift size always converges to any desired approximation of the equilibrium.

In summary, the state of the art on solution algorithms for Vickrey's departure time choice model offers solutions that converge effectively but not efficiently to the analytical solution (even for a single bottleneck) or scalable solutions for full networks that compromise convergence by allowing indifference bands around the analytical solution or sampling noise. Efficiently converging algorithms that are extendable to scale to full networks are desired.

## 1.3. Analogy to Route Choice

The route choice problem and the departure time choice problem are closely related to each other. In a deterministic equilibrium, both problems should respect Wardrop's first principle (Wardrop, 1952): the cost for all chosen alternatives (be it routes or departure times) should be equal while non-chosen alternatives have a higher (or at best, equal) cost. Both problems can be divided into two sub-problems. Where the first sub-problem consists of finding the set of alternatives that will be used and has a combinatorial character, the second sub-problem finds the flow distribution over this set and characterizes as a search/optimization in a continuous flow distribution space.

As discussed in the previous section, many state-of-the-art departure choice algorithms are inspired by route choice algorithms. However, there is an important difference between the two problems: the so-called separability of their cost function. The cost function for a route is dominantly separable, meaning that the main determinant of the route cost is the flow on the links of that route and flow on other links has little or no influence. In contrast, the cost for departing at a certain departure time is strongly non-separable, meaning that departure flows in other (notably: earlier) time intervals may have equal or even higher influence than that of the flows departing in the time interval considered. Mathematically, this means that the Jacobian $(\frac{\delta C}{\delta f})$ for the route choice problem is dominant on the diagonal while that of the departure time choice problem has asymmetric, dominant non-diagonal elements. Some insights into the structure for the latter will be given in the next subsection. According to (Dafermos, 1980), problems with a symmetrical Jacobian or an asymmetrical Jacobian with dominant diagonal elements converge faster than an asymmetrical Jacobian with non-dominant diagonal elements; if the latter converge at all. The reason can be easily understood. Many solution algorithms are based on the notion of swapping, where flow is reduced on alternatives with higher costs (assuming that this will reduce cost) and increased on lower-cost alternatives (assuming this will increase cost). Both effects contribute to equalizing costs towards Wardrop equilibrium. Non-separability, however, challenges these assumptions: costs may not decrease despite flow reduction; flow increase may not cause cost increase; and last but not least: alternatives not involved in the swap may see their cost changed as well, which may unintentionally (and in most algorithms that disregard non-separability: uncontrollably) push them away from equilibrium.

The main goal of this paper is to develop a numerical algorithm that uses all sensitivities in the Jacobian, both separable and non-separable components, such that a much more efficient swap is proposed.

# 2. Non-Separability

In this section, the structure of the Jacobian or sensitivity matrix $(S)$ is discussed. As in the rest of the paper (and most of the literature on departure time choice), we have chosen to discretize time in departure time intervals. Sensitivity $S_{a,b}$ denotes the increase in cost for departure time interval $b$ for an extra traveller in departure time interval $a$. Let us consider a typical rush with one congestion

period around the desired arrival time (time interval 7 in this example). The capacity of the bottleneck ($Q_{max}$) is set to 100 vehicles per time interval. The departures and arrivals per time interval are sketched in Figure 1.

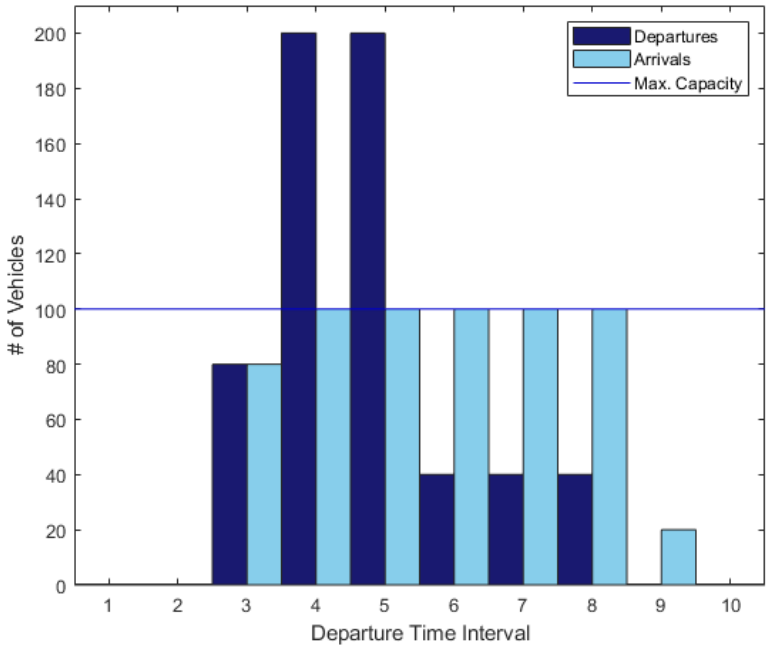



Figure 1 Departures and Arrivals per Time interval of a Simple Rush

Let us now investigate the structure of the sensitivity matrix, which for this example is given in Table 1. Note that, as recommended by (Blumberg-Nitzani and Bar-Gera, 2014; Heydecker and Verlander, 1999), the cost of the last traveller in a departure time interval is taken as the cost of that interval. The sensitivity matrix is an upper triangular matrix, meaning that departures during an interval have no influence on the cost of earlier intervals. There is a difference between the rows where the corresponding time interval has a queue and where it doesn't have a queue. If there is no queue (rows 1, 2, 3, 9 and 10), the partial derivatives are zero. Note that this assumes the travel time to only be a function of the queue length, otherwise, only the partial derivative of the travel time function would be found. If there is a queue (rows 4 to 8), there will be an influence over all later time intervals that still have the same queue. In this example, the influence lasts until time interval 8 as in time interval 9, the queue is gone. The magnitude of the influence is depending on the desired arrival time. $S_{4,4}$ and $S_{4,5}$ have different values because currently departure time interval $4$ arrives too early while time interval $5$ arrives too late. If there would be an extra traveller in departure time interval $4$, the travellers in interval $4$ will now experience more delay but arrive more on time, while travellers in interval $5$ will now be even later, on top of the extra delay. If, as in the original Vickrey model, the travel time is only a function of the queue length, the whole sensitivity matrix contains only four different values: (i) extra delay for travellers who were and still are too early, (ii) extra delay for travellers who were too early but now are too late, (iii) extra delay for travellers who were and still are too late and (iv) no extra delay as there is no influence. Due to discretization, in this simple example, case (ii) does not appear. In the continuous case or other discrete examples, some departing time interval can experience (ii).

Table 1 Sensitivity Matrix of a Simple Example

| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 3 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | 0 | 0 | 0 | $\frac{\alpha-\beta}{Q_{max}}$ | $\frac{\alpha+\gamma}{Q_{max}}$ | $\frac{\alpha+\gamma}{Q_{max}}$ | $\frac{\alpha+\gamma}{Q_{max}}$ | $\frac{\alpha+\gamma}{Q_{max}}$ | 0 | 0 |
| 5 | 0 | 0 | 0 | 0 | $\frac{\alpha+\gamma}{Q_{max}}$ | $\frac{\alpha+\gamma}{Q_{max}}$ | $\frac{\alpha+\gamma}{Q_{max}}$ | $\frac{\alpha+\gamma}{Q_{max}}$ | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | $\frac{\alpha+\gamma}{Q_{max}}$ | $\frac{\alpha+\gamma}{Q_{max}}$ | $\frac{\alpha+\gamma}{Q_{max}}$ | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | $\frac{\alpha+\gamma}{Q_{max}}$ | $\frac{\alpha+\gamma}{Q_{max}}$ | 0 | 0 |
| 8 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | $\frac{\alpha+\gamma}{Q_{max}}$ | 0 | 0 |
| 9 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

The discussed sensitivity matrix is conditional on the current departure flows and expresses a local sensitivity to infinitesimally small flow increments. If flow increments increase, the sensitivity values in the matrix are correct until some linearity condition is broken, meaning that at some point the increase in cost will no longer be linear with the given flow increment. One of the linearity conditions is when capacity is exceeded so that now, a queue builds up in the time interval. E.g., if in departure time interval 3 more than 20 travellers are added, a queue will grow in this time interval, causing a different sensitivity than when only one traveller is added. Another linearity condition is the inverse, while removing travellers, the sensitivity changes when the number of removed travellers is so large that the queue disappears from a time interval. These linearity conditions can be easily calculated when the bottleneck capacity $Q_{max}$ is known or can be observed by the queue lengths.

This section has given some useful insights to the problem. However, for a full network assignment, the travel time typically depends also on the flow (e.g., pre-saturation delays at intersections) and not only on the queue length. This will result in a more complex sensitivity matrix. Moreover, as mentioned above, the capacity of a network, with multiple bottlenecks that interact in different ways, is much harder to know.

# 3. A Family of Discrete Simultaneous-Swap Algorithms

In this section the structure of a new family of algorithms for numerically solving the single bottleneck case with sensitivities is proposed. During equilibration a departure rate is transformed to a new departure rate that is closer to equilibrium. The general structure of such equilibration is sketched in Figure 2. There are two main components: the Network Loading & Valuation and the Co-ordinated Departure Rate Adjustment. This last component is the new family of algorithms proposed in this paper and will be discussed in more detail in the next section. For each of the three submodules of this component, we will discuss the input, outputs, the goal of the submodule and its generic

properties. The main purpose for an algorithm of this new family is to calculate a new departure rate for each class that is closer to the user equilibrium, given an initial departure rate and corresponding costs. Running the overall framework iteratively should then produce the equilibrium departure rates. After each submodule is discussed generically, an instance of the family of algorithms will be discussed in detail in the full paper.

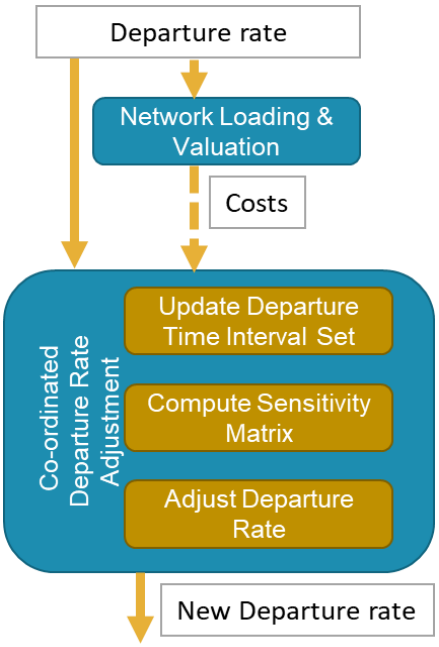



Figure 2 General Framework

## 3.1. Properties of the new Family of Algorithms

**Discrete** In this paper, we discretize time in departure time intervals. An analogue framework for continuous time is conceivable but is out of scope for this paper.

**Simultaneous** In order to account for the different effects and correlation among time intervals of the same class but also among different classes, the new update direction towards the equilibrium is determined simultaneously over all time intervals and classes.

## 3.2. Network Loading & Valuation

To use an instance of the newly proposed family of algorithms, it should be combined with a module (external to our algorithm family) that calculates travel costs. Given a departure rate as an input, this module calculates the perceived cost of a traveller for each simulated departure interval. This is done in two different steps. First, there is a network loading, which propagates traffic through the bottleneck. A property of a network loading is how queues are handled (is there spillback, is it just a vertical queue, …). Second, there is a valuation of the perceived costs. Each simulated time interval is given a cost that consists of the travel cost and the schedule delay cost.

## 3.3. Co-ordinated Departure Rate Adjustment

Given the departure rates and costs, this module adjusts the departure rates to new feasible departure rates that are closer to equilibrium. This adjustment is made in a coordinated way, and for every time interval and each class simultaneously. This requires three separate actions that are discussed in their respective submodule: (i) Update Departure Time Interval Set, (ii) Compute Sensitivity Matrix and (iii) Adjust Departure Rate. These submodules will be discussed separately.

Note that the simultaneity of the adjustments based on the full sensitivity matrix is what distinguishes this family of algorithms from the state of the art. It is also what makes these algorithms efficient: determining a set of coordinated and simultaneous adjustments that anticipate both separable and non-separable impacts. This means: for letting the costs in departure time interval $i$ converge to the target, we are not confined to changes to departure flows in interval $i$ *only*; instead, *any* interval $i'$ can contribute to this convergence if the sensitivity matrix indicates that the cost in interval $i$ is sensitive to departure flow changes by that class in interval $i'$. Moreover, by determining and balancing them all simultaneously, we avoid under- or overshooting the target (at least: within the assumptions of linear cost sensitivity). Such overshoot is otherwise difficult to prevent, as multiple adjustments may be aimed at closing the same gap with respect to the target and uncoordinated adjustments would be ignorant about how much of this gap will be closed by other adjustments once they are jointly implemented in an updated network loading.

### 3.3.1 Update Departure Time Interval Set

Given the departure rate and the costs for each simulated interval and class, this submodule selects for each class the set of departure time intervals to which non-zero departure flows can be assigned for equilibration in subsequent submodules. Depending on the implementation, the properties of this set can differ. E.g., does the set consist of consecutive intervals? Using consecutive intervals may seem logical, but it does have a consequence on the found adjustments; the type of consequences is depending on the implementations of the other submodules. The Adjust Departure Rate submodule should be able to handle those time intervals. In general, all selected intervals can have an influence on the search direction, depending on the constraints and projection in the Adjust Departure Rate submodule. Another important property of this submodule is its frequency of updating: in each iteration or only when certain criteria are met (going from a simple counter to complex criteria depending on the cost levels).

### 3.3.2 Compute Sensitivity Matrix

Given the current departure rate and interval set, this submodule estimates the sensitivity matrix ($S$ $[M \times M]$, with $M$ the number of time intervals in the set). The matrix itself indicates how the cost of all intervals in the set over all classes would change if there would be an extra departure in a given interval. There are different methods to estimate this sensitivity matrix, each differing in complexity, level of accuracy and computation time.

### 3.3.3 Adjust Departure Rate

Given the current departure rate, the interval set and the sensitivity matrix, this submodule calculates a feasible adjustment that aims to bring the costs closer to the equilibrium conditions. A feasible departure rate contains only non-negative departure flows for each interval and complies with the conservation of travellers. This means that no more than the original number of departures of a time interval can switch to another interval. The equilibrium conditions depend on whether travellers are

modelled as deterministic or stochastic. In the deterministic case, the equilibrium conditions determine that the cost of all time intervals in the set are equal. In the stochastic case, the equilibrium conditions are met when the new departure rates generate congestion costs that – according to the departure choice model – in turn cause these same departure rates. In both the deterministic and stochastic case, deviation from equilibrium can be measured by a gap function, which is a function of the congestion costs (that are departure flow dependent).

Adjusted departure rates can be found by searching which departure rates cause the gap function to be zero. In this search, the flow-dependent congestion costs are approximated using the sensitivity matrix as a local linear approximation of the network loading.

### Activation of Time Intervals

When an earlier time interval is added to the interval set, it has no departures. Therefore, its local sensitivity might be zero (e.g., when the network loading only considers delays by saturated bottlenecks, then a first flow increment will never exceed any bottleneck's capacity and produce no delay). Depending on the implementation of the other submodules and the specific situation, this new interval can attract no departures or can attract too many departures, as the linear cost approximation with zero sensitivity for this interval will assume costs to remain constant. This leads either to a situation where the problem does not converges or to a situation where the departure rates after swapping are worse due to overshooting. The Adjust Departure Rate sub-module should be able to handle the activation of new Time Intervals.

# 4. Numerical Example: Literature Case Study

In this section we will show the potential of the algorithms of our new algorithmic family by solving a numerical example from literature.

Table 2 Parameter Values for Literature Example

| | |
|---|---|
| $\alpha$ | $\$10/h$ |
| $\beta$ | $\$5/h$ |
| $\gamma$ | $\$15/h$ |
| $t^*$ | $2h$ or Time interval 40 |
| $N$ | $5100\ veh$ |
| $Q_{max}$ | $3533.6\ veh/h$ |
| $\Delta t$ | $0.05h$ |

In this numerical example, the case study from literature (Daly and Bar-Gera, 2021; Guo et al., 2017) is revisited. It represents a typical morning rush, all relevant parameters are given in Table 2. Note that Guo's example works with $N = 6000$ and $Q_{max} = 3000\ veh/h$, Daly & Bar-Gera argues that this may lead to over-optimistic impressions and have changed these parameters slightly. Therefore, the case study parameters by Daly & Bar-Gera are taken. We initialize by uniformly distributing the travellers from time interval 15 to 45. The costs found in equilibrium found by our algorithm are plotted in Figure 3, while Figure 4 and Figure 5 plot the departures and arrivals per time interval and their cumulative, respectively.

The algorithm is thus effective and finds the equilibrium. The real contribution of this paper is the efficiency of the new algorithmic family. Daly & Bar-Gera find the best results for their min-max algorithm with a dividing shift size of $1{,}05$. For reaching an RMS lower than $10^{-10}$, they need $\sim 1800$ iterations. Our algorithm requires 10 iterations (0,15 seconds). Figure 6 plots the convergence of this example.

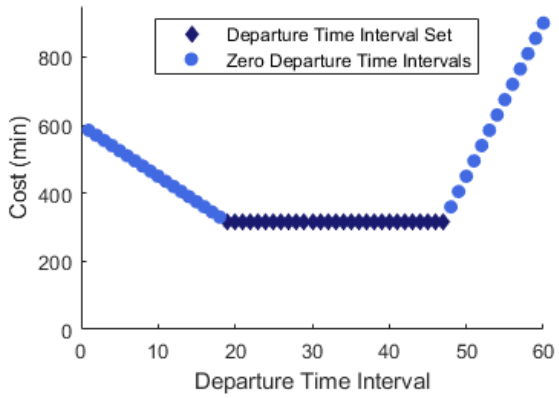



Figure 3 Equilibrium Costs for the single class Literature Example

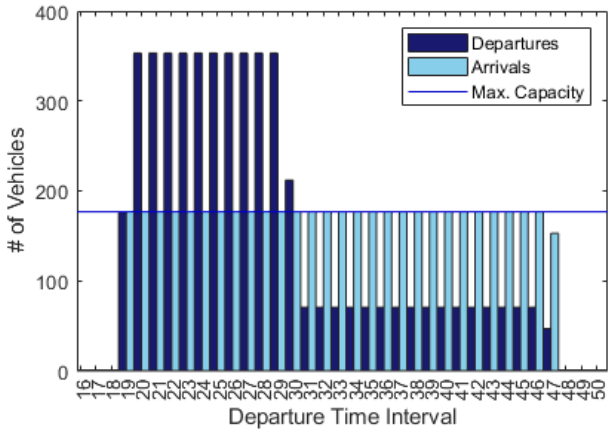



Figure 4 Departures and Arrivals for the single class Literature Example

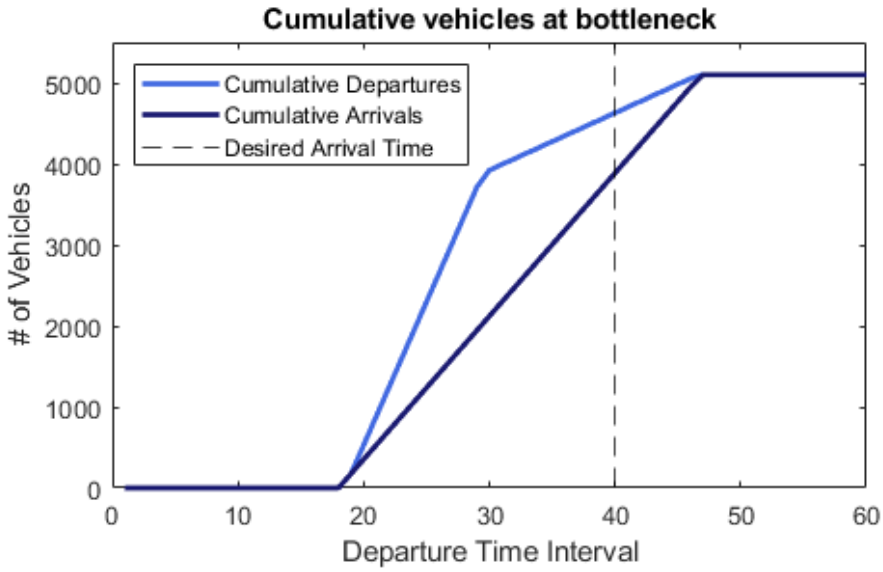



Figure 5 Cumulative Departures and Arrivals for the single class Literature Example

One iteration of the proposed algorithm requires more computations than the ones discussed in literature. Our algorithm called the network loading 347 times in total. This is still lower than what is reported in the cited literature, as they need one call for each iteration, hence 1800 calls. Note that most of our network loading calls are for the sensitivity calculation, which we implemented as a simple finite-difference method. This number can be further reduced by using smarter sensitivity methods like marginal computation (Corthout et al., 2014; Qian and Zhang, 2011).

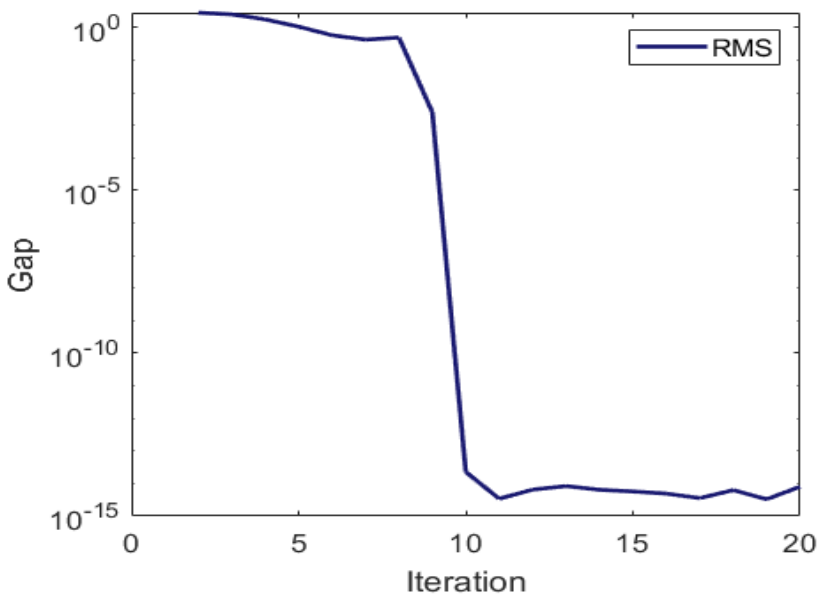



Figure 6 Convergence for the Day & Bar-Gera case

The reason the RMS is barely moving during the first 8 iterations is because during these iterations, the algorithm is finding the right set of time intervals. Once they have been found, the algorithm converges fast. The time interval activation is used in these first 8 iterations, while the last two iterations use only the local sensitivity to fine tune further the exact equilibrium costs.

# 5.Conclusion

In this paper, a new framework to solve a single bottleneck departure time choice problem with discrete departure time intervals is proposed. We analysed non-separability of the problem as a reason for the poor convergence of existing solution algorithms, and as a basis for our novel solution. The new family of algorithms solves the problem with departure rate adaptations for all time intervals simultaneously. The latter is what makes the family novel and unique: it takes all separable and non-separable influences of all departure time intervals on the cost of one interval into account, while existing algorithms only consider the influence of the own interval.

An algorithm of the family consists of multiple submodules, for which the input, the output and the purpose have been discussed. We demonstrate on a numerical example from literature how the algorithms of the new family solve the departure time choice problem much more efficiently (with potential to even more efficient solution if smarter solutions to compute sensitivity would be used). In the full paper, the pseudo-code and different implementations for each sub-module will be discussed. In the full paper, the family of algorithms will also be adjusted to be multi-class. Finally, an important challenge remains in conceiving network-wide algorithms that exploit the advantages of the efficient single-bottleneck solutions.

# 6. References


Arnott, R., de Palma, A., Lindsey, R., 1989. Schedule Delay and Departure Time Decisions with Heterogeneous Commuters.

Ben-Akiva, M., Cyna, M., de Palma, A., 1984. Dynamic model of peak period congestion. Transp. Res. Part B 18, 339–355. https://doi.org/10.1016/0191-2615(84)90016-X

Blumberg-Nitzani, M., Bar-Gera, H., 2014. The effect of signalised intersections on dynamic traffic assignment solution stability. Transp. Transp. Sci. 10, 622–646. https://doi.org/10.1080/18128602.2012.751682

Corthout, R., Himpe, W., Viti, F., Frederix, R., Tampère, C.M.J., 2014. Improving the efficiency of repeated dynamic network loading through marginal simulation. Transp. Res. Part C Emerg. Technol. 41, 90–109. https://doi.org/10.1016/j.trc.2013.12.009

Dafermos, S., 1980. Traffic Equilibrium and Variational Inequalities. Transp. Sci. 14. https://doi.org/10.1287/trsc.14.1.42

Daganzo, C.F., 1985. Uniqueness of a Time-Dependent Equilibrium Distribution of Arrivals At a Single Bottleneck. Transp. Sci. 19, 29–37. https://doi.org/10.1287/trsc.19.1.29

Daly, A., Bar-Gera, H., 2021. Exploration of Maximum-to-Minimum Shifts in Generic Methods for Departure Time Choice Models. Transp. Res. Rec. J. Transp. Res. Board 2675, 907–915. https://doi.org/10.1177/03611981211013348

de Palma, A., Marchal, F., 2002. Real Cases Applications of the Fully Dynamic METROPOLIS Tool-Box: An Advocacy for Large-Scale Mesoscopic Transportation Systems 2, 347–369.

de Palma, A., Marchal, F., Nesterov, Y., 1997. METROPOLIS: Modular system for dynamic traffic simulation. Transp. Res. Rec. J. Transp. Res. Board 1607, 178–184. https://doi.org/10.3141/1607-24

Guo, R.-Y., Yang, H., Huang, H.-J., 2018. Are We Really Solving the Dynamic Traffic Equilibrium Problem with a Departure Time Choice? Transp. Sci. 52, 603–620. https://doi.org/10.1287/trsc.2017.0764

Guo, R.Y., Yang, H., Huang, H.J., Li, X., 2017. Day-to-day departure time choice under bounded rationality in the bottleneck model. Transp. Res. Procedia 23, 551–570. https://doi.org/10.1016/j.trpro.2017.05.031

Han, K., Friesz, T.L., Yao, T., 2016. Existence of Simultaneous Route and Departure Choice Dynamic User Equilibrium ☆.

Han, K., Friesz, T.L., Yao, T., 2013. A partial differential equation formulation of Vickrey's bottleneck model, part I: Methodology and theoretical analysis. Transp. Res. Part B Methodol. 49, 55–74. https://doi.org/10.1016/j.trb.2012.10.003

Heydecker, B.G., Verlander, N., 1999. Calculation of dynamic traffic equilibrium assignments, in: Proceedings of the European Transport Conference. pp. 79–91.

Huang, W., Viti, F., Tampere, C., 2016. A dual control approach for repeated anticipatory traffic control with estimation of network flow sensitivity. J. Adv. Transp.

Iryo, T., 2019. Instability of departure time choice problem: A case with replicator dynamics. Transp. Res. Part B 126, 353–364. https://doi.org/10.1016/j.trb.2018.08.005

Iryo, T., 2008. An analysis of instability in a departure time choice problem. J. Adv. Transp. 42, 333–358. https://doi.org/10.1002/atr.5670420308

Jayakrishnan, R., Tsai, W.T., Prashker, J.N., Rajadhyaksha, S., 1994. A Faster Path-Based Algorithm for Traffic Assignment.

Lindsey, R., 2004. Existence, Uniqueness, and Trip Cost Function Properties of User Equilibrium in the Bottleneck Model with Multiple User Classes. Source Transp. Sci. 38, 293–314. https://doi.org/10.1287/trsc.l030.0045

Qian, Z. (Sean), Zhang, H.M., 2011. Computing Individual Path Marginal Cost in Networks with Queue Spillbacks: Transp. Res. Rec. https://doi.org/10.3141/2263-02

Ramadurai, G., Ukkusuri, S. V, Zhao, J., Pang, J.S., 2010. Linear complementarity formulation for single bottleneck model with heterogeneous commuters. Transp. Res. Part B Methodol. 44, 193–214. https://doi.org/10.1016/j.trb.2009.07.005

Schuster, P., Sigmund, K., 1983. Replicator Dynamics. J Theor Biol 100, 533–538.

Small, K.A., 2015. The bottleneck model: An assessment and interpretation. Econ. Transp. 4, 110–117. https://doi.org/10.1016/j.ecotra.2015.01.001

Smith, M.J., 1984. Existence of a Time-Dependent Equilibrium Distribution of Arrivals At a Single Bottleneck. Transp. Sci. 18, 385–394. https://doi.org/10.1287/trsc.18.4.385

Ubbels, B., Verhoef, E., 2005. Behavioural responses to road pricing. Empirical results from a survey among Dutch car owners •. Eur. Transp. Trasp. Eur. N 31, 101–117.

Vickrey, W., 1969. Congestion Theory and Transport Investment. Am. Econ. Rev. 59, 251–260.

Wardrop, J.G., 1952. Road Paper. Some Theoretical Aspects Of Road Traffic Research. Proc. Inst. Civ. Eng. 1, 325–362. https://doi.org/10.1680/ipeds.1952.11259

# Influence of Simulation Timestep Size and Pathfinding Interval Length on Path Patterns and Computational Cost in Dynamic Traffic Assignment

Siwei Hu[a], Pengyuan Sun[a], Daisik Nam[a,b,*],

R. Jayakrishnan[a], and Michael F. Hyland[a]

[a] *Department of Civil and Environmental Engineering,*

*University of California, Irvine, CA 92697-3600, USA*

[b]*Department of Logistics, Inha University,*

*Incheon, 22208, South Korea*

## Extended Abstract

*Keywords:* Dynamic Traffic Assignment, Path-based Gradient Projection Algorithm, Path Patterns, Simulation Timesteps, Pathfinding Intervals

## 1 Introduction

Although the pathfinding processes require significant computational resources in real-world Dynamic Traffic Assignment (DTA) applications (Adoni et al. 2018), the outcome of such pathfinding processes, the set of used paths at convergence, has rarely been studied in past DTA-related research. Hence, we investigate the equilibrium path patterns and their temporal evolution in DTA simulation in this paper to provide the path-level information required to reduce DTA's computational burden. For example, if for a certain OD pair, the equilibrium path set doesn't change as time evolves, then we don't need to conduct the pathfinding algorithm frequently. Instead, we could reduce the frequency of shortest path calculations to reduce DTA's computational cost. Specifically, in this paper, we investigate the impact of simulation timestep size and pathfinding interval length in DTA simulation on the used path patterns and its computational cost. We define simulation timestep size as the temporal resolution of the time-expanded network (e.g., if the simulation timestep size is 15 seconds, then traffic is updated every 15 seconds), and we define pathfinding interval length as the time gap between two consecutive shortest path calculations.

If the pathfinding intervals in a DTA model are very short (say 10 seconds), it is unlikely for an Origin-Destination-departure-Time (ODT) triad to have multiple shortest paths at equilibrium with the same travel time. However, if the pathfinding intervals are longer (say 1 or 2 minutes) multiple paths with equivalent travel times may exist at equilibrium. If the pathfinding intervals are too short, the pathfinding algorithm may repeatedly find the same physical road path (same sequence of physical links) many times, thereby adding unnecessary computations. Conversely, if the interval is too long, the pathfinding algorithm may overlook physical road alternative paths that may be available during some short time periods. Thus, we aim to study path patterns for different pathfinding interval lengths, to avoid unnecessarily repetitive computation for pathfinding and to accurately capture the path patterns.

* Corresponding author. E-mail address: namd@inha.ac.kr.

Along with the above-mentioned impact of pathfinding time interval length, it is also important to investigate the influence of different simulation timestep sizes on DTA's computation cost. Too long a simulation time step (typically more than 30 seconds or so) may limit the DTA model's ability to effectively capture traffic dynamics. Sub-second timesteps, as needed for microscopic traffic modeling, are computationally impractical in most DTA. Dealing with DTA's heavy computation while capturing the necessary equilibrium paths may require the investigation of path patterns under different pathfinding interval lengths and simulation timestep sizes.

To fill the aforementioned research gaps, we study equilibrium path patterns and computational costs under different pathfinding interval lengths and simulation timestep sizes. We adopt the simulation-based DTA model proposed by Nam (2019), which solves the Dynamic User Equilibrium (DUE) problem on a network model with mesoscopic traffic flow relations, using a path-based Gradient Projection (GP) algorithm (Jayakrishnan, Tsai, and Prashker 1994). We apply the algorithm to a realistic network centered on the I-10 freeway in Los Angeles (LA), California.

## 1.1 Overview of the DTA Model

Figure 1 gives an overview of Nam (2019)'s DTA model, which consists of an all-or-nothing assignment initialization, a pathfinding module, and a GP algorithm to adjust the path traffic loads. Since the equilibrium path sets are the focus of this paper, we study the DUE cases where all travelers sharing the same ODT triad will have the same travel time. The path-based GP algorithm stores the utilized path set automatically as the algorithm is executed, and usually converges in relatively few iterations (Jayakrishnan, Tsai, and Prashker 1994). For detailed formulation, readers are referred to Nam (2019).

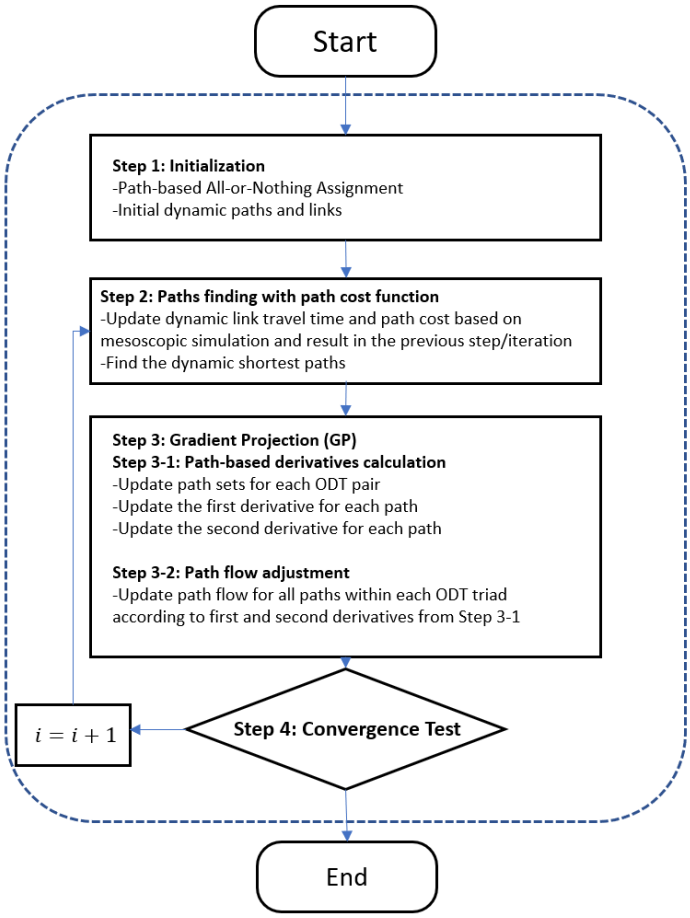



**Figure 1 Overview of Nam (2019)'s DTA model**

# 2 Results and Discussion

We apply Nam (2019)'s DTA model to a realistic network of 799 nodes and 1,927 links, centered on the I-10 freeway in LA. The experiments were on a computer with an i7-6800K CPU and 32 GB of RAM storage. This section first presents the temporal evolution of path patterns (number of paths, average path travel time, path appearance in this study) of short-, middle-, and long-distance ODs pairs. Later, this section also presents how path sets and computational time change under different pathfinding interval lengths and simulation timestep sizes.

## 2.1 Temporal Path Patterns of DTA

Figure 2 presents the temporal path patterns of short-, middle- and long-distance OD pairs. We have 3,712 OD pairs in total, and we divide them into short-, middle- and long-distance categories. We then select one OD pair in each distance category as a representative.

In terms of the number of paths, all three selected OD pairs from different categories find multiple paths, and the number of equilibrium paths might vary over time. The path travel times for all selected OD pairs increase as congestion increases during the analysis period. In terms of path appearance, which is the time for a path appearing in the equilibrium path set, 7 different paths are found between the short-distance OD representative at different times, while 8 different paths are found between the middle-distance OD representative at different times. And 3 different paths are found between long-distance OD representative at different times. For the short- and middle-distance OD representatives, more new paths are found when congestion increases. For long-distance OD representative, all 3 paths appearing in the equilibrium path set are found at the beginning of the simulation and no new paths are found as time evolves, but when congestion increases, path 1 is not used and only path 2 and path 3 are used. In general, as congestion grows, more new paths are found between the short- and middle-distance OD pairs. But for long-distance OD pairs, the equilibrium path sets are relatively stable as congestion evolves, and not many new paths are found.

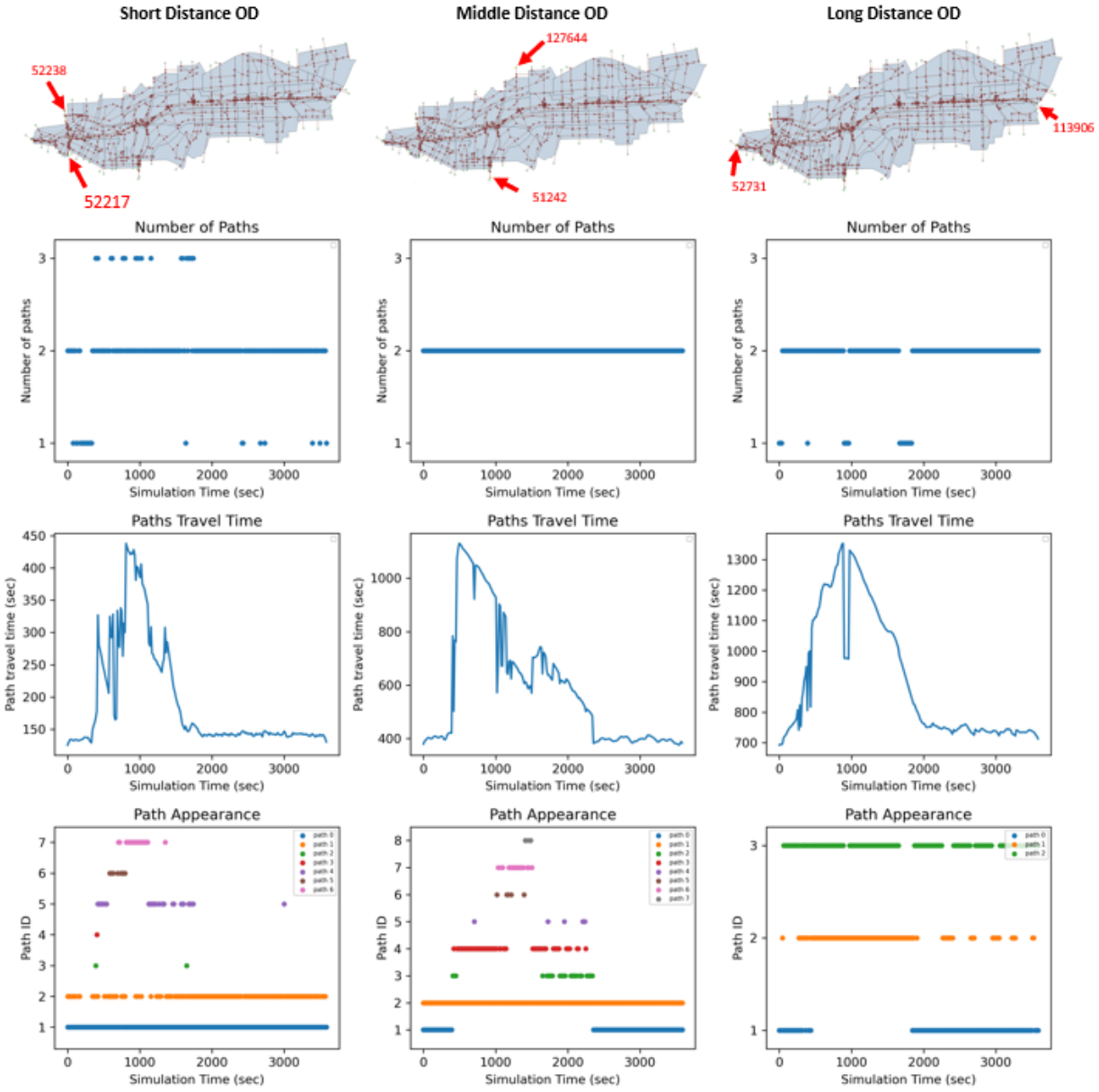



**Figure 2 Temporal path patterns of short-, middle-, and long-distance OD pairs**

## 2.2 Path patterns under different settings

Table 1 presents the total number of used paths found throughout the simulation, as well as the maximum, minimum, and the average number of used paths for OD pairs (paths from all time steps) and for ODT triad (paths from one time step) under different simulation timestep sizes and pathfinding interval lengths. We set pathfinding interval lengths equal to simulation timesteps when analyzing different simulation timesteps. When we analyze different pathfinding intervals, we set the simulation timestep size to 15 sec. Total OD pairs and ODT triads vary in Table 1 because there are 42.8% OD pairs with less than 1 trip (even to the magnitude of $10^{-5}$) in the demand input. In the demand loading process, some OD pairs round up to 1 trip, while some round down to 0, based on the random numbers generated. Thus, the number of total OD pairs might vary. From a computational perspective, computational time decreases as simulation timestep size and pathfinding interval length increase. In general, when the simulation timestep size

increases from 15 seconds to 30 seconds, the total number of used paths at the OD pair level and the average number of used paths at the ODT triad level increase. In contrast, when the pathfinding interval length increases from 30 seconds to 90 seconds, the average number of used paths at both OD pair and ODT triad levels generally decreases.

### 2.3 Discussion

This study presents the temporal evolution of the path patterns of a path-level DTA model proposed by Nam (2019). The results indicate that the number of equilibrium paths found varies across different simulation timesteps, and as congestion increases, more new paths will appear. It should be noted that such results are not limited to the specific DTA implementation studied here. They may be generalized to other path-based DTA models with proper modeling of traffic dynamics. It is intuitive that under low demand, all traffic may be on the shortest path, while when demand increases, some traffic may divert to other paths with the same travel cost. Therefore, more new paths will appear when congestion increases - a phenomenon that is not captured well by uncapacitated BPR-like functions used in static models but will be present in any DTA framework incorporating proper traffic flow models that constraints link flows to effective maximum values. This study also analyzes path patterns at the origin-destination-departure-time triad level under different simulation timestep sizes and pathfinding interval lengths. Computational time decreases as simulation timestep size and pathfinding interval length increase. In general, the average number of used paths increases as the simulation timestep size increases, and it decreases as the pathfinding interval length increases. However, further investigation is still needed on developing different path-level metrics to understand whether the equilibrium path sets are properly captured under different experimental contexts.

## 3 Acknowledgements

This study was funded, in part, by PSR-20-31 TO 040 project from the State of California, Department of Transportation (Caltrans) and the USDOT Pacific Southwest Research University Transportation Center, whom the authors thank for supporting this research study. The authors remain solely responsible for the information in this manuscript.

## 4 References

Adoni, Wilfried Yves Hamilton, Tarik Nahhal, Brahim Aghezzaf, and Abdeltif Elbyed. 2018. "The MapReduce-Based Approach to Improve the Shortest Path Computation in Large-Scale Road Networks: The Case of A* Algorithm." *Journal of Big Data* 5 (1). https://doi.org/10.1186/s40537-018-0125-8.

Jayakrishnan, R, Wei T Tsai, and Joseph N Prashker. 1994. "A Faster Path-Based Algorithm for Traffic Assignment." In *Transportation Research Record 1443*, 75–83. https://escholarship.org/uc/item/2hf4541x.

Nam, Daisik. 2019. "Modelling and Optimization of Smart Mobility Systems with Agent Envy as a Paradigm for Fairness and Behavior." UC Irvine. https://doi.org/ProQuest ID: Nam_uci_0030D_15996. Merritt ID: ark:/13030/m5188f19.

**Table 1 Path patterns under different simulation timestep sizes and pathfinding interval lengths**

| | | OD Pair Level | | | | | ODT Triad Level | | | | | Computation |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Total path num | Max path num | Min path num | Avg path num | Total OD pairs | Total path num | Max path num | Min path num | Avg path num | Total ODT triads | Time (sec) |
| Simulation timestep sizes | 10 sec | 18403 | 52 | 1 | 4.897 | 3758 | 117764 | 3 | 1 | 1.596 | 73775 | 9,454 |
| | 15 sec | 16092 | 43 | 1 | 4.335 | 3712 | 101833 | 3 | 1 | 1.567 | 64993 | 6,809 |
| | 20 sec | 18220 | 46 | 1 | 4.885 | 3730 | 111282 | 3 | 1 | 1.654 | 67264 | 5,986 |
| | 25 sec | 19074 | 56 | 1 | 5.095 | 3744 | 110093 | 3 | 1 | 1.690 | 65130 | 4,817 |
| | 30 sec | 20034 | 55 | 1 | 5.342 | 3750 | 106318 | 3 | 1 | 1.700 | 62527 | 4,564 |
| Pathfinding interval lengths | 15 sec | 16092 | 43 | 1 | 4.335 | 3712 | 101833 | 3 | 1 | 1.567 | 64993 | 12,151 |
| | 30 sec | 17598 | 30 | 1 | 4.713 | 3734 | 114565 | 3 | 1 | 1.626 | 70472 | 10,460 |
| | 45 sec | 17083 | 27 | 1 | 4.602 | 3712 | 112910 | 3 | 1 | 1.614 | 69936 | 10,036 |
| | 60 sec | 17161 | 29 | 1 | 4.596 | 3734 | 113829 | 3 | 1 | 1.615 | 70472 | 9,023 |
| | 75 sec | 17071 | 31 | 1 | 4.572 | 3734 | 113496 | 3 | 1 | 1.611 | 70472 | 8,366 |
| | 90 sec | 16655 | 23 | 1 | 4.487 | 3712 | 112030 | 3 | 1 | 1.602 | 69936 | 7,947 |

# Session 4

Metro Passenger Routing Choices Estimation with A Fully Differentiable Simulation-Based Optimization (SBO) Approach
*Kejun Du, Enoch Lee, Qiru Ma, Zhiya Su and Hong K. Lo*

A Scalable Framework for Public Transit Assignment
*Prateek Agarwal, Tarun Rambha, Dhanus Lal and Ayush Kumar*

Modeling Multi-modal Curbside Usage in Dynamic Networks
*Jiachao Liu and Sean Qian*

# Metro Passenger Routing Choices Estimation with A Fully Differentiable Simulation-Based Optimization (SBO) Approach

Kejun Du[a], Enoch Lee[a], Qiru Ma[a], Zhiya Su[a], Hong K. Lo[a*]

[a]Department of Civil and Environmental Engineering, Hong Kong University of Science and Technology, Hong Kong, China

## 1. INTRODUCTION

Hong Kong is a highly dense city, with rapid growths in passenger travel demand, which burdens the urban transportation system. The Hong Kong government addressed this problem by introducing an extensive multi-modal public transportation network, among which the Mass Transit Rail (MTR) is the largest carrier. With about 90% of trips conducted through public transport in Hong Kong, MTR has significant influences on the overall Hong Kong transportation, as around 50% of public transportation trips in Hong Kong are served by MTR, and this percentage goes even higher for the peak period.

However, MTR, with its history of over 45 years, has inevitably experienced occasional incidents, disruptions, and serious congestion, which resulted in massive travel delays and significantly affected the city's mobility. Therefore, metro disruption mitigation is an essential element of urban emergency management (Zhang et al., 2021). For the sake of both disruption management and further MTR system development, it is of growing importance to grasp the travel demands and passenger behaviors in the MTR system. Currently, around 94% passengers access the MTR system with the Octopus smart card, which records passengers' origins and destinations with temporal information. However, the Octopus smart card information does not provide the critical routing information of passengers within the MTR network; within which multiple route choices are possible. Despite the lack of routing information, knowing specific passenger movements in the MTR system is essential to support train loading modeling and train schedule planning. More importantly, it could provide short-term origin-destination demand forecast in the event of a disruption, which is critical for service replanning to mitigate its impact.

Previous works have modeled passenger routing behaviors within metro systems. Some studies utilized smartcard data to infer passenger route choices based on the passenger travel time distributions. Qiao et al. (2013) analyzed the passenger route choices within the Beijing metro network based on passenger classification to improve the passenger behavior models to match with the actual situations. Wu et al. (2019) proposed a data-driven model of estimating passenger route choices in the metro system by clustering passenger travel times and assigning different routes to corresponding travel times. Kim et al. (2020) utilized smartcard transaction data to calibrate a passenger route choice model for a multi-modal network including a metro network. Lee et al. (2022) formulated the empirical cumulative distribution functions (ECDFS) of passenger travel times and allocated passengers to different routes with the highest probabilities, validated with smartcard and transfer gate data. Some other studies went for simulation-based optimizations (SBO), calibrating passenger route choices by replicating the real data. Mo et al. (2020) constructed an event-based simulation model for the Hong Kong MTR system and minimized the difference between the simulated and real passenger exit flows, calibrating the passenger route choice model parameters. Further, Mo et al. (2021) utilized multiple heuristic optimization algorithms for the MTR simulation-based optimization problem, calibrating a passenger route choice model with train cabin congestion considered.

However, exit flows and travel time distributions used in previous studies can't distinguish routes with similar travel times, new constraints are required for the optimization problem to further constrain the optimal solutions. Recently, MTR updated train-cars that provides train loadings per car, which is a very useful input. Train loadings can be

translated into passenger numbers approximately, which can then be converted into link flows between adjacent stations along the time. On the other hand, this study introduces the concept of relative utilities and applies logit model to estimate passenger route choices. To calibrate the relative utilities, this study formulates a data-driven simulation-based optimization framework, utilizing Octopus smart card tap-in and tap-out records, train loadings and possibly other information to reversely replicate metro passenger routing behaviors.

Previous simulation-based optimizations are limited for low dimensional problems due to their non-analytical agent-based simulation framework that make gradient estimation intractable. Moreover, perturbation-based numerical gradient computation methods are computationally prohibitive for high dimensional optimizations. Therefore, this study proposes a fully differentiable flow-based simulation model for the MTR system, which simulates passenger flows microscopically based on relative utilities, realizing queueing, platform congestions, and transfer delays. Then, this study introduces the Iterative Backpropagation (IB) to optimize the passenger route split ratios such that the simulation outputs can satisfyingly replicate those in the actual MTR records. For the forward process, the simulation model gives out link flows and other outputs based on relative utilities as variables, and the backpropagation process goes backward through the simulation to calculate the derivatives of the differences between the simulation outputs and real records with respect to the relative utilities. With the analytical derivatives, the perceived relative utilities are optimized with the Adam optimizer iteratively. The incoming passengers are considered to follow Poisson distributions, introducing stochasticity into the optimization process to mimic the idea of Stochastic Gradient Descent (SGD), so that saddle points can be avoided through the optimization process.

The proposed framework can provide insightful estimates to understand the actual behaviors and calibrate the relative utilities, which will form the basis to estimate passenger rerouting behaviors under disruption.

## 2. METHODOLOGY

The mathematical formulations of the flow-based MTR simulation model are presented below, starting from the concept of relative utilities. In this study, relative utilities are introduced instead of generalized utilities, because this study focuses on passenger route split ratios for each OD pair. Relative utilities $u_{i,j}^r$ for passengers on OD pair $i,j$ of route $r$ are defined as the relative preferences of the passengers on the specific route between the specific OD pair, and the corresponding passenger split ratio $\beta_{i,j}^r$ for route $r$ is considered to follow the logit model:

$$\beta_{i,j}^r = \frac{e^{u_{i,j}^r}}{\sum_r e^{u_{i,j}^r}}$$

The incoming passengers for each minute are assumed to follow the Poisson distribution, in which the mean is the average number of passengers traveling through the OD. This stochasticity is beneficial for the model calibration, since it eliminates saddle points with fluctuating gradients. Incoming passengers are specified for each OD pair $i,j$ and each simulated time step $t$:

$$p_{i,j}^t \sim Poisson\left(\lambda_{i,j}^t\right)$$

Then, the number of passengers incoming at time step $t$ choosing to travel between OD pair $i,j$ on route $r$ can be expressed as $\beta_{i,j}^r p_{i,j}^t$, which is the minimum passenger flow units simulated in the MTR system (which will not be split anymore). After calculating route choices for each OD pair, the simulation model will output passenger exit flows and link flows, which are defined for each discretized time slots. For each time slot $s$, the passenger exit flow between OD pair $i,j$ and the link flow between adjacent stations $n,m$ are denoted respectively as $q_{i,j}^s$ and $f_{n,m}^s$. Both the passenger exit flows and link flows come from multiple minimum passenger flow units added up together, which is the inner mechanism of the simulated MTR system. Therefore, to mathematically express the passenger exit flows and link flows, two indicator functions $I_{exit}\left((i,j),r,t,s\right)$ and $I_{link}\left((i,j),r,t,(n,m),s\right)$ are introduced:

$$I_{exit}\big((i,j),r,t,s\big)=\begin{cases}1\ if\ the\ flow\ unit\ \beta_{i,j}^{r}p_{i,j}^{t}\ exits\ from\ the\ system\ in\ time\ slot\ s\\ 0\ otherwise\end{cases}$$

$$I_{link}\big((i,j),r,t,(n,m),s\big)=\begin{cases}1\ if\ the\ flow\ unit\ \beta_{i,j}^{r}p_{i,j}^{t}\ passes\ between\ link\ (m,n)\ in\ time\ slot\ s\\ 0\ otherwise\end{cases}$$

The indicator functions are actually an alternative representation of the MTR simulation model, reflecting how passenger flows traveling through the MTR network dynamically. Then, the passenger exit flows and link flows can be analytically represented in the optimization problem formulation.

To calibrate the simulation model, all the concerned simulation outputs are summarized as $\boldsymbol{F}_{sim}$, and corresponding recorded real values are denoted as $\boldsymbol{F}_{real}$. The calibrating objective is to minimize the difference between the simulation outputs and the real values. Since they are defined as discretized time series of flow, mean squared error (MSE) is chosen to be the minimization objective function. The general structure of the data-driven simulation-based optimization framework with relative utilities is illustrated in Figure 2, which shows how data flow inside the MTR simulation model. The gradient calculations will flow backwards, or reversely through the forward passes, during the backpropagation process.

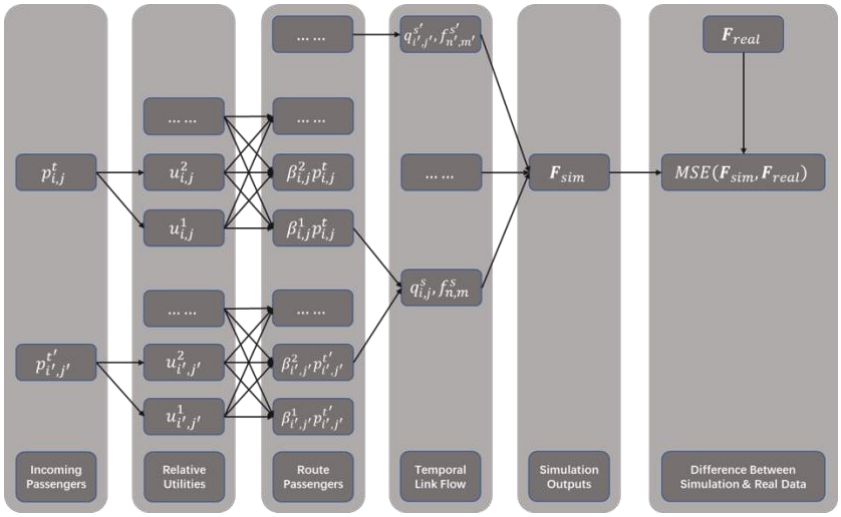


Figure 2. General structure of the proposed SBO framework

The problem of estimating passenger routing behaviors within the MTR system can be formulated as a minimizing problem the difference between the simulated MTR performance and the recorded data. In the following mathematical formulation, the sets of all OD pairs and all links are denoted as $\Theta$ and $\Omega$ respectively. The flow-based MTR simulation (FMS) model is denoted as $FMS(\cdot)$, which yields the indicator functions $I_{exit}(\cdot)$ and $I_{link}(\cdot)$. The recorded passenger exit flows and link flows are denoted as $\tilde{q}_{i,j}^{s}$ and $\tilde{f}_{n,m}^{s}$ respectively.

$$\min_{\boldsymbol{u}}\sum_{i,j\in\Theta}\left(q_{i,j}^{s}-\tilde{q}_{i,j}^{s}\right)^{2}+\sum_{n,m\in\Omega}\left(f_{n,m}^{s}-\tilde{f}_{n,m}^{s}\right)^{2}$$

Subject to:

$$I_{exit}\big((i,j),r,t,s\big),I_{link}\big((i,j),r,t,(n,m),s\big)=FMS\big((i,j),r,t,(n,m),s\big)$$

$$q_{i,j}^{s}=\sum_{t}\sum_{r}\beta_{i,j}^{r}p_{i,j}^{t}I_{exit}\big((i,j),r,t,s\big)$$

$$f_{n,m}^{s}=\sum_{t}\sum_{i,j}\sum_{r}\beta_{i,j}^{r}p_{i,j}^{t}I_{link}\big((i,j),r,t,(n,m),s\big)$$

$$\boldsymbol{u}=\left(\boldsymbol{u}_{1,1},\ \boldsymbol{u}_{1,2},\ \ldots,\ \boldsymbol{u}_{i,j}\right)$$

$$\boldsymbol{u}_{i,j}=\left(u_{i,j}^{1},u_{i,j}^{2},\ \ldots,u_{i,j}^{r}\right)$$

$$\beta_{i,j}^{r} = \frac{e^{u_{i,j}^{r}}}{\sum_{r} e^{u_{i,j}^{r}}}$$

The passenger flow calculations are continuous and differentiable, allowing the gradients of the mean squared errors (MSE) with respect to the relative utilities to be analytically derived. Therefore, this study proposes to adopt the Iterative Backpropagation (IB) algorithm to progressively optimize the relative utilities with analytical gradients and the gradient-based Adam optimizer, which has exhibited significant stability and adaptivity in optimization.

In this study, TensorFlow 2.0 is used for both the Iterative Backpropagation (IB) process and the Adam optimizer implementation. The automatic backpropagation mechanism built in TensorFlow 2.0 is utilized to automatically track the passenger flow units and generate gradients throughout the whole simulation process, which substantially improve the efficiency and accuracy and alleviate the burden of redeployment.

The gradients can also be derived analytically and manually, embedded in the optimization algorithm. For illustration, the gradients of the mean squared errors of the passenger exit flows and link flows with respect to the relative utilities are shown as following, using the chain rule of derivatives.

$$\frac{\partial\left(q_{i,j}^{s} - \tilde{q}_{i,j}^{s}\right)^{2}}{\partial u_{i',j'}^{r'}} = \begin{cases} \dfrac{\partial\left(q_{i,j}^{s} - \tilde{q}_{i,j}^{s}\right)^{2}}{\partial q_{i,j}^{s}} \sum_{r} \left( \dfrac{\partial q_{i,j}^{s}}{\partial \beta_{i,j}^{r}} \dfrac{\partial \beta_{i,j}^{r}}{\partial u_{i,j}^{r'}} \right) & if\ (i',j') = (i,j) \\ 0\ otherwise \end{cases}$$

$$\frac{\partial\left(q_{i,j}^{s} - \tilde{q}_{i,j}^{s}\right)^{2}}{\partial q_{i,j}^{s}} \sum_{r} \left( \frac{\partial q_{i,j}^{s}}{\partial \beta_{i,j}^{r}} \frac{\partial \beta_{i,j}^{r}}{\partial u_{i,j}^{r'}} \right) = 2\left(q_{i,j}^{s} - \tilde{q}_{i,j}^{s}\right) \sum_{r} \left( \left( \sum_{t} p_{i,j}^{t} I_{exit}\left((i,j), r, t, s\right) \right) \frac{\partial \beta_{i,j}^{r}}{\partial u_{i,j}^{r'}} \right)$$

$$= 2\left(q_{i,j}^{s} - \tilde{q}_{i,j}^{s}\right) \left( \sum_{r \neq r'} \left( -\left( \sum_{t} p_{i,j}^{t} I_{exit}\left((i,j), r, t, s\right) \right) \beta_{i,j}^{r} \beta_{i,j}^{r'} \right) \right.$$

$$\left. + \left( \sum_{t} p_{i,j}^{t} I_{exit}\left((i,j), r', t, s\right) \right) \left(1 - \beta_{i,j}^{r'}\right) \beta_{i,j}^{r'} \right)$$

$$\frac{\partial\left(f_{n,m}^{s} - \tilde{f}_{n,m}^{s}\right)^{2}}{\partial u_{i,j}^{r'}} = \frac{\partial\left(f_{n,m}^{s} - \tilde{f}_{n,m}^{s}\right)^{2}}{f_{n,m}^{s}} \sum_{r} \left( \frac{\partial f_{n,m}^{s}}{\partial \beta_{i,j}^{r}} \frac{\partial \beta_{i,j}^{r}}{\partial u_{i,j}^{r'}} \right)$$

$$= 2\left(f_{n,m}^{s} - \tilde{f}_{n,m}^{s}\right) \sum_{r} \left( \left( \sum_{t} \sum_{i,j} p_{i,j}^{t} I_{link}\left((i,j), r, t, (n,m), s\right) \right) \frac{\partial \beta_{i,j}^{r}}{\partial u_{i,j}^{r'}} \right)$$

$$= 2\left(f_{n,m}^{s} - \tilde{f}_{n,m}^{s}\right) \left( \sum_{r \neq r'} \left( -\left( \sum_{t} \sum_{i,j} p_{i,j}^{t} I_{link}\left((i,j), r, t, (n,m), s\right) \right) \beta_{i,j}^{r} \beta_{i,j}^{r'} \right) \right.$$

$$\left. + \left( \sum_{t} \sum_{i,j} p_{i,j}^{t} I_{link}\left((i,j), r', t, (n,m), s\right) \right) \left(1 - \beta_{i,j}^{r'}\right) \beta_{i,j}^{r'} \right)$$

After the calibration, the passenger routing behaviors are reconstructed with the calibrated relative utilities. Although the relative utilities do not bear specific physical meanings, they can be viewed as the passengers' perceived relative route utilities. This SBO framework can be applied to any other utility-based passenger behavior models with the relative utilities in the future.

## 3. PRELIMINARY NUMERICAL RESULTS

The framework is evaluated under some preliminary numerical experiments, where the mean passenger OD demands are extracted from the Octopus card data and the simulated incoming passenger numbers are sampled from the Poisson distribution. Train travel times between stations and passenger transfer walking times are setup according to the information provided by MTR. The train schedules are defined arbitrarily for preliminary experiments.

The simulation time duration is one hour, and the link flows (train loadings) are recorded for every 10 minutes of two links on the MTR Tsuen Wan Line (TWL). The learning rate of the Adam optimizer is set to be 0.2 and 150 optimization steps are implemented, the convergence process is visualized in terms of the minimization of mean squared error (MSE) and mean absolute percentage error (MAPE) in Figure 3 and 4. The MAPE of the link flows after 150 optimization steps drops from nearly 50% to less than 3%, yielding satisfying accuracies of convergence. Hence, the calibration significantly reduced the difference between simulated flow and real flow.

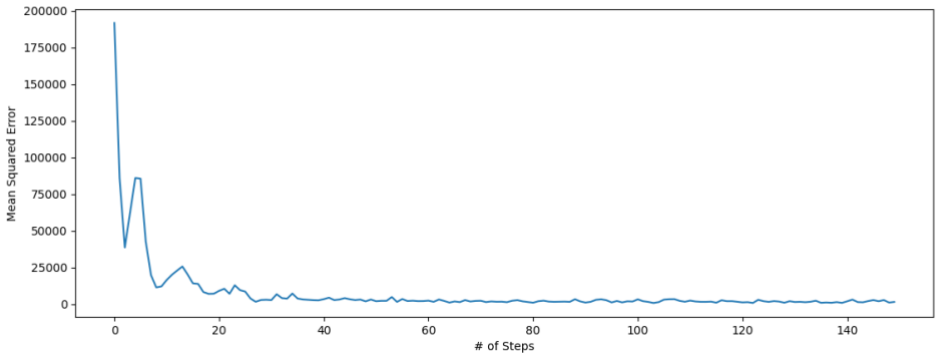



Figure 3. Mean square error convergence

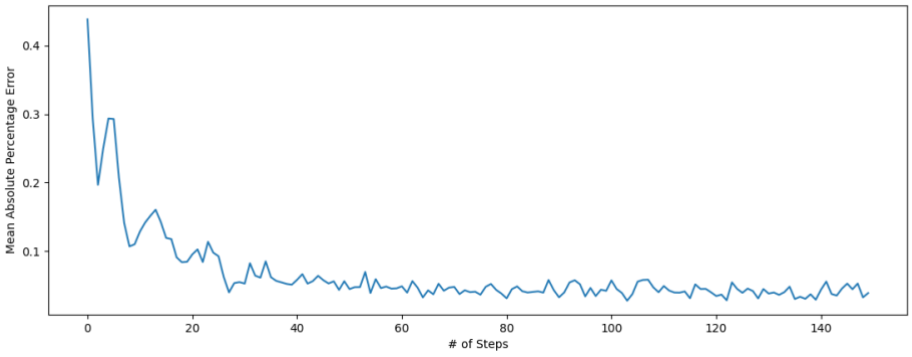



Figure 4. Mean absolute percentage error convergence

Multiple target link flows are chosen for evaluating the robustness of convergence for the proposed framework. Two different scenarios of target link flows are chosen for visualizing the results, where the target link flows, the default simulated link flows before calibration and the calibrated link flows are shown respectively in Figure 5 and 6. Clearly, the calibrated link flows can well replicate the target link flows, which proves the capability of the proposed framework in calibrating passenger route choices within the metro system.

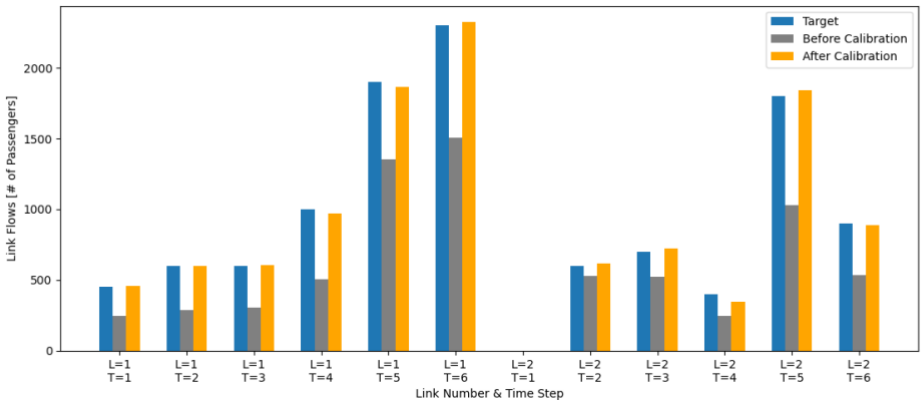



Figure 5. Calibration results (target flow scenario 1)

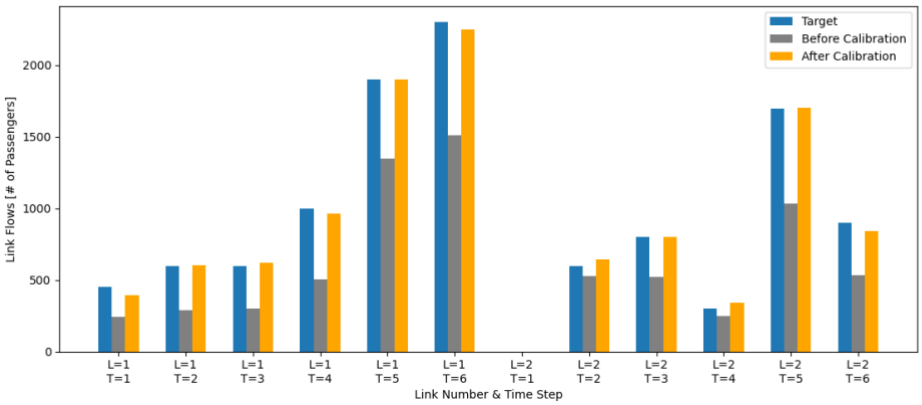



Figure 6. Calibration results (target flow scenario 2)

## 4. REFERENCES

Kim, I., Kim, H. C., Seo, D. J. and Kim, J. I. (2020). Calibration of a transit route choice model using revealed population data of smartcard in a multimodal transit network. Transportation, 47(5), 2179-2202.

Lee, E. H., Kim, K., Kho, S. Y., Kim, D. K., & Cho, S. H. (2022). Exploring for Route Preferences of Subway Passengers Using Smart Card and Train Log Data. Journal of Advanced Transportation, 2022.

Mo, B., Ma, Z., Koutsopoulos, H. N., & Zhao, J. (2020). Capacity-constrained network performance model for urban rail systems. Transportation Research Record, 2674(5), 59-69.

Mo, B., Ma, Z., Koutsopoulos, H. N., & Zhao, J. (2021). Calibrating path choices and train capacities for urban rail transit simulation models using smart card and train movement data. Journal of Advanced Transportation, 2021.

Qiao, K., Zhao, P., & Qin, Z. P. (2013). Passenger route choice model and algorithm in the urban rail transit network. Journal of Industrial Engineering and Management (JIEM), 6(1), 113-123.

Wu, J., Qu, Y., Sun, H., Yin, H., Yan, X. and Zhao, J. (2019). Data-driven model for passenger route choice in urban metro network. Physica A: Statistical Mechanics and Its Applications, 524, 787-798.

Zhang, S.Y., Lo, H.K., Ng, K.F. Chen, G. 2021. Metro system disruption management and substitute bus service: a systematic review and future directions. Transport Reviews, 41(2), 230-251.

# A Scalable Framework for Public Transit Assignment

Prateek Agarwal[1], Tarun Rambha[1,2], Dhanus M Lal[3], and Ayush Kumar[4]

[1]Department of Civil Engineering, Indian Institute of Science
[2]Center for infrastructure, Sustainable Transportation and Urban Planning (C*i*STUP), Indian Institute of Science
[3]Department of Mathematics, Indian Institute of Science
[4]Department of Computer Science and Engineering, Manipal Institute of Technology

## 1 Introduction

Public transit systems form a vital component of urban mobility. They offer a sustainable, equitable, and economical solution to transporting people. For transit to be an attractive option for travel and commuting, operators constantly aim to improve service quality. However, the benefits of such supply-side improvements heavily depend on inherently complex user behavior. Unlike personal mobility users, transit travelers typically use several criteria to choose their routes/journeys. Example criteria include in-vehicle travel time, waiting duration, walking time, cost, number of transfers, type of service (metro vs. buses), and crowding/comfort. Operators, on the other hand, have to decide routes, schedules, and fares (or modifications to existing ones) to improve key performance indicators while accounting for the changes in passenger choices that their actions may trigger. These choices are typically made in an ad-hoc manner, particularly in data-sparse environments such as cities in developing countries where smart cards and real-time GPS data are less prevalent. These cities are also plagued by crowding and reliability issues. A broad question of interest in this context is — "How can operators make better decisions to improve public transit?". This research aims to answer a critical facet of this question — the transit assignment problem.

The transit assignment problem predicts travelers' route choices for a given supply configuration in a transit network. While these choices can be observed using technological solutions such as onboard cameras and mobile apps, answers to "what-if" questions remain elusive. The problem requires addressing several issues summarized below, which are elaborated on in the next section.

- Given a set of transit schedules and network characteristics, can we determine what routes will be chosen by a particular individual?
- How do collective selfish choices of agents translate to flows and crowding in transit systems?
- What actions can an operator take to improve service levels, ridership, and revenue?

The assignment problem has been widely studied as network games in road routing, and solution concepts and algorithms that settle at a Nash equilibrium have been proposed. A related social optimum problem attempts to distribute flows to minimize system-level objectives. While similar ideas are in transit literature in two varieties, frequency-based (Schmöcker et al., 2011) and schedule-based assignment (Hamdouch and Lawphongpanich, 2008), they have been relatively less successful in terms of accuracy, scale, and computational efficiency. (Literature also points to Braess-like paradoxes where an increase in transit capacity can worsen system performance (Jiang and Szeto, 2016).) Frequency-based models are helpful for smaller subsets of routes where wait times for transfers are not critical. Scheduled-based approaches, on the other hand, have gained traction recently with increased data availability. A key difference in road routing and transit is that the supply exhibits a staccato-like time dependency. Capacity between stops exists only when buses pass through them. Further, transit-capacity constraints are notoriously hard to model. A passenger may choose a journey with a transfer, but the chances of boarding the transfer bus or finding a seat in it may be affected by passenger trips that originate at other locations in the future. Bus capacities are also small in magnitude compared to roadway capacities, making it challenging to formulate continuous mean-field approximations that can be cast as convex optimization problems. While there are a few variational inequality

approximation models, they remain highly intractable for medium to large networks (Hamdouch et al., 2011). The proposed work addresses the gaps in this classic problem by taking advantage of recent advances in transit routing.

The full gamut of public transportation planning necessitates understanding the role of demand elasticity that supply changes may induce, departure time choices, the effect of new-age mobility solutions such as ride-sharing and hailing services, impacts of real-time information, bundled multi-modal mobility options, etc. While important, these problems are excluded from the current scope because the algorithms for transit assignment are not computationally quick and mature enough to embed them within other larger mathematical models. This research will serve as a conduit for addressing such problems in the future.

## 2 Methodology

The first step of the proposed transit assignment model involves predicting the route choices of a single transit traveler. To this end, we use data from public sources and the Bengaluru Metropolitan Transport Corporation (BMTC), one of India's largest public transportation systems with over 6000 buses, 5000 routes, and 27,000 trips. This information is fused with land-use data to build a multinominal logit model that determines passenger preferences and weights for different journey attributes such as time, cost, and transfers. However, the application of these models requires a choice set. Existing literature uses simple heuristics that do not discover journeys that passengers might find on routing apps such as Google Maps. They are also computationally inefficient since they build time-expanded graphs and run a variant of Dijkstra's algorithm. Instead, we use state-of-the-art transit routing algorithms such as the multi-criteria RAPTOR (and McRAPTOR) Delling et al. (2015) to generate a set of Pareto-optimal journeys that form the choice set.

The basic version of RAPTOR finds Pareto-optimal journeys based on two criteria: arrival time and the number of transfers. It does not involve any preprocessing stage and works directly on the timetable network by storing the transit timetable in the form of trips. RAPTOR works in rounds, where each round is divided into two phases: a *route scan* and a *transfer scan*. The $k^{th}$ round computes arrival time at reachable stops with exactly $k-1$ transfers. Labels at every stop (initialized to infinity) are of the form: $(\tau_0(s), \tau_1(s), \ldots, \tau_k(s))$, where $\tau_k(s)$ denotes the earliest arrival time at stop $s$ with $k$ trips. A round $k$ involves the following two phases:

1. *route scan* phase: This phase starts by collecting all routes serving the stops whose labels were updated in the previous round (known as marked stops). Next, labels of the stops along the route are updated (if possible) using the earliest trip that can be boarded on it. A stop whose label is updated is added to the marked stop list.
2. *transfer scan* phase: This step involves scanning the footpaths (or walking links) originating from the stops present in the marked stop list. Stops whose arrival time is improved using these footpath connections are added to the marked stop list.

Likewise, Round $k+1$ starts its route scan phase using the marked stop list obtained from the transfer scan phase of Round $k$. The algorithm proceeds similarly and terminates when the maximum transfer limits are reached, or when no stop labels are updated. Stop labels can then generate Pareto-optimal journeys to all the stops using a quick post-processing step.

To find the efficiency frontier of an individual, we use McRAPTOR (an extension of RAPTOR, which can handle other objectives) with the following criteria: arrival time, number of transfers, in-vehicle travel time, walking time, and cost.

To address the second question, we build a digital twin of a transit system by simulating transit travelers, i.e., their origins, destinations, and departure times. To assign them to routes, we create a network loading process in which the logit model is employed to load travelers one by one to journeys on their respective efficiency frontiers. Figure describes the framework for the proposed methodology. McRAPTOR is used for *Choice set generation*. The passenger exits the system if no journey is possible for the given origin, destination, and departure time. Else, an MNL model is employed to select the preferred journey. If the passenger cannot board the preferred journey (due to overcrowding), the departure time is updated, and the choice set is generated again. Otherwise, passengers are assumed to board the bus until the next transfer stop. Operators can then use the post-processing function to answer crowding, congestion, and ridership-related questions.

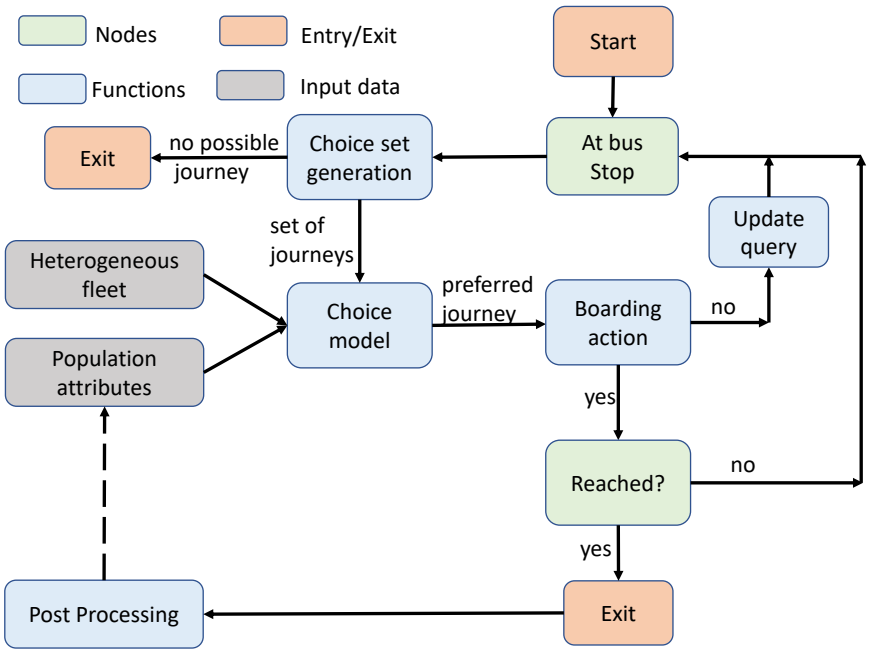



Figure 1: Framework used for scalable public transit assignment

# 3 Results and Future Work

All algorithms were implemented in C++ (with OpenMP parallelization) and compiled with optimization flags on a 128-core Intel Xeon Gold CPU clocked at 3.0 GHz with 512 GB RAM. The choice set is generated based on five criteria mentioned earlier. The maximum transfer limit was set to three. The input data can be classified into two categories: supply and demand side. Supply data consists of fleet and service characteristics of the transit agency – BMTC. Three types crush loads/capacity were assumed (60, 80, and 100). The demand data (i.e., origin, destination, and departure time) was generated synthetically using data from a four-step model.

Table 1 summarizes the results. The demand column shows the number of passengers simulated in the system. The next column shows the total runtime in seconds. The values in parenthesis show the total time required for choice set generation, choice model, and boarding action. The third column shows the number of iterations. An iteration comprises three steps: choice set generation + choice model application + boarding action. The maximum number of iterations was set to 200. Columns failure and reached show the total number of people who failed to reach and those who reached their destination, respectively. A passenger is said to have failed if the choice set generation results in no journeys (e.g., when no trip is available from the stop after the passenger's departure time). Queries that fail thus indicate the lack of public transit and such passengers are assumed to switch to personalized modes of travel. The last column shows the number of people who have not reached their destination within the limit of maximum iterations due to capacity restrictions . Figure 2 shows a heat map of stops belonging to overcrowded trips in the time intervals 6–7 PM, 7–8 PM, and 8–9 PM, respectively. A trip is classified as overcrowded if the volume-to-capacity ratio of 70% of its link is more than 0.8.

# 4 Future Work

The proposed framework assumed fixed capacity constraints, and updates routes in an online manner. It however fails to account for day-to-day learning resulting from crowded experiences. To overcome this issue, we plan to initialize subsequent rounds with an additional criterion — a penalty function that reflects the crowding levels. This criterion would prevent overcrowded trip segments from being part of Pareto-optimal journeys and thus would have reduced ridership in subsequent rounds. However, the solutions from this procedure can oscillate. To converge to a

| Demand | Runtime | Iterations | Failure | Reached | Denied |
|---|---|---|---|---|---|
| **100,000** | 393 (99+159+135) | 200 | 23,598 | 76,400 | 2 |
| **500,000** | 1767 (727+586+453) | 200 | 195,678 | 304,177 | 145 |
| **1,000,000** | 3520 (1,879+935+713) | 200 | 512,051 | 487,668 | 281 |

Table 1: Assignment Statistics (*Demand*: Number of passengers, *Runtime*: Total runtime (in seconds), *Iterations*: Total number of iterations required, *Failure*: Passengers who failed to board, *Reached*: Passengers who reached their destination, *Denied*: Passengers who are denied boarding and are still in the system)

(a) 6–7 PM (b) 7–8 PM (c) 8–9 PM

Figure 2: Stops belonging to overcrowded trips

steady state, we plan to experiment with two methods — (1) use the crowding criteria for only those travelers who were denied boarding or had to spend a significant duration of their journey in crowded buses, and (2) design penalty functions that get "damped" in subsequent iterations akin to step-sizes in stochastic approximation algorithms. The proposed methods will be validated using ticketing data which is available on a subset of route segments. This process will likely produce a feasible solution, but a notion of a stochastic user equilibrium is needed to characterize its properties. (We call it stochastic since a logit probability distribution governs the route choice.) Finally, the model will be tested to predict ridership from changes to routes and schedules (also called route rationalization) and discounted fare strategies using variable neighborhood searches. Since closed-form expressions for the operator's objectives cannot be conceived without making the model overly stylized, we plan to build a simulation-assisted heuristic that perturbs the current supply levels using different exchange/randomization techniques.

**Acknowledgments:** The authors would like to thank interns – Aditya Shankar Garg and Kalash Abhay Kala for initial assistance with setting up parts of the code.

# References

Delling, D., T. Pajor, and R. F. Werneck (2015). Round-Based Public Transit Routing. *Transportation Science 49*(3), 591–604.

Hamdouch, Y., H. Ho, A. Sumalee, and G. Wang (2011). Schedule-based transit assignment model with vehicle capacity and seat availability. *Transportation Research Part B: Methodological 45*(10), 1805–1830.

Hamdouch, Y. and S. Lawphongpanich (2008). Schedule-based transit assignment model with travel strategies and capacity constraints. *Transportation Research Part B: Methodological 42*(7-8), 663–684.

Jiang, Y. and W. Szeto (2016). Reliability-based stochastic transit assignment: Formulations and capacity paradox. *Transportation Research Part B: Methodological 93*, 181–206.

Schmöcker, J.-D., A. Fonzone, H. Shimamoto, F. Kurauchi, and M. G. Bell (2011). Frequency-based transit assignment considering seat capacities. *Transportation Research Part B: Methodological 45*(2), 392–408.

# Modeling Multi-modal Curbside Usage in Dynamic Networks

Jiachao Liu[a], Sean Qian[a,b]

[a]*Department of Civil and Environmental Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, United States*
[b]*Heinz College, Carnegie Mellon University, Pittsburgh, PA 15213, United States*

*Keywords:* Dynamic traffic assignment, multi-modal dynamic user equilibrium, curbside management, mesoscopic traffic simulation

## 1. Introduction

The explosive growth of ride-hailing (RH) services, the ubiquity of e-commerce with on-demand deliveries, and the rise of micro-mobility services have transformed curb spaces into more valuable public infrastructure for which multi-modal transportation fights. Curb utilization by various traffic modes keeps increasing pressure on the already oversaturated curbs and transportation system, further altering travelers' choices in routes, clogging streets, and polluting the urban environment. Therefore, it is critical to analyze the system-level impacts associated with curb utilization by multi-modal transportation and to better understand the interactions among each curb user to facilitate the optimal management of this scarce public infrastructure.

There exist many studies on building dynamic traffic assignment (DTA) frameworks to model multi-modal travelers' behaviors. One can refer to the literature review in Pi et al. (2019). Notably, some researchers focus on modeling parking-related driving behaviors in the dynamic network, such as on-street parking, cruising, and corresponding pricing strategies (Arnott et al. 1991, Qian et al. 2012, Zheng & Geroliminis 2016, Leclercq et al. 2017, Gu et al. 2020, 2021). However, the research gap lies in that most models do not simultaneously consider various types of stops and occupancies at curb spaces by multiple modes, especially the location choice of stops near curbside spaces. Curb spaces are not only used by private vehicles for on-street parking but also commercial RH services and commercial trucks to make stops for picking up/dropping off (PU/DO) passengers and loading/unloading freights. Each mode has a profound impact regarding when, where, and how travelers make trips. To better understand the interactions between different curb users, travelers' choice behaviors related to curb utilization should be embedded in the framework, and this helps to predict how the travel pattern of curb users will change spatially and temporally with the implementation of various curb management policies such as dynamic pricing and space allocation. On the other hand, dynamic network loading (DNL) plays a vital role in the simulation-based DTA framework, but to our best knowledge, none of the existing DNL models integrates spatial usage of curbs into network models and is not able to model the dynamics of curb space utilization of vehicles and the effect of potential curb/road congestion (i.e., double parking) on the network traffic.

Given this, our study proposes a curb-aware network modeling framework that encapsulates the route choice, competition, and interactive effect among different curb users simultaneously and embeds the dynamics of curb usage and the associated network impact into the mesoscopic DNL model. The contributions are summarized as follows:

- The proposed DTA framework models the travelers' choice behaviors related to curb usage. A multi-modal dynamic user equilibrium model (DUE) is formulated among three modes: private driving, ride-hailing, and commercial trucks. We model the choices of parking locations for driving, curb PU/DO locations for ride-hailing, and curb loading/unloading locations for commercial trucks simultaneously. Curbside pricing and space restrictions are also embedded in the generalized cost of different modes. It offers a holistic approach to assessing how travelers' choices will be changed and how system-level performance will be impacted when implementing curbside management strategies.
- Refined curb space searching and usage dynamics are integrated into the mesoscopic DNL model for estimating dynamic traffic conditions (i.e., travel time). This enables us to relate the curb usage condition (i.e., occupancy) with the traffic flow dynamics to encapsulate the dynamic traffic effect of curbside stops. An increasing number of curbside stops will lead to potential curbside congestion due to the limited capacity of curbside spaces, further influencing traffic on the driving lanes. The increased travel time caused by curb congestion is embedded in the utility functions of different modes of interest and further impacts the route choice, stop location choice of ride-hailing and commercial trucks, and parking location choice of

*Email addresses:* jiachaol@andrew.cmu.edu (Jiachao Liu), seanqian@cmu.edu (Sean Qian)

driving. In addition, it offers an approach to examine how flow dynamics will be impacted if curb capacity is changed due to space regulation.

- Real-world curb usage data is used to calibrate the proposed framework together with other multi-source traffic sensor data. The parameters of the proposed link dynamic model with curb usage can be calibrated by the classical traffic data and emerging curb usage data which identifies individual vehicle arrivals and departures on curb spaces. A framework of dynamic origin-destination demand estimation (DODE) is constructed through computational graphs using the muti-source data.

## 2. Model Formulation

### 2.1. Preliminaries

Consider a general roadway network denoted by $G = (N, A)$, which consists of a node set $N$ and a link set $A$. Define $T_d = [1, 2, ..., T]$ as the assignment horizon and time $t \in T_d$.$R$ and $S$ are sets of origin nodes and destination nodes. For each O-D pair $rs$, $r \in R$ and $s \in S$, there are three categories of demand, represented by three modes, departing at time $t$: private driving demand $q^{rs}_{PD,t}$, RH demand $q^{rs}_{RH,t}$ and commercial truck demand $q^{rs}_{CT,t}$. It is assumed no mode choice is considered in this study. Between each O-D pair $rs$, travelers of each mode make choices of routes associated with stop location choice on curb spaces. The problem of interest, which is the goal of the DTA problem, is to allocate spatial-temporal flows $f^{rs}_{M,k,t}$ of different modes considering the effect of curb utilization on the multi-modal network given O-D demand $q^{rs}_{M,t}$.

We model the route choice associated with curb stop location choice separately for different modes. Private driving travelers only stop once and can choose to park at their destinations or on curb spaces nearby, namely parking location choice. The choice is embedded in the route sets used. Each route between an OD pair for private driving mode is denoted by either parking at the destination or a curb near the destination. Each route has a corresponding generalized travel cost defined in Section 2.2.

For ride-hailing vehicles and commercial trucks, they may have multiple successive stops for PU/DO or load/unloading. To model this, we use definition of tour to represent the whole route between origin and destination node. Each tour consists of several successive passenger or commercial delivery trips and each trip starts from curb space around an intermediate destination and ends at another curb space. In our study, the generated routes for ride-hailing and trucks are in the tour format with predefined intermediate destinations between each OD pair. Each tour has a generalized travel cost, and vehicles choose between tours.

### 2.2. Dynamic User Equilibrium

In this study, we assume no mode choice is considered, and travelers in the transportation system are fully aware of traffic conditions. The generalized travel cost of three modes is defined and used to govern route choices following user equilibrium.

A driving trip must start from an origin node $r \in R$ and end at either a destination node $s \in S$ (parking at the garage at destination) or curb spaces nearby (on-street parking or off-street parking at curbs), and these two options might have different parking fees. The generalized cost of a driving trip is defined as

$$c^{rs}_{PD,k,t} = \max\left[\gamma(t + w^{rs}_{k,t} - t^{*,rs}_{PD}), \beta(t^{*}_{PD} - t - w^{rs}_{k,t})\right] + \alpha w^{rs}_{k,t} + p_c\sigma_c + (1 - \sigma_c)p_s + \xi,\ \forall k \in P^{rs}_{PD} \tag{1}$$

where the first term represents the cost of arriving early or late in a general setting. $t^{*,rs}_{M}$ denotes the expected arrival time. $\gamma$ and $\beta$ are the unit cost of arriving early or late. $w^{rs}_{k,t}$ denotes the travel time on path $k$ obtained from the dynamic network loading. $\sigma_c$ is the indicator variable and equals one if the driving traveler parks the car at the curb space, i.e., on-street parking, and will pay a parking fee of $p_c$, otherwise $\sigma_c = 0$ when he/she parks the car right at the destination and pay a parking fee of $p_s$. $\xi$ is the other cost for the driving mode.

A ride-hailing trip starts from a curb space around an intermediate destination $r \in R'$ (passenger PU) and ends at a curb space around another intermediate destination $s \in S'$ (passenger DO), and it is a fraction of ride-hailing tour. PU and DO will use curb spaces to perform quick stops and might have location-based curb use fees. The generalized cost of a ride-hailing trip is defined as

$$c^{rs}_{RH,k,t} = \max\left[\gamma(t + w^{rs}_{k,t} - t^{*,rs}_{RH}), \beta(t^{*}_{RH} - t - w^{rs}_{k,t})\right] + \alpha w^{rs}_{k,t} + \rho^{rs}_{k,t}(d^{rs}_{k,t}, w^{rs}_{k,t}) + p^{PU}_{c} + p^{DO}_{c} + \alpha l^{rs}_{k},\ \forall k \in P^{rs}_{RH} \tag{2}$$

where the first term represents the cost of arriving early or late and $w^{rs}_{k,t}$ denotes the travel time similar to Equation 1. $\rho^{rs}_{k,t}$ is the service fee function related to the distance $d^{rs}_{k,t}$ and travel time $w^{rs}_{k,t}$ of trip $k \in P^{rs}_{RH}$. $p^{PU}_{c}$ and $p^{DO}_{c}$ are pick-up and drop-off curb use fees respectively. $l^{rs}_{k}$ is walking distance from intermediate destinations to PUDO curbs.

Different from driving and ride-hailing, the generalized cost of a commercial truck is tour-based, meaning that a truck driver makes route choices considering travel cost of all trips in a tour simultaneously, defined in Equation 3.

$$c^{rs}_{CT,k,t} = \max\left[\gamma(t + w^{rs}_{k,t} - t^{*,rs}_{CT}), \beta(t^{*}_{CT} - t - w^{rs}_{k,t})\right] + \alpha w^{rs}_{k,t} + p_f d^{rs}_{k,t} + \sum_{i}^{N_k} p^{i}_{c},\ \forall k \in P^{rs}_{CT} \tag{3}$$

where $p_f d^{rs}_{k,t}$ is distance-related fuel cost for truck mode and $p^i_c$ is $i$-th curb use fee and total $N_k$ times of curb deliveries.

The user equilibrium conditions of the whole system are

$$\begin{aligned}
&c^{rs}_{PD,k,t} - \mu^{rs}_{PD,k,t} = 0 \ \text{ if } \ \forall k \in P^{rs}_{PD}, r, s \in R, S, f^{rs}_{PD,k,t} > 0 \\
&c^{rs}_{PD,k,t} - \mu^{rs}_{PD,k,t} \geq 0 \ \text{ if } \ \forall k \in P^{rs}_{PD}, r, s \in R, S, f^{rs}_{PD,k,t} = 0 \\
&c^{rs}_{RH,k,t} - \mu^{rs}_{RH,k,t} = 0 \ \text{ if } \ \forall k \in P^{rs}_{RH}, r, s \in R', S', f^{rs}_{RH,k,t} > 0 \\
&c^{rs}_{RH,k,t} - \mu^{rs}_{RH,k,t} \geq 0 \ \text{ if } \ \forall k \in P^{rs}_{RH}, r, s \in R', S', f^{rs}_{RH,k,t} = 0 \\
&c^{rs}_{CT,k,t} - \mu^{rs}_{CT,k,t} = 0 \ \text{ if } \ \forall k \in P^{rs}_{CT}, r, s \in R, S, f^{rs}_{CT,k,t} > 0 \\
&c^{rs}_{CT,k,t} - \mu^{rs}_{CT,k,t} \geq 0 \ \text{ if } \ \forall k \in P^{rs}_{CT}, r, s \in R, S, f^{rs}_{CT,k,t} = 0
\end{aligned} \tag{4}$$

Note that in Equation 4, the path flow of ride-hailing mode is between intermediate destinations $r', s' \in R', S'$ (i.e., path flow of trips), based on the definitions in Section 2.1, we can map path flow of trips to path flow of tours. Here we do not consider the passenger-service mapping problem and assume all passengers who choose ride-hailing services can be served. Consider an arbitrary tour of ride-hailing mode $u \in U^{rs}_{RH}$ where $r, s \in R, S$ and it consists of $N_u$ successive trips. Consider a trip $i \in [N_u]$ starting and ending at $r'_i$ and $s'_i$ respectively. Between the intermediate destination pair $r'_i s'_i$, there are $K^{r'_i s'_i}$ routes with flow of $f^{r'_i s'_i}_{RH,k,t}$ where $k \in [K^{r'_i s'_i}]$. These path flows should follow the UE condition defined in Equation 4 with constraints of

$$\sum_{k \in K_i} p^{r'_i s'_i}_{RH,k,t} = 1, \ p^{r'_i s'_i}_{RH,k,t} = \frac{f^{r'_i s'_i}_{RH,k,t}}{q^{r'_i s'_i}_{RH,t}} \tag{5}$$

The path flow of tour $u$ between OD pair $rs$ will be

$$\begin{aligned}
f^{rs}_{RH,u,t} &= \prod_i p^{r'_i s'_i}_{RH,n(i),t'} \cdot q^{rs}_{RH,t} \\
t' &= t + \sum_{m=0}^{i-1} w^{r'_i s'_i}_m
\end{aligned} \tag{6}$$

where $n(i)$ is a function mapping trip indicator $i$ to the corresponding indicator in the route set between intermediate destination pair $r'_i s'_i$. $w^{r'_i s'_i}_m$ is the travel time of trip $m$. We use $\psi$ to denote the mapping function defined in Equation 5 and 6.

The equivalent VI problem is formulated as follows

$$\begin{aligned}
&\text{Find } f^{*rs}_{PD,k,t}, f^{rs}_{RH,k,t}, f^{rs}_{CT,k,t} \text{ such that} \\
&\sum_{r,s \in R,S} \sum_{t \in T_d} \left\{ \sum_{k \in P^{rs}_{PD,t}} c^{rs}_{PD,k,t}(f^{*rs}_{PD,k,t}) \cdot \left( f^{rs}_{PD,k,t} - f^{*rs}_{PD,k,t} \right) + \sum_{k \in P^{rs}_{CT,t}} c^{rs}_{CT,k,t}(f^{*rs}_{CT,k,t}) \cdot \left( f^{rs}_{CT,k,t} - f^{*rs}_{CT,k,t} \right) \right\} + \sum_{r',s' \in R,S} \sum_{t \in T_d} \sum_{k \in P^{r's'}_{PD,t}} c^{r's'}_{RH,k,t}(f^{*r's'}_{RH,k,t}) \cdot \left( f^{r's'}_{RH,k,t} - f^{*r's'}_{RH,k,t} \right) \geq 0 \\
&\text{where } \sum_{k \in P^{rs}_{M,t}} f^{rs}_{M,k,t} = q^{rs}_{M,t}. \ r, s \in R, S, \ t \in T_d, \ M \in \{PD, RH, CT\}. \\
&f^{rs}_{RH,k,t} = \psi(f^{r's'}_{RH,k',t})
\end{aligned} \tag{7}$$

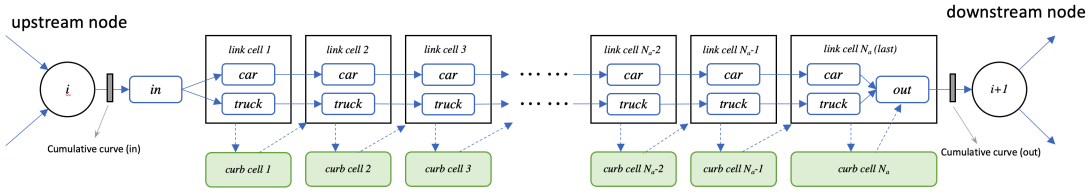
### 2.3. Dynamic Network Loading Considering Curb Space Usage

Mesoscopic dynamic network loading (DNL) integrates models encapsulating traffic flow propagations over nodes and links and output link-based cumulative flow curves, which record the number of arriving and departing vehicles on each link. In our study, curbside stopping behaviors for different vehicles are integrated into the link dynamics of DNL, and the impact will be captured since the output of traffic conditions is related to curb usage. The link flow dynamics model in this study extends the multi-class traffic flow model proposed in Qian et al. (2017), which generalizes the cell transmission model (CTM) and considers two vehicle classes with different traffic flow characteristics. Each link is divided into cells and a curb cell is introduced for each link cell to represent the curbside space to model the curb space usage shown in Figure 1. Each curb cell is used to store vehicles stopping in the curb space or double parking on the side lane. The capacity of curb space in curb cell $i$ of link $i$ is denoted by $\kappa^i_a$, and the number of parking vehicles at time $t$ is denoted as $\nu^i_a(t)$. Hence the occupancy of curb cell $i$ is $o^c_a(i,t) = \nu^i_a(t)/\kappa^i_a$. Similarly, we can define the occupancy of the side lane used for double parking, denoted by $o^d_a(i,t)$. The last cell of each link also stores private driving vehicles which park at the corresponding off-street parking lot, and these vehicles are assumed to have no impact on the link flow dynamics (i.e., no impact on the FDs of cells).

A predefined curb space searching process for different users within a link updates the number of stopping vehicles at each curb cell used to compute $o^c_a(i,t)$ and $o^d_a(i,t)$. Each vehicle will be assigned a predefined aggressive parameter $\theta \in (0,1)$ to indicate

the maximal stop-checking location on the curb link. When entering the destination link, it checks the curb space availability starting from the first curb cell to the last checking location determined by $floor(\theta \cdot \text{cell num})$. If any cell has available space, it joins the curb stopping array at the cell and records the start time. If reaching the maximal checking location and still no space, it will double park at the current curb cell.

Figure 1: Cell transmission model with curb usage




We use a pragmatic approach with a truncated fundamental diagram (FD) in which we treat the curb-stopping effect as partial lane closure and relate the FD change with curb occupancy. If curb cell $i$ has stopping vehicles at curb space, the FD used will be truncated by reducing the flow capacities, critical densities, and jam densities for separate classes. The reduction scalar of cell $i$ on link $a$ at time $t$, denoted by $\eta_a(i,t)$ is determined by the curb occupancy $o_a^c(i,t)$ and occupancy of the double parking lane $o_a^d(i,t)$ of the cell.

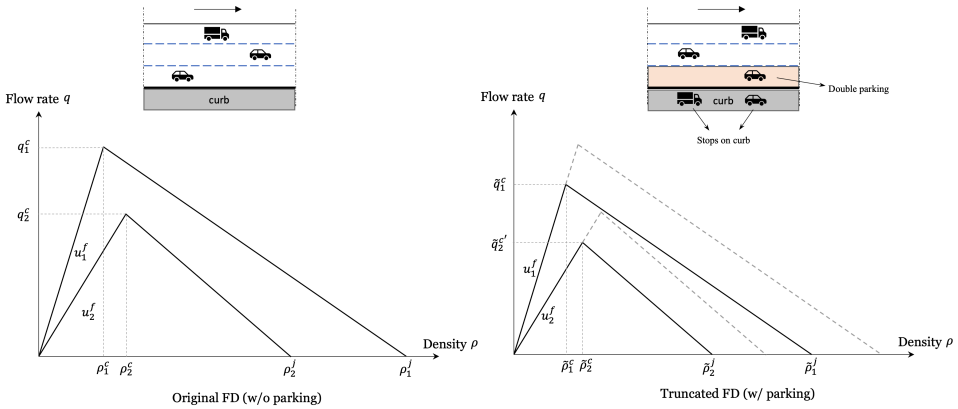
$$\eta_a(i,t) = \frac{n_a - \tau_c o_a^c(i,t) - \tau_d o_a^d(i,t)}{n_a} \tag{8}$$

where $n_a$ is the lane number of link $a$. $\tau$ is the effect scalar of curbside stops and double parking, and the choice should satisfy $\tau_c o_a^c(i,t) \leq 1$, $\tau_d o_a^d(i,t) \leq 1$ and $n_a - \tau_c o_a^c(i,t) - \tau_d o_a^d(i,t) \geq 0$. The truncated FD has the following reduced parameters

$$\tilde{q}_{m,a}^c(i,t) = \eta_a(i,t) q_{m,a}^c(i,t), \quad \tilde{\rho}_{m,a}^c(i,t) = \eta_a(i,t) \rho_{m,a}^c(i,t), \quad \tilde{\rho}_{m,a}^j(i,t) = \eta_a(i,t) \rho_{m,a}^j(i,t) \tag{9}$$

The illustration of FDs before and after truncation is shown in Figure 2. The flow flux between cells is computed using adjacent cells' corresponding FD parameters. If cell $i$ has curb stops and parameters of FD are truncated, the flow flux entering and leaving this cell may be smaller before truncation; hence link travel time might be increased. Detailed proofs can be found in the full paper. In this way, link dynamics are related to curb conditions, and curb usage might contribute to traffic congestion on the link. For each link, cumulative curves are built by counting and adding up the number of arriving and departing vehicles that do not stop at the curb; hence first-in-first-out (FIFO) is ensured. One can refer to Pi et al. (2019) for more details of dynamic models in our mesoscopic simulation.

Figure 2: Illustration of truncated FD considering curb occupancy




## 3. Dynamic Origin-destination demand estimation (DODE)

The DODE problem aims to minimize the differences between observations (observed traffic flow $\boldsymbol{y}$, observed travel time $\boldsymbol{z}$, observed curbside arrival $\boldsymbol{a}$ and departure $\boldsymbol{d}$) from real-world detectors and mesoscopic models with the constraints of DTA

framework which depicts travels' behaviors in the network, shown in Equation 10 in a vectorized form.

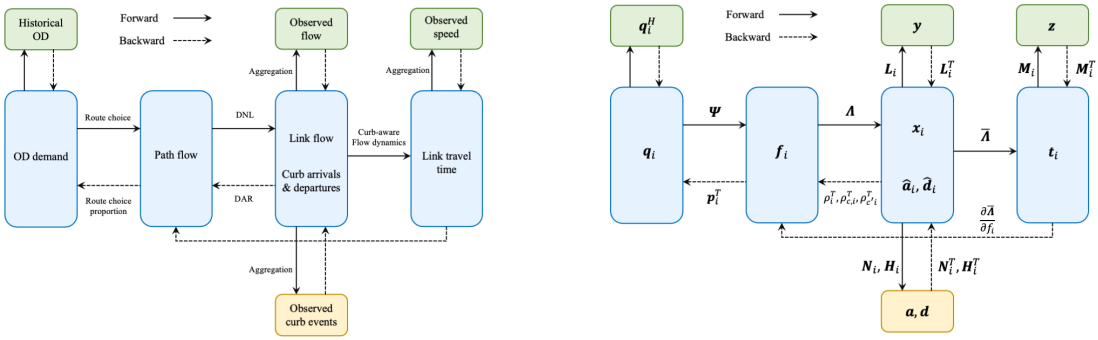
$$
\begin{aligned}
\min_{\{\boldsymbol{q}_i\}_i} \quad \mathscr{L} &= w_1 \underbrace{\left(\left\|\boldsymbol{y} - \sum_{i\in D} \boldsymbol{L}_i \boldsymbol{x}_i\right\|_2^2\right)}_{\mathscr{L}_1} + w_2 \underbrace{\left(\left\|\boldsymbol{z} - \sum_{i\in D} \boldsymbol{M}_i \boldsymbol{h}_i\right\|_2^2\right)}_{\mathscr{L}_2} + w_3 \underbrace{\left(\left\|\boldsymbol{a} - \sum_{i\in D} \boldsymbol{N}_i \boldsymbol{a}_i\right\|_2^2\right)}_{\mathscr{L}_3} + w_4 \underbrace{\left(\left\|\boldsymbol{d} - \sum_{i\in D} \boldsymbol{H}_i \boldsymbol{d}_i\right\|_2^2\right)}_{\mathscr{L}_4} \\
&= w_1 \underbrace{\left(\left\|\boldsymbol{y} - \sum_{i\in D} \boldsymbol{L}_i \boldsymbol{\rho}_i \boldsymbol{p}_i \boldsymbol{q}_i\right\|_2^2\right)}_{\mathscr{L}_1} + w_2 \underbrace{\left(\left\|\boldsymbol{z} - \sum_{i\in D} \boldsymbol{M}_i \bar{\Lambda}\left(\boldsymbol{\rho}_i \boldsymbol{p}_i \boldsymbol{q}_i\right)\right\|_2^2\right)}_{\mathscr{L}_2} + w_3 \underbrace{\left(\left\|\boldsymbol{a} - \sum_{i\in D} \boldsymbol{N}_i \boldsymbol{\rho}_{a,i} \boldsymbol{p}_i \boldsymbol{q}_i\right\|_2^2\right)}_{\mathscr{L}_3} + w_4 \underbrace{\left(\left\|\boldsymbol{d} - \sum_{i\in D} \boldsymbol{H}_i \boldsymbol{\rho}_{d,i} \boldsymbol{p}_i \boldsymbol{q}_i\right\|_2^2\right)}_{\mathscr{L}_4} \\
\text{s.t.} \quad & \{\boldsymbol{h}_i, \boldsymbol{\rho}_i, \boldsymbol{\rho}_{a,i}, \boldsymbol{\rho}_{d,i}\}_i = \Lambda(\{\boldsymbol{f}_i\}_i) \\
& \boldsymbol{f}_i = \boldsymbol{p}_i \boldsymbol{q}_i \\
& \boldsymbol{p}_i = \Psi_i(\{\boldsymbol{c}_i\}_i, \{\boldsymbol{h}_i\}) \\
& \{\boldsymbol{c}_i\} = \mu(\{\boldsymbol{h}_i\}, \boldsymbol{c}^e) \\
& \boldsymbol{q}_i \geq 0, \forall i \in D
\end{aligned}
\tag{10}
$$

We use three different DAR matrices $\boldsymbol{\rho}_i$, $\boldsymbol{\rho}_{a,i}$ and $\boldsymbol{\rho}_{d,i}$ to map multi-class path flow $\boldsymbol{f}_i$ to link flow $\boldsymbol{x}_i$, curb arrival flow $\boldsymbol{a}_i$ and curb departure flow $\boldsymbol{d}_i$. The problem is constructed and solved on a computational graph shown in Figure 3. In the forward process, multi-class OD demand $\boldsymbol{q}_i$ is used as the input and the DTA problem is solved based on the DNL model $\Lambda$ and route choice model $\Psi$. Route choice portions $\boldsymbol{p}_i$, DAR matrices $\boldsymbol{\rho}$, $\boldsymbol{\rho}_{a,i}$ and $\boldsymbol{\rho}_{d,i}$, and link travel time $\boldsymbol{h}_i$ are obtained and loss function can be computed. In the backward process, the gradients are computed for each equation in the forward process and the gradient of loss function with respect to multi-class demand $\boldsymbol{q}_i$ is

$$
\begin{aligned}
\frac{\partial \mathscr{L}}{\partial \boldsymbol{q}_i} = & -2w_1 \boldsymbol{p}_i^T \boldsymbol{\rho}_i^T \boldsymbol{L}_i^T \left(\boldsymbol{y} - \sum_{i'\in D} \boldsymbol{L}_{i'} \boldsymbol{\rho}_{i'} \boldsymbol{p}_{i'} \boldsymbol{q}_{i'}\right) - 2w_2 \boldsymbol{p}_i^T \frac{\partial \Lambda_i(\{\boldsymbol{f}_i\}_i)}{\partial \boldsymbol{f}_i} \boldsymbol{M}_i^T \left(\boldsymbol{z} - \sum_{i'\in D} \boldsymbol{M}_{i'} \boldsymbol{h}_{i'}\right) \\
& - 2w_3 \boldsymbol{p}_i^T \boldsymbol{\rho}_{c,i}^T \boldsymbol{N}_i^T \left(\boldsymbol{a} - \sum_{i'\in D} \boldsymbol{N}_{i'} \boldsymbol{\rho}_{c,i'} \boldsymbol{f}_{i'}\right) - 2w_4 \boldsymbol{p}_i^T \boldsymbol{\rho}_{c',i}^T \boldsymbol{H}_i^T \left(\boldsymbol{d} - \sum_{i'\in D} \boldsymbol{H}_{i'} \boldsymbol{\rho}_{c',i'} \boldsymbol{f}_{i'}\right)
\end{aligned}
\tag{11}
$$

where The term $\frac{\partial \Lambda_i(\{f_i\}_i)}{\partial f_i}$ is the derivative of link travel time with respect to path flow and can be computed following the works of calculating path marginal cost (PMC). Details can be found in Qian & Zhang (2011).

Figure 3: Computational graph illustration




(a) Computational graph structure

(b) Computational graph for all variables

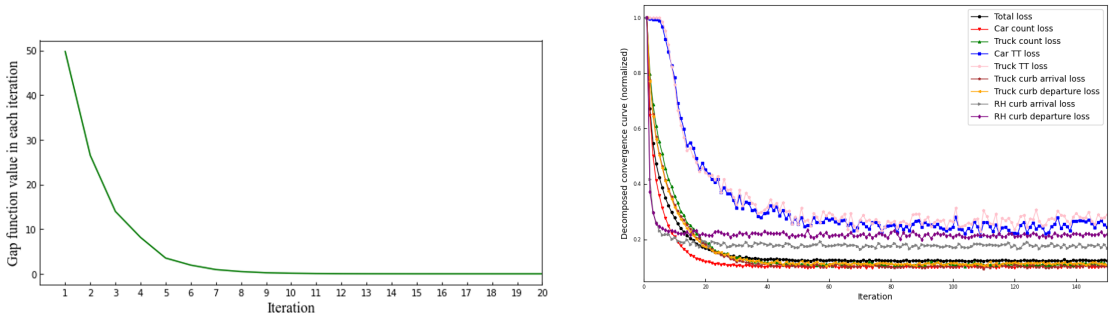
## 4. Numerical experiments

Two networks are used to test our proposed DUE and DODE framework: the Braess network and the real-world Pittsburgh network. The convergence performance of DUE and DODE on the Braess network can be seen in Figure 4. The simulated traffic conditions matches observed conditions with a satisfactory accuracy, shown in Figures 5 and 6.

## 5. Acknowledgement

This research is supported by a National Science Foundation grant CMMI-1931827, and a Department of Energy grant DOE-EE0009659.

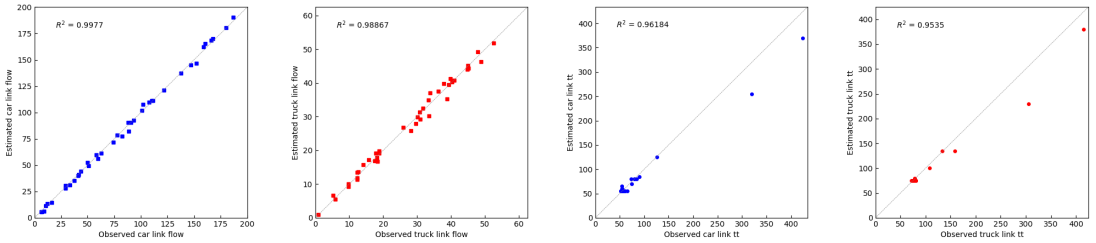
Figure 4: Convergence performance on Braess network




(a) Convergence performance of DUE

(b) Convergence performance of DODE

Figure 5: Estimated vs Observed link-level traffic flow and travel time



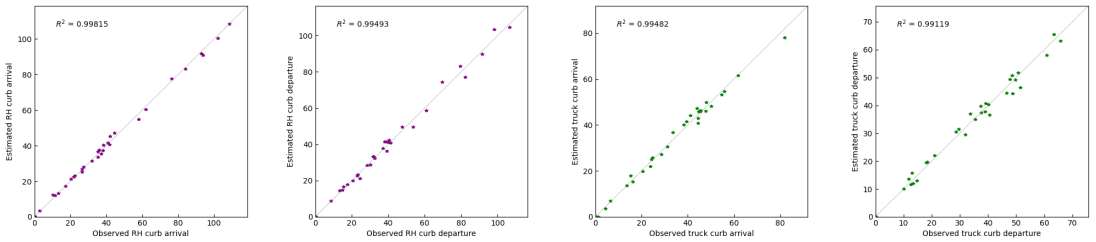

(a) Car link flow (vehs/15min)

(b) Truck link flow (vehs/15min)

(c) Car link travel time (secs)

(d) Truck link travel time (secs)

Figure 6: Estimated vs Observed curb arrivals and departures




(a) Car curb arrival flow (vehs/15min)

(b) Car curb departure flow (vehs/15min)

(c) Truck curb arrival flow (vehs/15min)

(d) Truck curb departure flow (vehs/15min)

## References

Arnott, R., de Palma, A. & Lindsey, R. (1991), 'A temporal and spatial equilibrium analysis of commuter parking', *Journal of Public Economics* **45**(3), 301–335.

**URL:** *https://www.sciencedirect.com/science/article/pii/0047272791900306*

Gu, Z., Najmi, A., Saberi, M., Liu, W. & Rashidi, T. H. (2020), 'Macroscopic parking dynamics modeling and optimal real-time pricing considering cruising-for-parking', *Transportation Research Part C: Emerging Technologies* **118**, 102714.

Gu, Z., Safarighouzhdi, F., Saberi, M. & Rashidi, T. H. (2021), 'A macro-micro approach to modeling parking', *Transportation Research Part B: Methodological* **147**, 220–244.

**URL:** *https://www.sciencedirect.com/science/article/pii/S019126152100059X*

Leclercq, L., Sénécat, A. & Mariotte, G. (2017), 'Dynamic macroscopic simulation of on-street parking search: A trip-based approach', *Transportation Research Part B: Methodological* **101**, 268–282.

Pi, X., Ma, W. & Qian, Z. S. (2019), 'A general formulation for multi-modal dynamic traffic assignment considering multi-class vehicles, public transit and parking', *Transportation Research Part C: Emerging Technologies* **104**, 369–389.

Qian, Z. S., Li, J., Li, X., Zhang, M. & Wang, H. (2017), 'Modeling heterogeneous traffic flow: A pragmatic approach', *Transportation Research Part B: Methodological* **99**, 183–204.

Qian, Z. S., Xiao, F. E. & Zhang, H. (2012), 'Managing morning commute traffic with parking', *Transportation Research Part B: Methodological* **46**(7), 894–916.

**URL:** *https://www.sciencedirect.com/science/article/pii/S019126151200015X*

Qian, Z. & Zhang, H. M. (2011), 'Computing individual path marginal cost in networks with queue spillbacks', *Transportation Research Record* **2263**(1), 9–18.

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.

# Session 5

Queue replacement principle for corridor problems with heterogeneous commuters
*Takara Sakai, Takashi Akamatsu and Koki Satsukawa*

Lane Choice Equilibrium in Macroscopic Multi-Queue Node Model for Multilane Intersections
*Xiaolin Gong, Michiel C.J. Bliemer and Mark P.H. Raadsen*

The Challenge of Computing the Cell Transmission Model Loading with Exact First-In-First-Out
*Genaro Jr. Peque and Hillel Bar-Gera*

Dynamic Traffic Assignment with Physics-Informed Deep Learning
*Shakiba Naderian, Ohay Angah, Yiran Zhang and Xuegang Ban*

# Queue replacement principle for corridor problems with heterogeneous commuters*

Takara Sakai[a,*], Takashi Akamatsu[a], and Koki Satsukawa[a,*]

[a]*Graduate School of Information Sciences, Tohoku University, Japan*
**Corresponding authors*
*E-mail: takara.sakai.t1@dc.tohoku.ac.jp, satsukawa@tohoku.ac.jp*

## 1 Introduction

The single-bottleneck model proposed by Vickrey (1969) describes morning commute traffic as departure time choice equilibrium. The corridor problem is a spatial extension of this model, which considers multiple tandem bottleneck networks and describes the spatial distribution and interaction of the congestion. However, in the corridor problem, complex case divisions rapidly increase with the number of bottlenecks when the equilibrium condition is directly analyzed, like most previous analyses of the single-bottleneck case, as noted in Arnott and DePalma (2011). Hence, the number of systematic analyses of the corridor problem with an arbitrary number of bottlenecks is very limited; consequently, there are few theoretical findings regarding this problem, unlike the single-bottleneck model (Li et al. 2020).

To address this issue, we propose an indirect analysis approach focusing on one of the remarkable properties of the single-bottleneck model, which we refer to as the *queue replacement principle (QRP)*. The QRP means that the queuing delay pattern in the dynamic user equilibrium (DUE) state is equal to the optimal pricing pattern in the dynamic system optimal (DSO) state (Arnott 1998, Vickrey 1969). This principle enables us to derive the DUE queueing and flow pattern indirectly using the optimal pricing pattern, which is the solution to the linear programming (DSO problem) (Akamatsu et al. 2021, Iryo and Yoshii 2007). If the QRP holds in the corridor networks, the theoretical relationships between the DSO and DUE states can be clearly understood, and systematic theoretical findings for the corridor problem can be obtained.

This paper clarifies the condition under which the QRP holds for corridor problems with heterogeneous commuters with different values of schedule delay. As might be inferred from the QRP concept, this condition is equivalent to the existence condition of the flow pattern that does not contradict the DUE condition when the optimal pricing pattern replaces the queuing delay pattern. To clarify the existence condition mathematically, we first derive the closed-form optimal pricing pattern. We then substitute the analytical solution into the DUE problem and analyze the feasibility of the DUE flow to clarify the condition where the QRP holds.

## 2 DSO and DUE problems

We consider the freeway corridor network with $N$ origins and single destination shown in Figure 1. Each link $i \in \mathcal{N} \equiv \{1, \ldots, N\}$ connecting from the origin $i$ has a bottleneck with a

**Keywords:* departure time choice problem, corridor networks, heterogeneous value of schedule delay, dynamic user equilibrium, dynamic system optimum

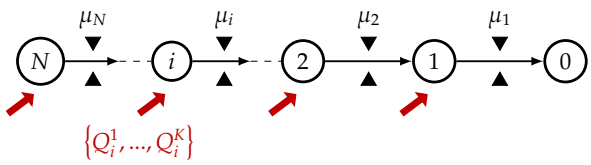



**Figure 1:** Corridor network.

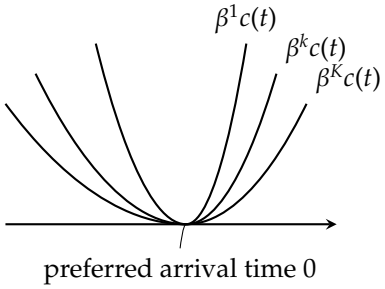



**Figure 2:** Schedule delay cost functions.

finite capacity $\mu_i$. We assume that $\mu_i < \mu_{i-1}$, $\forall i \in \mathcal{N} \setminus \{1\}$ for the sake of simplicity. The queue evolution and associated queueing delay are modeled by the standard point queue model with the first-in-first-principle.

Continuum commuters travel from every origin to the destination during the morning rush-hour $\mathcal{T}$. They are classified into $K$ groups based on the values of the schedule delay (Figure 2). The schedule delay cost function of the commuters in a group $k \in \mathcal{K} \equiv \{1, \dots, K\}$ is represented as $\beta^k c(t)$, where $c(t)$ is a convex function whose value becomes zero when $t = 0$ (preferred arrival time); $\beta^k \in (0, 1]$ is a parameter representing the heterogeneity of commuters and satisfies the following relationship: $\beta^K < ... < \beta^1 = 1$. The mass of commuters in group $k$ departing from origin $i$ referred to as $(i,k)$-commuters) is denoted by $Q_{i,k}$. We denote by $q_{i,k}(t)$ an arrival flow of $(i,k)$-commuters whose destination arrival time is $t$, and $q_i(t) = \sum_{k \in \mathcal{K}} q_{i,k}(t)$ an aggregated arrival flow with respect to the origin $i$.

We formulate the DSO and DUE problems in the Lagrangian-like coordinate system to avoid the complex nested structure of commuters' trip costs (for a detailed discussion, see Akamatsu et al. 2015). In this system, the DSO problem is formulated as a linear programming (LP), which determines the arrival (departure) flow pattern that minimizes the total trip cost without queues (Fu et al. 2022, Osawa et al. 2018). Its optimal conditions are described as follows:

$$0 \le q_{i,k}^{\mathrm{SO}}(t) \perp \left\{ \sum_{j;j \le i} p_j^{\mathrm{SO}}(t) + \beta^k c(t) - \rho_{i,k}^{\mathrm{SO}} \right\} \ge 0 \qquad \forall k \in \mathcal{K}, \quad \forall i \in \mathcal{N}, \quad \forall t \in \mathcal{T}, \tag{1}$$

$$0 \le p_i^{\mathrm{SO}}(t) \perp \left\{ \mu_i - \sum_{j;j \ge i} q_j^{\mathrm{SO}}(t) \right\} \ge 0 \qquad \forall i \in \mathcal{N}, \quad \forall t \in \mathcal{T}. \tag{2}$$

$$0 \le \rho_{i,k}^{\mathrm{SO}} \perp \left\{ \int_{\mathcal{T}} q_{i,k}^{\mathrm{SO}}(t) \mathrm{d}t - Q_{i,k} \right\} \ge 0 \qquad \forall k \in \mathcal{K}, \quad \forall i \in \mathcal{N}, \tag{3}$$

where $p_i(t)$ and $\rho_{i,k}$ are the Lagrange multipliers for bottleneck capacity and demand conservation constraints, respectively. They can be interpreted as the optimal dynamic congestion price on every link for achieving the DSO flow pattern and the commuting cost of $(i,k)$-commuters under the optimal pricing scheme.

Meanwhile, the DUE problem is formulated as a linear complementarity problem (LCP) to find the flow and queueing delay patterns where each commuter chooses their departure time so as to minimize their experienced trip cost. The LCP consists of the following departure

time choice condition and the queueing condition in addition to the demand conservation condition, which has a similar form to Eq. (3):

$$0 \le q_{i,k}^{\mathrm{UE}}(t) \perp \left\{ \sum_{j;j\le i} w_j^{\mathrm{UE}}(t) + \beta^k c(t) - \rho_{i,k}^{\mathrm{UE}} \right\} \ge 0 \qquad \forall k \in \mathcal{K}, \quad \forall i \in \mathcal{N}, \quad \forall t \in \mathcal{T}, \tag{4}$$

$$0 \le w_i^{\mathrm{UE}}(t) \perp \left\{ \mu_i \dot{o}_i^{\mathrm{UE}}(t) - \sum_{j;j\ge i} q_j^{\mathrm{UE}}(t) \right\} \ge 0 \qquad \forall i \in \mathcal{N}, \quad \forall t \in \mathcal{T}, \tag{5}$$

where $w_i(t)$ is the queueing delay on link $i$ which commuters with destination-arrival time $t$ experienced; moreover, $o_i^{\mathrm{UE}}(t) = t - \sum_{j;j\le i} w_j^{\mathrm{UE}}(t)$. An overdot denotes the derivative of the variable with respect to the destination arrival time $t$.

## 3 Queue replacement principle

We first introduce the formal definition of the queue replacement principle in the corridor network, as follows:

**Definition 3.1** (Queue replacement principle)**.** The queuing delay pattern in the DUE state is equal to the optimal pricing pattern in the DSO state, that is,

$$w_i^{\mathrm{UE}}(t) = p_i^{\mathrm{SO}}(t), \quad \forall i \in \mathcal{N}, \quad \forall t \in \mathcal{T}. \tag{6}$$

The QRP shows that, for a given optimal pricing pattern in the DSO state, a DUE solution exists even if the values of the pricing pattern replace those of the queueing delay pattern. This implies that we can clarify the condition where the QRP holds by investigating the existence condition of such a DUE solution. Based on this concept, we first derive the optimal pricing pattern analytically. We then substitute the analytical solution into the DUE problem and analyze the feasibility of the DUE to clarify the condition where the QRP holds.

To solve the DSO problem analytically, one of the most troublesome issues is the complex spatial interaction between flows on multiple bottlenecks; this makes it difficult to derive the solution (i.e. flow and cost pattern) on the multiple bottlenecks simultaneously. To avoid this difficulty, we focus on the *bottleneck-decomposition property*. Specifically, under some technical assumptions, the following relationship between the aggregated arrival flows is obtained:

$$q_i^{\mathrm{SO}}(t) > 0 \quad \Rightarrow \sum_{j;j\ge i} q_j^{\mathrm{SO}}(t) = \mu_i, \quad \forall i \in \mathcal{N}. \tag{7}$$

This relationship shows that when $q_i^{\mathrm{SO}}(t)$ is positive, i.e. commuters departing from the origin $i$ exist, the flow from the immediate upstream origin $i+1$ is equal to the bottleneck capacity $\mu_{i+1}$. This implies that we can solve the DSO problem sequentially from upstream to downstream origins and bottlenecks while taking the upstream flow as a given condition, i.e. the DSO problem can be decomposed into single-bottleneck problems.

Interestingly, the decomposed single bottleneck problems have the same structure as the optimal transport problem that can be analytically solvable by the theory of optimal transport (Akamatsu et al. 2021, Rachev and Rüschendorf 1998). Therefore, by applying the

theory to the decomposed bottleneck problems sequentially, we can derive the DSO solution and the optimal pricing pattern analytically.

We then derive the condition where the QRP holds by combining the optimal pricing pattern and the DUE problem. First, we substitute the value of $p_i^{SO}(t)$ into $w_i^{UE}(t)$ based on Eq. (6), and derive the DUE flow and cost pattern from the equilibrium conditions shown in the previous section. We next check whether the derived flow pattern is feasible or not by examining its non-negativity condition. Consequently, we derive the following theorem:

**Theorem 3.1** (Sufficient QRP condition)**.** The QRP holds if the schedule delay cost function satisfies the following conditions:

$$\frac{\mu_{(i+1)}}{\mu_i} - 1 < \dot{c}(t) < \frac{\mu_i}{\mu_{(i+1)}} - 1 \qquad \forall i \in \mathcal{N} \setminus \{N\}, \quad \forall t \in \mathcal{T}. \tag{8}$$

It should be noted that the analytical solution to the DUE problem is derived through the derivation of this sufficient QRP condition. Thus, we here establish not only the QRP condition itself but also the constructive way for deriving the DUE solution from the DSO solution.

## 4 Discussion

The QPR enables the easy analysis of the efficiency and equity of the optimal pricing scheme. Specifically, we see that the pricing that mimics the queueing delay pattern in the DUE state archive the first best state and the Pareto improvement simultaneously when the QRP holds.

**Theorem 4.1** (Pareto improvement)**.** Suppose that the QRP condition (8) holds. If the road manager imposes the dynamic pricing equal to the queuing delay $\{w_i^{\mathrm{UE}}(t)\}_{t\in\mathcal{T}}$ at all bottlenecks $i \in \mathcal{N}$, the total system cost can be minimized without increasing anyone's equilibrium commuting cost.

In addition, the QRP facilitates the examination of instances in which pricing is instituted solely for select bottlenecks and the evaluation of equilibrium under partial bottleneck pricing. We consider the case in which the optimal pricing is introduced to some bottlenecks $\mathcal{N}^{\mathrm{P}} \subseteq \mathcal{N}$ (hereafter partial bottleneck pricing: PBP). Interestingly, if the pricing bottleneck set $\mathcal{N}^{\mathrm{P}}$ satisfies a certain condition, the queueing delay pattern at no-pricing bottleneck under the PBP equals that in the DUE state, i.e., $w_i^{\mathrm{PBP}}(t) = w_i^{\mathrm{UE}}(t)$, $\forall i \in \mathcal{N} \setminus \mathcal{N}^{\mathrm{P}}$. Moreover, using this queueing delay pattern, we can construct the equilibrium flow pattern under the PBP in a similar to derive the DUE solution. By comparing this solution with the DUE and DSO solutions, we can find the Pareto improvement property of PBP.

These results suggest that the QRP does not only contribute to obtaining an analytical solution, but also enables us to clearly understand the efficiency and equity of an optimal pricing scheme from the perspective of welfare analysis. Given the usefulness of the QRP, it is another significant challenge to analyze whether the QRP holds in more general settings, for example, considering the other heterogeneities of commuters and a network with route choice.

## Acknowledgements

This work was supported by JSPS KAKENHI Grant Numbers JP20J21744, JP21H01448, and JP20H02267.

## References

Akamatsu, T., Wada, K., Hayashi, S., 2015. The corridor problem with discrete multiple bottlenecks. Transportation Research Part B: Methodological 81, 808–829.

Akamatsu, T., Wada, K., Iryo, T., Hayashi, S., 2021. A new look at departure time choice equilibrium models with heterogeneous users. Transportation Research Part B: Methodological 148, 152–182.

Arnott, R., 1998. Congestion tolling and urban spatial structure. Journal of regional science 38, 495–504.

Arnott, R., DePalma, E., 2011. The corridor problem: Preliminary results on the no-toll equilibrium. Transportation Research Part B: Methodological 45, 743–768.

Fu, H., Akamatsu, T., Satsukawa, K., Wada, K., 2022. Dynamic traffic assignment in a corridor network: Optimum versus equilibrium. Transportation Research Part B: Methodological 161, 218–246.

Iryo, T., Yoshii, T., 2007. Equivalent optimization problem for finding equilibrium in the bottleneck model with departure time choices, in: 4th IMA International Conference on Mathematics in TransportInstitute of Mathematics and its Applications, trid.trb.org.

Li, Z.C., Huang, H.J., Yang, H., 2020. Fifty years of the bottleneck model: A bibliometric review and future research directions. Transportation Research Part B: Methodological 139, 311–342.

Osawa, M., Fu, H., Akamatsu, T., 2018. First-best dynamic assignment of commuters with endogenous heterogeneities in a corridor network. Transportation Research Part B: Methodological 117, 811–831.

Rachev, S.T., Rüschendorf, L., 1998. Mass Transportation Problems: Volume I: Theory. Springer Science & Business Media.

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

# Lane Choice Equilibrium in Macroscopic Multi-Queue Node Model for Multilane Intersections

Xiaolin Gong*, Michiel C.J. Bliemer, Mark P.H. Raadsen
Institute of Transport and Logistics Studies, University of Sydney, Australia
xiaolin.gong@sydney.edu.au, michiel.bliemer@sydney.edu.au, mark.raadsen@sydney.edu.au
* Corresponding author

*Extended abstract submitted for presentation at the 9th International Symposium on Dynamic Traffic Assignment (DTA), July 10-12, 2023, United States*

May 19, 2023

Keywords: macroscopic traffic assignment model, link FIFO, node model, lane choice equilibrium

## 1. INTRODUCTION

In macroscopic Dynamic Traffic Assignment (DTA), node models play an essential role as part of network loading. In network loading, (i) travel demand for each path is assumed given (via a route choice model), (ii) the link model propagates flows downstream and queues upstream, (iii) the node model determines *realised* link outflows depending on *desired* link outflows (referred to as sending flows) while taking available link inflow capacities into account (referred to as receiving flows). The main outputs of network loading are link inflows and outflows as well as travel time delays. In this work we propose a novel node model that accounts for multi-lane intersections by explicitly considering lane configurations and relating it to an equilibrium-based formulation.

Consider a network consisting of nodes and links, where each node $n$ has a set of incoming links $I$ and outgoing links $J$. For each node, define link sending flows by vector $\mathbf{s}^\bullet = [s_i^\bullet]_{i\in I}$, link receiving flows by vector $\mathbf{r}^\bullet = [r_j^\bullet]_{j\in J}$, and movement-specific turn fractions by matrix $\mathbf{f}^\bullet = [f_{ij}^\bullet]_{i\in I, j\in J}$, where $(i,j)$ is a movement from incoming link $i \in I$ to outgoing link $j \in J$. Link exit capacities are defined by vector $\mathbf{c}^\bullet = [c_i^\bullet]_{i\in I}$. The node model output is given by movement-specific outflow rates $\mathbf{v}^\bullet = [v_{ij}^\bullet]_{i\in I, j\in J}$ via an implicit function $\Gamma(\cdot)$ that describes the lane-aggregated node model,

$$\mathbf{v}^\bullet = \Gamma\left(\mathbf{s}^\bullet, \mathbf{f}^\bullet, \mathbf{r}^\bullet, \mathbf{c}^\bullet\right). \tag{1}$$

Daganzo (1994, 1995) was one of the first to study node models. Since then, several researchers have made extensions (Jin & Zhang, 2003; Tampère et al., 2011; Flötteröd & Rohde, 2011; Wright et al., 2017), where Tampère et al. (2011) proposed a generic first-order node model that applies to the intersection with any combination of links. Typically, existing node models assume that all vehicles wait in the same *first-in-first-out* (FIFO) queue, no matter their turn/movement. If one movement becomes congested, all other movements from the same link are equally affected. This overlooks the fact that for multi-lane intersections different queues and delays per movement can arise. Hence, while a *single-queue* (SQ) approach is suitable for single lane intersection it cannot adequately model general multilane configurations. Because the majority of intersections, with significant delays, are not single lane in nature, it is imperative to address these limitations in order to develop more practical solutions.

There is currently no satisfactory and practically viable methodology to address this link-based FIFO restriction , despite the fact that node models account for the majority of queueing delays produced in DTA models. Wright et al. (2017) partially relax the strong link FIFO assumption by capturing interactions between lanes via a single exogenous parameter, but do not explicitly consider approach lane configurations nor multiple queues. Fig. 1(a) shows a typical example of a multi-lane intersection. Fig. 1(b) shows the SQ node model representation with link-based queues. Vehicles on link 1 that are unable to turn left prevent straight moving vehicles (in the model, not in reality) from exiting the link, leading to unrealistic traffic flows and delays.

This work builds upon the SQ node model of Tampère et al. (2011) while proposing the first *multi-queue* (MQ) node model by accounting for the lane configuration to achieve the link-based FIFO relaxation. We first construct a lane-expanded node that can account for separate queues per lane as illustrated in Fig. 1(c). Then, we describe the MQ node model where we perform a lane-aggregation to obtain separate queues for each movement as illustrated in Fig. 1(d).

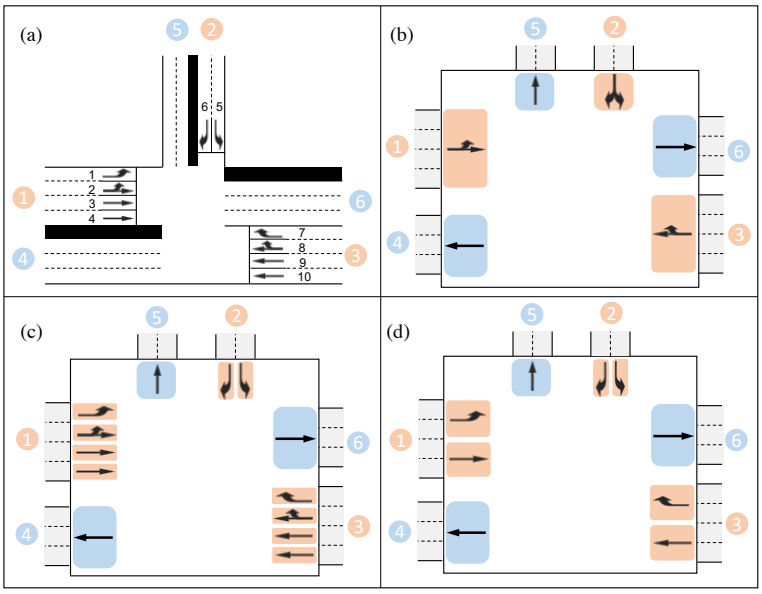



**Fig. 1. (a) Three-legged multilane intersection, (b) link-based queues, (c) lane-based queues, (d) movement-based queues.**

In the proposed MQ node model, each approach lane can in theory have a separate queue. We propose the concept of a *lane choice equilibrium*, in which drivers cannot exit the link earlier by unilaterally switching lanes. In equilibrium, all used lanes for a movement have the same delay, which is smaller than the delay of any unused lane. A lane choice equilibrium at the node level is a necessary condition for a route choice equilibrium at the network level (which is generally the objective in traffic assignment for strategic planning purposes).

## 2. METHODOLOGY

The node model is time invariant (similar to Tampère et al. (2011)), we therefore, for notational convenience, can omit time $t$ from all variables. Consequently, the model is compatible with event-based and time-discretised DTA network loading models, as well as semi-dynamic and/or static network loading methods.

**Proposed MQ node model: lane-expansion**

Each incoming link $i \in I$ has a set of approach lanes denoted by $M_i$. Let $M = \bigcup_i M_i$ be the set of all incoming approach lanes of node *n*. Let $M_{ij} \subseteq M_i$ denote the (sub)set of lanes on incoming link $i \in I$ that allow the movement towards $j \in J$.

The MQ node model assumes the same four inputs as the SQ node model, but now considers lane-specific turn fractions $\mathbf{f} = [f_{ij}^m]_{i\in I, j\in J, m\in M_i}$ and lane-specific exit capacities $\mathbf{c} = [c_i^m]_{i\in I, m\in M_i}$ for which hold $\sum_{m\in M_i} f_{ij}^m = f_{ij}^\bullet$ and $\sum_{m\in M_i} c_i^m = c_i^\bullet$. Lane- and movement-specific outflow rates $\mathbf{v} = [\mathbf{v}_{ij}^m]_{i\in I, j\in J, m\in M_i}$ can be computed using implicit function $\Gamma_{\mathrm{MQ}}(\cdot)$ where lanes are now explicitly considered,

$$\mathbf{v} = \Gamma_{\mathrm{MQ}}\left(\mathbf{s}^\bullet, \mathbf{f}, \mathbf{r}^\bullet, \mathbf{c}\right). \tag{2}$$

Based on $\mathbf{s}^\bullet$ and $\mathbf{f}$ we can compute lane-specific sending flow $\tilde{\mathbf{s}} = [\tilde{s}_i^m]_{i\in I, m\in M_i}$,

$$\tilde{s}_i^m = \min\left\{\sum_{j\in J'} f_{ij}^m s_i^\bullet, c_i^m\right\}, \qquad \forall m \in M_i, \forall i \in I. \tag{3}$$

Further, based on $\mathbf{f}$ we can compute lane-specific turn fractions $\tilde{\mathbf{f}} = [\tilde{f}_{ij}^m]_{i\in I, j\in J, m\in M_i}$,

$$\tilde{f}_{ij}^m = \frac{f_{ij}^m}{\sum_{j\in J'} f_{ij}^m}, \qquad \forall i \in I, \forall j \in J', \forall m \in M_i. \tag{4}$$

Using $\tilde{\mathbf{s}}$ and $\tilde{\mathbf{f}}$, we can reformulate MQ node model $\Gamma_{\mathrm{MQ}}(\cdot)$ into a lane-expanded SQ node model $\Gamma_{\mathrm{SQ}}(\cdot)$ ($\Gamma_{\mathrm{SQ}}(\cdot) = \Gamma(\cdot)$), where each lane is now treated as a separate link with its own queue,

$$\Gamma_{MQ}\left(\mathbf{s}^{\bullet},\mathbf{f},\mathbf{r}^{\bullet},\mathbf{c}\right)=\Gamma_{SQ}\left(\tilde{\mathbf{s}},\tilde{\mathbf{f}},\mathbf{r}^{\bullet},\mathbf{c}\right). \tag{5}$$

In other words, the conventional SQ node model can be used by changing input variables $(\mathbf{s}^{\bullet},\mathbf{f})$ to $(\tilde{\mathbf{s}},\ \tilde{\mathbf{f}})$.

**Proposed MQ node model: lane aggregation**

To determine $\mathbf{f}$ endogenously in the node model, we propose a lane choice equilibrium approach, similar to Wardrop's equilibrium law for route choice. Equilibrium turn fractions, denoted by $\bar{\mathbf{f}}$, can be found by solving the following variational inequality problem (VIP):

$$\sum_i \sum_j \sum_{m\in M_{ij}} h_{ij}^m(\bar{\mathbf{f}})\left(f_{ij}^m-\bar{f}_{ij}^m\right)\geq 0, \qquad \forall \mathbf{f}\in F, \tag{6}$$

where $\mathbf{h}=[h_{ij}^m]_{i\in I,j\in J,m\in M_{ij}}$ is an appropriate generalized lane exit cost function that reflects queuing delays, and $F$ is the set of feasible turn fractions with non-negativity constraints $f_{ij}^m\geq 0$.

When a lane is capacity constrained, we propose the following generalized continuous lane exit queuing delay:

$$h_{ij}^m(\mathbf{f})=\frac{\sum_{j\in J'} f_{ij}^m s_i^{\bullet}}{\sum_{j\in J'} v_{ij}^m}-1, \forall i\in I, \forall j\in J', \forall m\in M_{ij} \tag{7}$$

The ratio indicates the proportion of the sending flow that can flow out. The larger this ratio, the more constrained the flow is, and the larger the queuing delay. In equilibrium, queuing delay will be the same across all used lanes for a given movement, which indicates that vehicles flow out in the same proportion as that they arrive at the node.

When a lane is in free flow, we propose the concept of *pressure proportionality* to obtain the unique solution. The smaller the ratio $\tilde{s}_i^m/c_i^m$, the larger the outflow potential and therefore the more attractive the lane will be.

Fig. 2 displays the transformation and relationships of all model variables, which clearly shows that the SQ node model is at the heart of the MQ node model.

## 3. RESULTS AND DISCUSSION

Consider an intersection as discussed in Tampère et al. (2011), where each incoming link is considered as a single lane (see Fig. 3(a)). Link capacities are 1000, 2000, 1000, 2000, 1000, 2000, 1000, 2000 veh/h on link 1 to link 8, respectively. Let us assume the lane configurations as shown in Fig. 3(b), where link 2 and link 4 have two lanes for three movements, link 1 and link 3 have one lane for three movements. The capacity of each lane is 1000 veh/h. The example is now examined with the novel MQ node model. The equilibrium solution in the MQ node model is obtained after 4 iterations (see Table 1).

According to the equilibrium solution in Table 1, lane 6 on link 4 is uncongested while lane 5 on the same link is congested. For movement (4, 6) all vehicles choose lane 6 since it has less travel cost. As a result, shared lane 5 on link 4 is only used for the movement (4, 7). The equilibrium solution in the MQ node model shows link 2 is fully congested, which is the same as the result of Tampère et al. (2011). In equilibrium, vehicles going straight on link 2 are distributed to lanes 2 and 3 under equilibrium lane-specific split proportions $\bar{\mathbf{f}}$, enabling equal travel costs on both lanes.

Table 2 shows the accepted outflows from both the SQ and MQ node model, where the overall outflow in MQ model (4211 veh/h) is 4.21% higher than that in SQ node model (4041 veh/h). For movement (4, 6) the SQ node model only allows 645 veh/h to exit link 4, whereas the MQ node model allows 800 veh/h. These results show how the MQ node model, by explicitly considering the approach lane configuration and relaxing the link-FIFO restriction, yields significantly different results compared to the SQ node model.

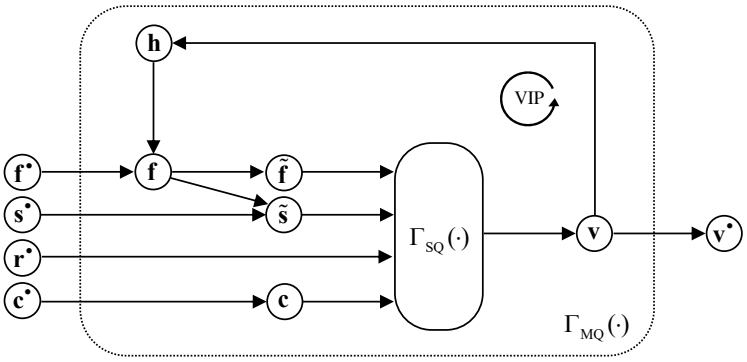
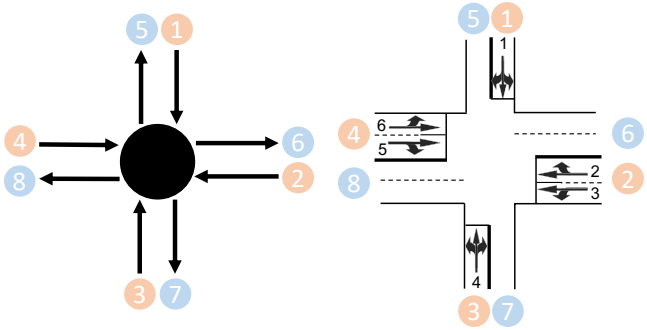



**Fig. 2. Schematic diagram of variable relationship between SQ and MQ node model.**




(a) (b)

**Fig. 3. (a) Intersection example from Tampère et al. (2011) (b) Intersection example with lane configuration**

The applied iterative solution algorithm will be provided in the presentation, and we will also present an extension to include traffic signal control. The novel MQ node model proposed in this work is directly compatible with existing macroscopic network loading models in OmniTRANS and AIMSUN. Hence, the presented approach presents governments and consultants with an effective method to accurately depict delays and traffic flows, especially at multi-lane motorway junctions and urban intersections.

In future research, we will explore extensions to include interaction with pedestrians, movement-specific outflow capacities and multiple vehicle types. Further, empirical data will be used to validate the outcomes of the proposed model.

**Table 1. Equilibrium results in MQ node model.**

| | | $\bar{\mathbf{f}}$ | | | | $\mathbf{v}(\bar{\mathbf{f}})$ | | | | $\mathbf{h}(\bar{\mathbf{f}})$ |
|---|---|---|---|---|---|---|---|---|---|---|
| *i* | *m* *j* | 5 | 6 | 7 | 8 | 5 | 6 | 7 | 8 | |
| 1 | 1 | 0 | 0.100 | 0.300 | 0.600 | 0 | 50 | 150 | 300 | 0 |
| 2 | 2 | 0.050 | 0 | 0 | 0.495 | 70 | 0 | 0 | 696 | 0.423 |
| 2 | 3 | 0 | 0 | 0.150 | 0.305 | 0 | 0 | 211 | 428 | 0.423 |
| 3 | 4 | 0.125 | 0.125 | 0 | 0.750 | 96 | 96 | 0 | 575 | 0.044 |
| 4 | 5 | 0 | 0 | 0.471 | 0 | 0 | 0 | 639 | 0 | 0.251 |
| 4 | 6 | 0.058 | 0.471 | 0 | 0 | 100 | 800 | 0 | 0 | 0 |

**Table 2. The accepted outflow in SQ and MQ node model.**

| *i* | *j* | 5 | 6 | 7 | 8 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| | | outflow in SQ node model | | | | outflow in MQ node model | | | |
| 1 | | 0 | 50 | 150 | 300 | 0 | 50 | 150 | 300 |
| 2 | | 69 | 0 | 206 | 1096 | 70 | 0 | 211 | 1124 |
| 3 | | 100 | 100 | 0 | 600 | 96 | 96 | 0 | 575 |
| 4 | | 81 | 645 | 645 | 0 | 100 | 800 | 639 | 0 |

## 4. REFERENCES


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.

Daganzo, C. F. (1995). The cell transmission model, part II: Network traffic. *Transportation Research Part B: Methodological*, *29*(2), 79–93.

Flötteröd, G., & Rohde, J. (2011). Operational macroscopic modeling of complex urban road intersections. *Transportation Research Part B: Methodological*, *45*(6), 903–922.

Jin, W. L., & Zhang, H. M. (2003). On the distribution schemes for determining flows through a merge. *Transportation Research Part B: Methodological*, *37*(6), 521–540.

Tampère, C. M. J., Corthout, R., Cattrysse, D., & Immers, L. H. (2011). A generic class of first order node models for dynamic macroscopic simulation of traffic flows. *Transportation Research Part B: Methodological*, *45*(1), 289–309.

Wright, M. A., Gomes, G., Horowitz, R., & Kurzhanskiy, A. A. (2017). On node models for high-dimensional road networks. *Transportation Research Part B: Methodological*, *105*, 212–234.

# The Challenge of Computing the Cell Transmission Model Loading with Exact First-In-First-Out

Genaro Jr. Peque[a], and Hillel Bar-Gera[b,*]

[a]*LocationMind Inc., Tokyo, Japan.*
[b]*Department of Industrial Engineering and Management, Ben Gurion University of the Negev, Beer-Shave 8410501, Israel*

Extended Abstract

*Keywords:* Dynamic Network Loading (DNL), Cell Transmission Model (CTM), First-In-First-Out (FIFO).

# 1 Introduction

Multi-commodity dynamic network loading (DNL) considers all traveler choices as given, including origins, destinations, departure times, modes and routes. Given these choices, and a description of the road network, the DNL produces predictions of traffic patterns, focusing on congestion and its progression in time and space. Many commonly used DNL models, including the popular cell transmission model – CTM (Daganzo, 1994; 1995; 2005), consider traffic as continuous (see chapter 9 in Boyles et al., 2021). The advantages of continuous-traffic models include: 1. deterministic outcome, without the stochastic noise of simulations; 2. identification of cause and effect is less complicated; 3. insights that are useful for intervention design and evaluation. On the other hand, continuous-traffic models have several limitations since certain complexities are easier to represent in discrete simulation models. Among these limitations, addressing the condition of first-in-first-out (FIFO) in continuous-traffic models is not trivial (e.g., Blumberg and Bar-Gera, 2009).

Daganzo (1995) proposed to approximate FIFO in CTM (denoted here as A-FIFO-CTM), by merging flows (of the same commodity) when they enter the same cell in the same time step. Other flow merging strategies can be considered for FIFO approximation. Evaluation and comparison of alternative FIFO approximation options requires a benchmark, where FIFO is maintained exactly, at the network level (denoted here as E-FIFO-CTM).

* Corresponding author. E-mail address: *bargera@bgu.ac.il*.

Bar-Gera and Carey (2022) discussed the conceptual challenges in developing a method for E-FIFO-CTM, and particularly the issue of partial overlaps between traffic components (also denoted as cohorts or quanta). At the conceptual level, the difference between the two methods may seem to be a technical book-keeping issue. The research presented here shows that the computational challenge is very different. While implementing the standard CTM (A-FIFO-CTM) is a reasonable project for an undergraduate course, E-FIFO-CTM is far more complex.

To address the computational challenge, we developed a novel approach for depicting traffic movement through nodes, and particularly through merge nodes, using the concept of match matrices. Our novel comprehensive formulation addresses all corner cases (e.g., traffic components of zero magnitude). The method will be discussed in detail in the full paper, while the remainder of this extended abstract consists of a general rigorous formulation for E-FIFO-CTM (section 2), description of our main test network (section 3), and illustrative solution evaluation results (section 4).

# 2 Formulation

The foundation of this research is a novel rigorous formulation that covers both the traditional approximate FIFO version of CTM (A-FIFO-CTM), which is based on Daganzo (1995), as well as the recently proposed E-FIFO-CTM (Bar-Gera and Carey, 2022). Part of the exposition here is merely a concise reminder of key equations from the above-mentioned previous studies (eq. 1-4). Node conditions (eq. 5-6) capture the same principles proposed previously in a manner that simplifies the solution verification process considerably. Aggregate occupancies and flows are divided into components and sub-components. The connection between aggregate and disaggregate traffic representations are ensured by eq. (7-12), while disaggregate flow conservation is given by (13). Finally, the main novelty of the presented formulation is the representation of disaggregated node behavior using the concept of match matrices (eq. 14-15).

Consider a network consisting of a set of nodes, N, and a set of cells ,C , connecting these nodes. Let $C_{in}(n) \subseteq C$ and $C_{out}(n) \subseteq C$ be the sets of cells entering and leaving node $n \in N$, respectively. Input data include the functions $q_i^{in}$ and $q_i^{out}$, the receiving and sending capacities respectively of cell $i \in C$, as well as $\alpha_i$, its "guaranteed merging share." Time step $t = 1, \dots, T$ represents the interval $[(t-1)\cdot\delta, t\cdot\delta]$. As in CTM, the main aggregate variables are: $x_{it}$ – occupancy; $u_{it}$ – entering flow; and $v_{it}$ – exiting flow. Aggregate constraints are:

(1) $x_{i,t+1} = x_{it} + u_{it} - v_{it} \quad \forall i \in C, 1 \le t \le T,$

(2) $\sum_{i \in C_{in}(n)} v_{it} = \sum_{i \in C_{out}(n)} u_{it} \quad \forall n \in N, 1 \le t \le T.$

(3) $u_{it} \le q_i^{in}(x_{it}) \quad \forall i \in C, 1 \le t \le T,$

(4) $v_{it} \le q_i^{out}(x_{it}) \quad \forall i \in C, 1 \le t \le T,$

(5) $|B_t^s \cap C_{in}(n)| + |B_t^r \cap C_{out}(n)| \ge 1;\ B_t^r = \left\{i \in C \middle| u_{it} = q_i^{in}(x_{it})\right\};\ B_t^s = \left\{i \in C \middle| v_{it} = q_i^{out}(x_{it})\right\};\ \forall n \in N, 1 \le t \le T.$

(6) $v_{it} \ge \min\left(q_i^{out}(x_{it}), \alpha_i \cdot u_{jt}\right) \quad \forall i \in C_{in}(n)\ ,\mathrm{j} \in \mathrm{C_{out}}(n), n \in N, 1 \le t \le T.$

Every occupancy value, $x_{it}$, is divided into $G_{it}$ occupancy components (OC), denoted $x_{itg}$, $g = 1, \dots, G_{it}$. Similarly, every flow value, $u_{it}$ for $i \in C_{out}(n)$ or $v_{it}$ for $i \in C_{in}(n)$, is divided into $H_{nt}$ flow components (FC), denoted $u_{ith}$ or $v_{ith}$, $h = 1, \dots, H_{nt}$. Naturally, component aggregation constraints must be satisfied:

(7) $\sum_{g=1}^{G_{it}} x_{itg} = x_{it} \quad \forall i \in C, 1 \le t \le T,$

(8) $\sum_{h=1}^{H_{nt}} u_{ith} = u_{it} \quad \forall i \in C_{out}(n), n \in N, 1 \le t \le T,$

(9) $\sum_{h=1}^{H_{nt}} v_{ith} = v_{it} \quad \forall i \in C_{in}(n), n \in N, 1 \le t \le T.$

Occupancy and flow components are further divided by their route ($r \in R$) into subcomponents, i.e., occupancy subcomponent (OSC) and flow subcomponent (FSC) respectively. Subcomponent aggregation constraints are:

(10) $\sum_{r \in R} x_{itgr} = x_{itg} \quad \forall 1 \le g \le G_{it}, i \in C, 1 \le t \le T,$

(11) $\sum_{r \in R} u_{ithr} = u_{ith} \quad \forall 1 \le h \le H_{nt}, i \in C_{out}(n), n \in N, 1 \le t \le T,$

(12) $\sum_{r \in R} v_{ithr} = v_{ith} \quad \forall 1 \le h \le H_{nt}, i \in C_{in}(n), n \in N, 1 \le t \le T,$

The disaggregate node flow conservation constraint is:

(13) $u_{ithr} = v_{i'thr} \; \forall 1 \le h \le H_{nt}, i \in C_{in}(n); \; i' \in C_{out}(n); i, i' \subseteq r \in R, n \in N, 1 \le t \le T.$

The main difference between E-FIFO-CTM and A-FIFO-CTM is in the disaggregate cell flow conservation. Our innovative unifying formulation is based on alternative notations for flow and occupancy subcomponents. For each combination of cell $i \in C_{out}(n) \cap C_{in}(n')$, time step $t$ and route $r$, four vectors $\boldsymbol{Y}(itr), \boldsymbol{Z}(itr), \boldsymbol{U}(itr), \boldsymbol{V}(itr)$ by setting $Y_g(itr) = x_{itgr} \; \forall 1 \le g \le G_{it}; Z_g(itr) = x_{i,t+1,g'r} \; \forall 1 \le g' \le G_{i,t+1};$ $U_h(itr) = u_{ithr} \; \forall 1 \le h \le H_{nt}; V_h(itr) = v_{ith'r} \; \forall 1 \le h \le H_{n't}$. Using these notations, and properly defined match matrices, $\boldsymbol{M}(it)$, the general cell flow conservation constraint can be written as:

(14) $[\boldsymbol{V}(itr) \quad \boldsymbol{Z}(itr)] = [\boldsymbol{Y}(itr) \quad \boldsymbol{U}(itr)] \cdot \boldsymbol{M}(it) \; \forall i \in C, 1 \le t \le T, r \in R.$

The construction of match-matrices in A-FIFO-CTM ensures that $H_{nt} = 1 \; \forall \, n \in N, 1 \le t \le T$, while in E-FIFO-CTM they ensure that within each time-step each flow materializes at a constant rate, and therefore at any merge node $n$ (i.e. if $|C_{in}(n)| > 1$):

(15) $\frac{v_{ith}}{v_{it}} = \frac{v_{i'th}}{v'_i t} \qquad \forall i, i' \in C_{in}(n); 1 \le h \le H_{nt}; \; 1 \le t \le T.$

# 3 Numerical experiments

Our main test network, presented in figure 1, is composed of 54 nodes, 48 cells, and 18 links. The main finding in the evaluation of the solution is the growth in the number of traffic components (i.e. rows in the solution file) as a function of the number of time steps in the model, as shown in figure 2. The solution for 20 time steps requires ~1E4 traffic components, ~1E5 traffic components for 50 time steps, and nearly 1E6 traffic components for 100 time steps. This growth pattern is almost as quick as a constant-rate exponential growth, thus verifying the concerns regarding the viability of exact FIFO solutions for practical purposes, with large-scale networks, especially when the modelling duration requires a large number of time steps. Nevertheless, the ability to produce such solutions on small to medium problems can be used as a benchmark for evaluating the quality of computationally efficient approximations.

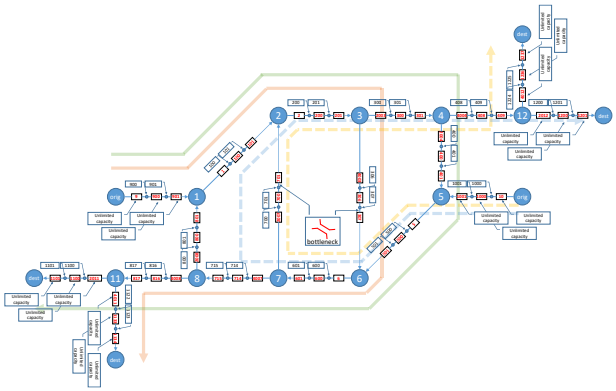



Figure 1: Test network.

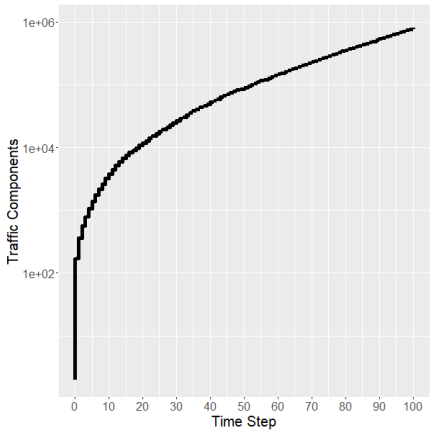



Figure 2. Number of traffic components as a function of model duration (log scale)

# References

Bar-Gera, H., Carey, M., 2022. Forthcoming in Transportation Research Part B. Constructing a cell transmission model solution adhering fully to first-in-first-out conditions.

Blumberg, M., Bar-Gera, H., 2009. Consistent node arrival order in dynamic network loading models. Transport. Res. Part B 43 (3), 285–300. https://doi.org/10.1016/ j.trb.2008.07.002.

Boyles, S.D., Lownes, N.E., Unnikrishnan, A., 2021. Transportation network analysis volume 1. Version 0.89. https://sboyles.github.io/blubook.html.

Daganzo, C.F., 1994. The cell transmission model: a dynamic representation of highway traffic consistent with the hydrodynamic theory. Transport. Res. Part B 28 (4), 269–287.

Daganzo, C.F., 1995. The cell transmission model part II: network traffic. Transport. Res. Part B 29 (2), 79–93.

Daganzo, C.F., 2005. A variational formulation of kinematic waves: basic theory and complex boundary conditions. Transport. Res. Part B 39, 187–196.

**Dynamic Traffic Assignment with Physics-Informed Deep Learning**

Shakiba Naderian (naderian@uw.edu), Ohay Angah (oangah@uw.edu), Yiran Zhang (yiranz94@uw.edu),
Xuegang (Jeff) Ban * (banx@uw.edu)
University of Washington
*: corresponding author

## 1. Introduction

Dynamic traffic assignment (DTA) models normally contain two major components: a behavioral model to describe users' travel choices (route, departure-time, mode, etc.) and a dynamical system model to describe the (physical) traffic dynamics. It is the integration of these two mathematically very distinct modeling components that make DTA one of the most challenging problems in transportation science. So far, dynamical systems (in the forms of ordinary differential equations (ODEs) or partial differential equations (PDEs) or their discrete-time forms) have been applied to describe traffic dynamics, which respect traffic physics while imposing difficulties in model analysis and computation.

Recent advances in DTA are featured by applying learning methods to learn/approximate traffic dynamics, such as Kriging models (Patwary et al., 2021; Song et al., 2018), neural networks (NN) (Chen et al., 2015; Fan et al., 2022), Autoregressive integrated moving average and neural network (ARIMA-NN) hybrid models (Ye et al., 2010), and hybrid support vector machine models (Wang & Shi, 2013). Although learning-based models have shown capabilities of approximating complex patterns from data, they need a large amount of data for training compared to physical models; moreover, they are sensitive to data noises and lacking physical interpretations. In this study, we develop a physics-informed deep learning (PIDL) framework to approximate traffic dynamics and apply the PIDL method to DTA. The PIDL framework leverages the advantage of both the existing physical models and deep neural networks that are powerful for estimation. This can help produce more accurate predictions, while at the same providing more interpretations to the model structure and the results. In this abstract, we present the key methods and structure of the PIDL-based DTA modeling approach. More details of the methods and numerical experiments will be provided in the full paper.

## 2. Methodology

To develop the PIDL model, we apply the DTA model developed by Ma et al. (2018), in which the link-level dynamics of is extended to Network-level Dynamics to construct a (link-based) continuous-time Dynamic User Equilibrium (DUE) model. Using the formulation of double queue and nodal models for the links and nodes respectively, the proposed DTA model takes the interactions between incoming and outgoing links at a junction into account, based on which the traffic dynamics and travelers' departure-time and route choices are modeled. Figure 1 and Figure 2 below illustrate the double queue model and the nodal model, respectively. Readers can refer to Ma et al. (2018) for more details of the DUE model. Below we provide a brief description of the traffic dynamics modeling, which will help us develop the PIDL-based framework.

Consider a network with a set of links $(i,j) \in L$ between a set of nodes $N$, each link comes with characteristics including inflow capacity $(\underline{C_{ij}^{p}})$, exit flow capacity $(\underline{C_{ij}})$ that may be identical to the inflow capacity for a homogeneous link, queue capacity $(\underline{Q_{ij}})$, free flow travel time $(\tau_{ij}^{0})$, and shockwave travel time $(\tau_{ij}^{\omega})$. The dynamics of each link can be defined by its inflow $(p_{ij}(t))$, exit flow $(v_{ij}(t))$, and upstream

and downstream queues $(q_{ij}^{u}(t), q_{ij}^{d}(t))$ that regulate the inflow and exit flow of the link. The upstream and downstream queues are defined as follows:

$$q_{ij}^{u}(t) = p_{ij}(t) - v_{ij}\left(t - \tau_{ij}^{\omega}\right),$$

$$q_{ij}^{d}(t) = p_{ij}\left(t - \tau_{ij}^{0}\right) - v_{ij}(t).$$

The upstream and downstream queues can only capture traffic dynamics within a link. To consider network-level dynamics, a nodal model is often needed to distribute the incoming link flows (demand) to the outgoing links (supply) and calculate the amount of flow withholding at the entrance and exit of that link, in case of congestion and especially queue spillback, occurs. To do so, slack and auxiliary variables (also referred to as hidden variables in this study) are introduced to capture the withheld flow caused from flow spillback and insufficient capacity. In addition, the incoming flows to a junction are discharged based on predefined priorities in a way that the exit flow of an incoming link with the lowest priority would be withheld first if spillback occurs at the junction (Ma et al., 2018). Finally, based on the node model proposed, the hidden variables of the DTA model can be represented as complementarity constraints below:

$$0 \le q_{ij}^{d}(t) \perp \mu_{ij}(t) \ge 0,$$

$$0 \le \eta_{ij}^{Q}(t) \perp \underline{Q}_{ij}(t) - q_{ij}^{u}(t) \ge 0,$$

$$0 \le \eta_{ij}^{C}(t) \perp \underline{C}_{ij}^{p}(t) - p_{ij}(t) \ge 0,$$

where $\mu_{ij}(t)$ is a slack variable showing the withheld exit flow from link (i,j) caused by the lack of traffic on the link, and $\eta_{ij}^{Q}(t)$ and $\eta_{ij}^{C}(t)$ are slack variables showing withheld inflow to link (i,j) due to spillback and insufficient inflow capacity. The total withheld inflow to link (i,j) is $\eta_{ij}(t) = \eta_{ij}^{Q}(t) + \eta_{ij}^{C}(t)$. The main objective of using the above slack variables are better explained in Figure 1. The first complementarity condition above indicates that whenever queue forms at the downstream of the link, $\mu_{ij}(t)$ will be zero, and the exit flow of the link will be equal to the exit capacity if no spillback occurs from downstream. The other two complementarity constraints are to model the withheld inflows, due to (upper stream) queue storage capacity and inflow capacity respectively. If the inflow to link (i,j) is higher than the inflow capacity, or the upstream queue reaches the queue capacity, the inflow to the link will be withheld.

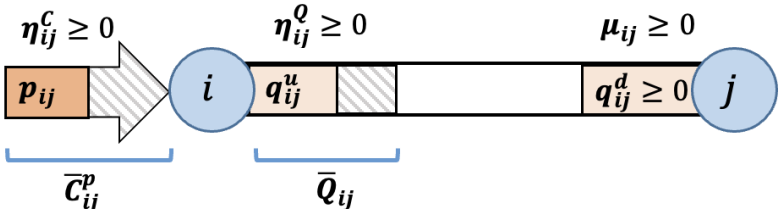



*Figure 1. Double queue model: slack variables determining the withheld inflow and exit flow on link (i,j)*

Finally, based on predefined discharging priorities, the exit flow of an incoming link $(n, i)$ with the $m^{th}$ priority is calculated by the nodal model. In this model, the amount of withheld inflow to the outgoing links $(\eta)$ are distributed as withheld exit flows of incoming links $(\delta)$ in a sequential order based on their priorities:

$$\underline{\delta}_{n_i^m,i}(t) = Min(\underline{C}_{n_i^m,i}(t) - \mu_{n_i^m,i}(t), \sum_{j\in(i,j)} \eta_{ij}(t) - \sum_{1\le\hat{m}\le m-1} \delta_{n_i^m,i}(t)), \qquad m = 1,\dots,M_i$$

The above equation indicates that if the exit flow of the link is greater than the withheld flow due to downstream congestion, the withheld exit flow will be equal to the latter. On the other hand, if the exit flow is already less than the flow that should be withheld, the withheld flow can only be equal to the exit flow $(\underline{C}_{n_i^m,i}(t) - \mu_{n_i^m,i}(t))$. The other auxiliary variable $(\delta_{n_i^m,i})$ is used to prevent nonnegativity of the withheld flow, and the exit flow is calculated through the following complementarity conditions:

$$0 \le v_{n_i^m,i}(t) \perp \sum_{j\in(i,j)} \eta_{ij}(t) - \sum_{1\le\hat{m}\le m-1} \delta_{n_i^m,i}(t) - \underline{\delta}_{n_i^m,i}(t) \ge 0, \qquad m = 1,\dots,M_i$$

$$0 \le \delta_{n_i^m,i}(t) \perp \delta_{n_i^m,i}(t) - \underline{\delta}_{n_i^m,i}(t) \ge 0, \qquad m = 1,\dots,M_i$$

where $M_i$ is the total number of incoming links. Figure 2 illustrates how due to downstream congestion, the exit flows from incoming links are withheld sequentially based on their discharging priorities that are previously defined.

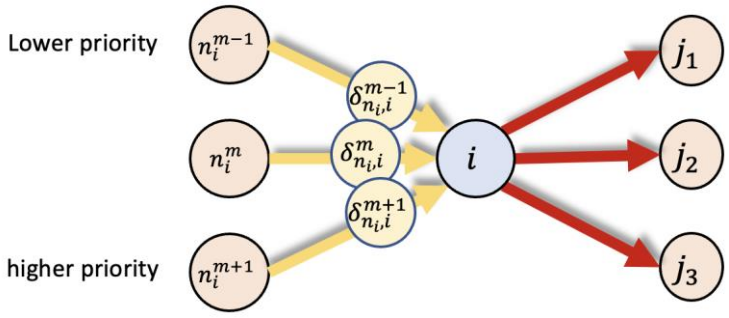



*Figure 2. withheld flows ($\delta_{n_i^m,i}$) of incoming links at junction i due to downstream spillback*

Lastly, the link travel time can be approximated using the following closed-form function:

$$\tau_{ij} = \tau_{ij}^0 + \frac{q_{ij}^d(t+\tau_{ij}^0)}{\underline{C}_{ij}}\left[1 + \frac{\mu_{ij}(t+\tau_{ij}^0) + \underline{\delta}(t+\tau_{ij}^0)}{\underline{C}_{ij}}\right]$$

Note that the link travel time is one of the key measures used in the departure-time and route choice models. As can be seen, the calculation of auxiliary and slack variables used in the travel time function is not a trivial task. In order to alleviate the computational burden of the time-consuming traffic dynamics model, we attempt to take advantage of the information that real-world data can provide and develop a physics-informed neural network capable of estimating the hidden state variables based on the input data. Such PIDL-based models have been applied successfully in the literature for other traffic dynamics problems such as car-following models (Mo et al., 2021). The main framework of the proposed PIDL model developed in this study is presented in Figure 3.

In this framework, two neural networks are developed, one for the calibration of the dynamic traffic assignment model parameters ($\underline{Q}_{ij}, \underline{C}_{ij}^{p}, \underline{C}_{ij}, \tau_{ij}^{0}, \tau_{ij}^{\omega}$) based on observed data (denoted as #1), and another for the estimation of hidden variables ($\delta_{ij}(t), \underline{\delta}_{ij}(t), \eta_{ij}(t), \eta_{ij}^{Q}(t), \eta_{ij}^{C}(t), \mu_{ij}(t)$) used in the travel time function (denoted as #2). A DTA model (e.g., the one presented in Ma et al., (2018)) sits in between the two deep neural networks, serving as the physical model to connect the two deep neural networks (thus "physical-informed" deep learning models). The DTA model can i) provide data for the hidden variables so that Neural Network #2 can be properly trained; and ii) generating output from the physical model so that both the physical model and Neural Network #1 can be co-trained. At the same time, key parameters of the DTA model (e.g., link inflow/exit flow capacities, upstream queue storage capacities, etc.) can also be calibrated during the training process. Neural Network #2 proposed here is unique, which helps train and generate the hidden variables. The input data into the model is a combination of observed and collocation variables including $p_{ij}(t)$, $v_{ij}(t)$, $q_{ij}^{d}(t)$, and $q_{ij}^{u}$. The output data are also the same variables at the next time step of the DTA model ($t + \Delta t$). In the numerical experiments, the observed information is derived from running simulations on the Downtown Seattle network, whereas the collocation data is generated by the link-based TA model in Ma et al. (2018).

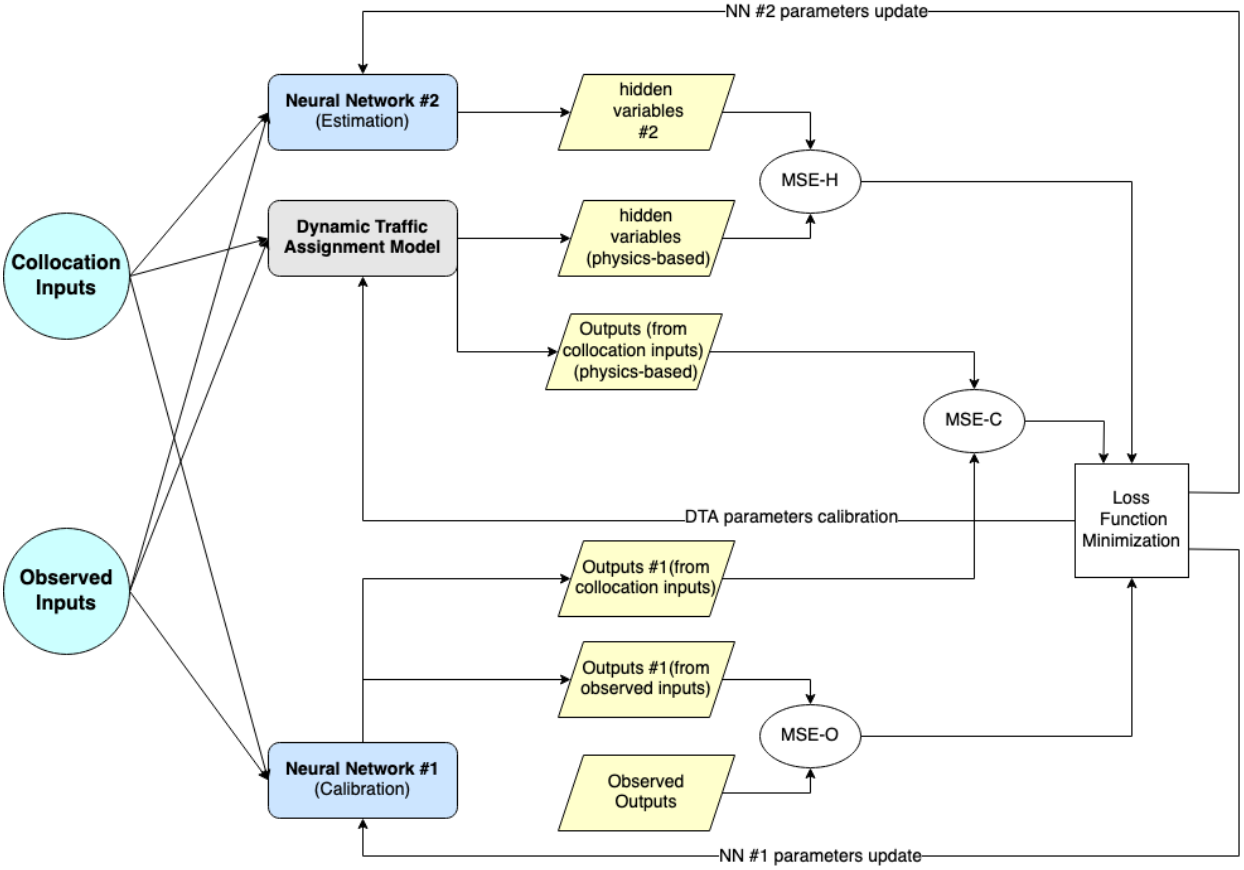



*Figure 3. Main framework of PIDL for DTA*

The proposed framework is trained using the Mean Square Error (MSE) criteria. At each training step, $MSE_O$ serves as a tool to train the neural network #1 based on the observed data and the neural network parameters ($\theta_1$) are estimated by minimizing this function. Similarly, $MSE_C$ is used for calibrating DTA

parameters $(h)$ using the neural network #1, and $MSE_H$ is for calibration of the neural network #2 parameters $(\theta_2)$:

$$MSE_O = \frac{1}{n_O}\left[\sum_{i=1}^{n_O} (output\ estimated\ from\ observed\ data\ by\ NN\#1\ -\ observed\ outputs)^2\right]$$

$$MSE_C = \frac{1}{n_C}\left[\sum_{i=1}^{n_C} (output\ estimated\ from\ collocation\ data\ by\ NN\#1 - output\ estimated\ from\ collocation\ data\ by\ DTA)^2\right]$$

$$MSE_H = \frac{1}{n_H}\left[\sum_{i=1}^{n_H} (hidden\ variables\ estimated\ by\ NN\#2\ -\ hidden\ variables\ estimated\ by\ DTA)^2\right]$$

The main purpose of the proposed framework is to minimize the loss function defined by summing up the mean square errors represented earlier:

$$Loss_{\theta_1,\theta_2,h} = \alpha MSE_O + \beta MSE_C + \gamma MSE_H,$$

$$\alpha + \beta + \gamma = 1.$$

Coefficients $\alpha$, $\beta$ and $\gamma$ are used to distribute the emphasis between the three objectives.

**3. Numerical Results**

We will first test the proposed PIDL-based DTA framework using the Sioux-Falls network, which has been widely tested in the literature, including the results and analysis of applying the DTA model in Ma et al. (2018). We will then implement the framework using the Downtown Seattle network bounded by Mercer Street at North, South Atlantic Street and Edgar Martinez Drive at South, Alaskan Way at West, and 12th Ave at the East side (Figure 4). The network was built and calibrated in Simulation of Urban Mobility (SUMO) using the Puget Sound Regional Council (PSRC) origin-destination (OD) demands and other observed data (such as loop data for volumes, travel time data, and transit data); see Ban et al. (2022).

Passenger vehicles are the focus of this study. We extract trajectories of passenger vehicles from the simulation outputs. The data is considered as the observed inputs shown in Figure 3. Meanwhile, we run the DTA model in Ma et al. (2018) given the same set of OD demands and properties of the network. The resulting assignments and travel times from the DTA model serve as the collocation inputs shown in Figure 3. These two datasets are used in the proposed framework to iteratively update the parameters of the neural networks and the DTA model until the convergence criterion is met. In the full paper, details of calibrating the PIDL models (the two neutral networks and the parameters of the DTA model) and the results and analysis of the PIDL-based DTA models will be presented. We will also compare the model results with pure model-based DTA models in Ma et al. (2018) to draw more insights.

*Figure 4. Downtown Seattle Network*

**4. Conclusions and Future Research**

In this study, we proposed a PIDL-based DTA framework that leverages the advantage of physics-based models and deep neural networks. In the proposed framework, we use a deep neural network to help calibrate the DTA model, while at the same the DTA model helps train another deep neural network for hidden variable estimation. The overall PIDL framework is expected to achieve higher calibration and variable estimation performance than using only data-driven neural networks or existing model-based DTA models. We will validate the proposed framework using the test cases of the Sioux-Falls network and the Downtown Seattle network. In the full paper, the following will be conducted and completed:

- A formulation of the proposed PIDL-based DTA framework
- Implementation of the proposed framework
- An evaluation of the DTA calibration and the variable estimation (especially for hidden variables)
- A comparison between using the proposed PIDL-based DTA framework and the existing linked-based DTA model.

**References**


Ban, X., Angah, O., Zhang, Y., Guo, Q., 2022. A Multiscale Simulation Platform for Connected and Automated Transportation System. Final Report Submitted to C2smart Tier 1 UTC, New York University.

Chen, X. (Michael), Zhu, Z., He, X., & Zhang, L. (2015). Surrogate-Based Optimization for Solving a Mixed Integer Network Design Problem. *Transportation Research Record, 2497(1)*, 124–136. https://doi.org/10.3141/2497-13

Fan, W., Tang, Z., Ye, P., Xiao, F., & Zhang, J. (2022). Deep Learning-Based Dynamic Traffic Assignment With Incomplete Origin–Destination Data. *Transportation Research Record*, 03611981221123805. https://doi.org/10.1177/03611981221123805

Ma, R., Ban, X., & Pang, J. S. (2018). A link-based differential complementarity system formulation for continuous-time dynamic user equilibria with queue spillbacks. *Transportation Science*, *52*(3), 564–592. https://doi.org/10.1287/trsc.2017.0752

Mo, Z., Shi, R., & Di, X. (2021). A physics-informed deep learning paradigm for car-following models. *Transportation Research Part C: Emerging Technologies*, *130*. https://doi.org/10.1016/j.trc.2021.103240

Patwary, A. U. Z., Huang, W., & Lo, H. K. (2021). A link-to-link segment based metamodel for dynamic network loading. *Transportation Research Part C: Emerging Technologies, 130,* 103286. https://doi.org/10.1016/j.trc.2021.103286

Song, W., Han, K., Wang, Y., Friesz, T. L., & del Castillo, E. (2018). Statistical metamodeling of dynamic network loading. *Transportation Research Part B: Methodological*, *117*, 740–756. https://doi.org/10.1016/j.trb.2017.08.018

Wang, J., & Shi, Q. (2013). Short-term traffic speed forecasting hybrid model based on Chaos–Wavelet Analysis-Support Vector Machine theory. *Transportation Research Part C: Emerging Technologies*, *27*, 219–232. https://doi.org/10.1016/j.trc.2012.08.004

Ye, Q., Wong, S. C., & Szeto, W. Y. (2010). Short-term traffic speed forecasting based on data recorded at irregular intervals. *13th International IEEE Conference on Intelligent Transportation Systems*, 1541–1546. https://doi.org/10.1109/ITSC.2010.5625184

# Session 6

Efficient Pedestrian and Bicycle Traffic Flow Estimation using Mobile Data Sources, Considering All Possible Paths
*Simanta Barman, Michael Levin and Raphael Stern*

A quasi-dynamic multi-region MFD stochastic user equilibrium model
*Lawrence Duncan, Thomas Rasmussen and David Watling*

Integrated Activity-Based Models With Dynamic Multimodal Simulation-Assignment With Macroscopic Congestion Functions
*Yuhan Zhou and Hani Mahmassani*

# Efficient Pedestrian and Bicycle Traffic Flow Estimation using Mobile Data Sources, Considering All Possible Paths

Simanta Barman, Michael W. Levin, Raphael Stern

Accurate estimate of traffic flow measures like annual average daily traffic (AADT) is vital to making decisions about roadway planning, safety, maintenance, operation etc. Most of the previous work on traffic flow estimation has been done for vehicular traffic. Most automated counters also mostly collected data for vehicular traffic. Methodology to inexpensively obtain an accurate estimate of traffic flow that is consistent with the number of productions and attractions is lacking in the literature. This is especially true for pedestrian and bicyclist traffic because of the high expenses of conducting household surveys and setting up traffic monitoring stations to collect the required data. In this study, we develop a methodology to inexpensively obtain a good estimate of pedestrian and bicyclist traffic flow. Instead of creating expensive travel behavior models, we focus on estimating the traffic flow from mobile data sources. This also allows us to avoid most of the privacy issues associated with travel behavior models based on household survey results. From the mobile data source we have access to the number of trips produced from and attracted to different regions, and link flows in a some of the links of the network. The number of trips produced and attracted are available for a larger region and are pre-adjusted based on characteristics of the regions in the available dataset. This gives us more confidence on the accuracy of that data. However, similar adjustment on the individual link flow data was not done because individual link information are not available to that same extent as region information. Therefore, accuracy of partially available link flows from the mobile data is unknown and may vary in different locations. To deal with erroneous data sources different techniques are used to estimate and keep improving an origin-destination (OD) matrix from the observed link flows to ultimately get a reasonable estimation of link flows. In essence, the two steps required for estimating the link flows include first estimating the OD matrix and then using a route choice model to solve trip assignment using the OD matrix to obtain the link flows. However, estimating the OD matrix based on link traffic counts is a highly underdetermined problem. This is because the number of links where traffic counts are available is much smaller than the number of OD pairs in a realistic network. In brief, the main objective of this research is to come up with an OD-matrix so that traffic flows predicted using that OD-matrix would match traffic flows from mobile-sourced data while ensuring production, attraction consistencies.

Most previous studies on this topic focused on vehicular traffic and formulated bi-level optimization problems with heuristics to estimate OD matrices. For networks with a large number of nodes, estimating the OD matrix becomes computationally expensive especially when solving an optimization problem that requires estimating the OD matrix and performing traffic assignment simultaneously to get the link flows. Along with estimation of traffic flows this research also tries to tackle the OD matrix estimation problem for pedestrians and bicyclists. Since pedestrian and bicycle traffic is usually very small it is assumed that travel time in different links do not depend on traffic on the links. This allows us to solve traffic assignment efficiently and get rid of the constraint from the original bi-level optimization problem where traffic assignment is solved which itself is an optimization problem. Since travel time is assumed to be constant, with logit loading, traffic as-

signment can be done without solving an optimization problem. By doing the traffic assignment portion of the OD matrix estimation problem using logit loading, we will first find an OD matrix and then use that OD matrix to determine the link flows. Then based on the partial link flow counts we will modify the OD matrix and hence link flows will also be updated. By doing this until no further improvement between partially available link flow counts and link flows from the OD matrix can be made, we will obtain the final OD matrix and final link flows.

The contributions of this study are as follows. First, we extract the trip productions, attractions and the traffic counts observed at different links from available mobile data sources. After extracting the number of trips produced and attracted for different regions of a network from the StreetLight's dataset, using the network topology we use trip distribution based on the gravity model to generate an initial OD-matrix. Then we formulate a non-linear convex optimization problem to improve the initial OD matrices for pedestrian and bicyclist trips so that the link flows resuling from the improved OD matrices matches the available partial link flows more accurately. In our model, we also enforce the consistency between the number of trips produced from a region and the number of trips attracted to a region with the OD matrix. We assume logit route choice model for traffic assignment both for pedestrians and bicyclists. Since other factors besides travel times also affect route choice, we consider all possible paths unlike previous studies where only $k$-shortest paths were considered to solve this problem. We show that logit assignment can be efficiently performed even with the consideration of all possible paths including cyclic paths. Methods from stochastic traffic assignment has shown that this can be performed efficiently while considering infinite paths without explicit path enumeration. We use a similar solution technique. We also present how to implement the solution algorithm using tensor operations using a graphics processing unit (GPU). Then we present the performance of our solution algorithm for a real network.

Consider a directed graph $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ where the sets of nodes and edges are represented by $\mathcal{N}$ and $\mathcal{E}$ respectively. Let $\mathcal{Z}$ be the set of zones or regions. Let the number of trips produced from zone $r \in \mathcal{Z}$ be $P_r$ and trips attracted to zone $s \in \mathcal{Z}$ be $A_s$. Denote the shortest path travel times from zone $r$ to zone $s$ with $T_{rs}$. We represent the partially available link flow by $\tilde{\mathbf{x}} = \left\{\tilde{x}_{ij}\ \forall (i,j) \in \tilde{\mathcal{E}}\right\}$ where $\tilde{\mathcal{E}} \subseteq \mathcal{E}$ is the set of links where we have link flow data. The graph $\mathcal{G}$ is obtained from OpenStreetMaps and other data like the production, attraction and partial link flows are available from different mobile sourced datasets.

The obvious approach is to assign trips to the shortest paths in terms of travel time. However, that approach has the following issues.

1. Multiple paths with equal or very similar travel times.

2. Relying solely on travel time to decide which path to take is not realistic because a traveller's decision to choose a path can be influenced by many different factors. Furthermore, perfect information about the travel times at different links are not available to the travellers anyway.

3. Number of paths in a network can be infinite if cycles are allowed, in other words if all possible paths are considered. Requiring total acyclic paths is a strong assumption and in reality pedestrians' and bicyclists' probability of using any path including the cyclic paths are non-zero. In addition, considering all possible paths including the cyclic paths allow us to solve the trip assignment more efficiently which will be discussed later.

Therefore, the most likely path flow formulation can be used instead where the probability of a path to have some flow is maximized. Initially, assuming only a single OD pair and $k$-shortest paths are equally likely to be used, the probability mass function (pmf) for a path flow $\mathbf{h} = [h_1, h_2, \cdots, h_k]$

can be represented by the multinomial distribution for demand $d$,

$$\Pr\{H = \mathbf{h}\} = \frac{d!}{\prod_{i=1}^{k} h_i!} \left(\frac{1}{k}\right)^k \tag{1}$$

Then maximizing that probability results in the most likely path flows formulation. Using Stirlings approximation further simplifying and adding all the OD pairs we get the objective function required for most likely path flow formulation. We also use all the possible paths $\Pi^{rs}$ for OD pair $(r, s)$ instead of $k$-shortest paths $\hat{\Pi}^{rs}$. Then to relax the assumption that all paths are equally likely, eq. (3) are added where disutility $c^\pi$ of path $\pi$ for each paths are considered.

$$\max_{\mathbf{h}} \quad \sum_{(r,s)\in\mathcal{Z}^2} \sum_{\pi\in\Pi^{rs}} h_\pi \log\left(\frac{h_\pi}{d^{rs}}\right) \tag{2}$$

$$\text{s.t.} \quad \sum_{\pi\in\Pi^{rs}} h_\pi c^\pi = c^*_{rs} d_{rs} \qquad \forall (r,s) \in \mathcal{Z}^2 \tag{3}$$

$$\sum_{\pi\in\Pi^{rs}} h_\pi = d_{rs} \qquad \forall (r,s) \in \mathcal{Z}^2 \tag{4}$$

$$h_\pi \geq 0 \qquad \forall (r,s) \in \mathcal{Z}^2 \ \forall \pi \in \Pi^{rs} \tag{5}$$

Here $d^{rs}$ is the demand for OD pair $(r, s)$, $h_\pi$ is the path flow for path $\pi$ and $c^*_{rs}$ is the average path disutility for OD pair $(r, s)$. Then adding the constraints for multiple OD pairs, flow dependent travel times we get most likely path flow formulation. Now Lagrangianizing the constraints (3), (4) and then analyzing the optimality conditions result in the logit loading where the probability of path flows is given by (6). Detailed relationship between most likely path flow and logit loading can be found in Boyles et al. [2022].

$$p^\pi = \frac{\exp(-\theta c^\pi)}{\sum_{\pi'\in\Pi} \exp(-\theta c^{\pi'})} = \prod_{(i,j)\in\pi} P_{ij}^{\delta^\pi_{ij}} \tag{6}$$

where, $P_{ij} = \Pr\left\{\text{Traveller uses } (i,j) \;\middle|\; \text{Traveller passed through } i\right\}$ and $\delta^\pi_{ij}$ is the number of times link $(i, j)$ is used in path $\pi$ (Note: if path $\pi$ is cyclic then there exists at least one link $(i, j) \in \pi$ such that $\delta^\pi_{ij} > 1$). Again, we assume that the link or path flows do not affect the disutility of a path since the number of bicyclists and pedestrians flow is usually small. The Markov property holds for the logit formula which implies the segment substitution property [Boyles et al., 2022]. Satisfying the segment substitution property for a set of reasonable paths imply that for any pair of reasonable paths which pass through the same two nodes, the paths formed by exchanging the segments between those nodes are also reasonable. We use the notation $\bigoplus$ to indicate joining segments. For example, if paths $\pi = \sigma_1 \bigoplus \sigma_2 \bigoplus \sigma_3$ and $\pi' = \sigma'_1 \bigoplus \sigma'_2 \bigoplus \sigma'_3$ are reasonable with $\sigma_2$ and $\sigma'_2$ starting and ending at the same node, then the paths $\sigma_1 \bigoplus \sigma'_2 \bigoplus \sigma_3$ and $\sigma'_1 \bigoplus \sigma_2 \bigoplus \sigma'_3$ are also reasonable. Let the set of segments be

$$\Sigma_{ab} = \{\sigma_1, \sigma_2, \cdots, \sigma_n\} \qquad \forall (a,b) \in N^2 \tag{7}$$

where, $\sigma_i = [a, \cdots, N_1] \bigoplus [N_1, \cdots, N_2] \bigoplus [N_2, \cdots, b]$. Notice that any other subpath starting in node $N_1$ and ending in node $N_2$ can substitute the middle term here to generate a different segment.

To simplify the logit loading further define the quantity $V_{ab}$ by

$$V_{ab} = \sum_{\sigma\in\Sigma_{ab}} \exp(-\theta c^\sigma) \qquad \forall (a,b) \in N^2 \tag{8}$$

where, $c^\sigma = \sum_{(i,j)\in\sigma} c_{ij}$ represents the cost or disutility of segment $\sigma$ and $c_{ij}$ is the cost or disutility for link $(i,j)$. By the Chapman-Kolmogorov equations the $n^{\text{th}}$ power of an adjacency matrix gives the number of walks or segments of length $n$. So, $V$ can be written as

$$V = \sum_{n=1}^{\infty} L^n = L + L(\sum_{n=1}^{\infty} L^n) = L + LV \tag{9}$$

$$\implies V = L(1-L)^{-1} = [[V_{ab}\ \forall b \in \mathcal{N}]\ \ \forall a \in \mathcal{N}] \tag{10}$$

where $L$ is the adjacency matrix of the weighted graph $\mathcal{G}$ and the weights are the cost of the links. Now eq. (10) allows us to calculate the quantity $V_{ab}$ for all $(a,b) \in \mathcal{N}^2$ which allows to further simplify the logit formula. Now the link flows can be obtained efficiently using eq. (6) and eq. (8), considering all possible path sets including the cyclic paths using the following equation

$$x_{ij} = \sum_{(r,s)\in\mathcal{Z}^2} \sum_{\pi\in\Pi^{rs}} \mathbb{1}_{\{ij\in\pi\}} d_{rs} p^\pi = \sum_{(r,s)\in\mathcal{Z}^2} \sum_{\pi\in\Pi^{rs}} \mathbb{1}_{\{ij\in\pi\}} d_{rs} \frac{\exp(-\theta c^\pi)}{V_{rs}} \tag{11}$$

$$= \sum_{(r,s)\in\mathcal{Z}^2} \frac{d_{rs}}{V_{rs}} \sum_{\sigma_1\in\Sigma_{ri}} \sum_{\sigma_2\in\Sigma_{js}} \exp\left(-\theta(c^{\sigma_1} + t_{ij} + c^{\sigma_2})\right) \tag{12}$$

$$= \sum_{(r,s)\in\mathcal{Z}^2} \frac{d_{rs}}{V_{rs}} \left[\sum_{\sigma_1\in\Sigma_{ri}} \exp(-\theta c^{\sigma_1})\right] \exp(-\theta t_{ij}) \left[\sum_{\sigma_2\in\Sigma_{js}} \exp(-\theta c^{\sigma_2})\right] \tag{13}$$

$$= \sum_{(r,s)\in\mathcal{Z}^2} d_{rs} \frac{V_{ri}\left[\exp(-\theta t_{ij})\right] V_{js}}{V_{rs}} \tag{14}$$

Replacing $t_{ij}$ with $\psi_i^r + t_{ij} - \psi_j^r$ to avoid taking the exponential of large number we get,

$$L_{ij}^r = \exp(-\theta t_{ij}) = \exp(-\theta(\psi_i^r + t_{ij} - \psi_j^r)) = \exp(\theta(\psi_j^r - \psi_i^r - t_{ij}))$$

where $\psi_i^r$ is node label set using Dijkstra's shortest path algorithm for one to all label setting. The superscript for $L$'s and $V$'s denote the source node for Dijkstra's label setting algorithm. To capture the effects of node labels using Dijkstra's algorithm from different source, eq. (14) is rewritten as

$$x_{ij} = \sum_{(r,s)\in\mathcal{Z}^2} d_{rs} \frac{V_{ri}^r L_{ij}^r V_{js}^r}{V_{rs}^r} \qquad \forall (i,j) \in \mathcal{E} \tag{15}$$

An initial OD matrix $\tilde{\mathbf{d}}$ can be obtained from trip distribution using the gravity model with matrix rebalancing using $\mathcal{G}, P, A$ and $T$. Since link counts $\tilde{x}_{ij}$ at links $\tilde{\mathcal{E}} \subset \mathcal{E}$ are available from mobile data sources, the estimation problem can be formulated as the following program.

$$\min \qquad f(\mathbf{d}) = \sum_{(i,j)\in\tilde{E}} \left( \sum_{(r,s)\in\mathcal{Z}^2} d_{rs} \frac{V_{ri}^r L_{ij}^r V_{js}^r}{V_{rs}^r} - \tilde{x}_{ij} \right)^2 \tag{16}$$

$$P_r = \sum_{s\in\mathcal{Z}} d_{rs} \qquad \forall r \in \mathcal{Z} \tag{17}$$

$$A_s = \sum_{r\in\mathcal{Z}} d_{rs} \qquad \forall s \in \mathcal{Z} \tag{18}$$

$$d_{rs} \geq 0 \qquad \forall (r,s) \in \mathcal{Z}^2 \tag{19}$$

Equation (16) minimizes the sum of squared errors between the partially available link flows and the link flows from the OD matrix. Equations (17) and (18) ensures production and attraction consistency and eq. (19) says that the OD matrix entries must be non-negative. The Hessian of the objective function is positive semi-definite which implies that we have a convex objective function. Since eqs. (17) to (19) are linear, the formulation is a convex program.

To solve this program first the eqs. (17) and (18) will need to be Lagrangianized with Lagrange multipliers $\alpha_r$ and $\beta_s$ respectively. That gives us the following

$$\mathcal{L}(\mathbf{d}, \alpha, \beta) = \sum_{(i,j)\in\tilde{\mathcal{E}}} \left( \sum_{(r,s)\in\mathcal{Z}^2} d_{rs} \frac{V^r_{ri} L^r_{ij} V^r_{js}}{V^r_{rs}} - \tilde{x}_{ij} \right)^2 + \sum_{r\in\mathcal{Z}} \alpha_r (P_r - \sum_{s\in\mathcal{Z}} d_{rs}) + \sum_{s\in\mathcal{Z}} \beta_s (A_s - \sum_{r\in\mathcal{Z}} d_{rs}) \tag{20}$$

**Algorithm 1** Projected Gradient Descent.

1: Initialize $\mathbf{d} \leftarrow \tilde{\mathbf{d}}$ using trip distribution
2: Determine the step size $t$ by a dynamic step size rule
3: Update $d \leftarrow \max\{0, d - t\nabla_d \mathcal{L}\}$
4: Update $\alpha \leftarrow [\alpha - t\nabla_\alpha \mathcal{L}]$
5: Update $\beta \leftarrow [\beta - t\nabla_\beta \mathcal{L}]$
6: Terminate if converged, else go to 2.

We present the solution algorithm developed using projected gradient descent and dynamic step size rule to solve the formulated optimization problem. Several properties of the optimization problem are also discussed in our paper. We present how the solution algorithm can be implemented using tensor operations as well. Then we show how the tensor operations can be performed using modern GPU. To determine the descent direction we first needed to compute the partial derivatives with respect to different OD pairs. Then using those partial derivatives a descent direction matrix need to be constructed for an iteration of the solution algorithm. We demonstrate how these numerous partial derivatives can be computed in parallel using tensor operations on a GPU. The large number of cores in modern GPUs allow processing large of blocks of data and simple computations in parallel faster than a central processing unit which usually has a significantly low core count. Being able to express and solve parts of the solution algorithm using tensors in GPUs gives a significant boost in performance in terms of speed. Furthermore, we present the performance of the solution algorithm for a real network. We also present the accuracy of the traffic flows estimated by our approach compared to manually collected traffic flow data for the real network.

## References

Stephen D. Boyles, Nicholas E. Lownes, and Avinash Unnikrishnan. *Transportation Network Analysis*, volume 1. 0.90 edition, 2022.

# A quasi-dynamic multi-region MFD stochastic user equilibrium model

Lawrence Christopher DUNCAN *[,1], Thomas Kjær RASMUSSEN [1], David Paul WATLING [2]

[1] Department of Technology, Management and Economics, Technical University of Denmark, Denmark
[2] Institute for Transport Studies, University of Leeds, United Kingdom

**Short summary**
Multi-region Macroscopic Fundamental Diagram (MFD) traffic equilibrium models have been developed as a computationally efficient alternative to classical link-based traffic equilibrium models with full disaggregate network representation. However, there are theoretical and behavioural limitations of existing multi-region MFD traffic equilibrium approaches, which thus far have been based on dynamic traffic assignment. Specifically, there are questions over how well-behaved the proposed approaches are in terms of producing continuous model outputs (crucial for real-life applications of the model), as well as being able to capture how drivers may base their regional path choices on 'realised' region travel times actually experienced, rather than instantaneous travel times at the time of departure. Motivated by these limitations, in this extended abstract we develop a novel quasi-dynamic multi-region MFD stochastic user equilibrium model, and demonstrate its theoretical and behavioural properties.

**Key Words:** macroscopic fundamental diagram, multi-regional system, quasi-dynamic traffic equilibrium model, traffic propagation, realised travel times

## 1 Introduction

A common issue for transport policy analyses that model travel behaviour on large-scale (e.g. national-scale) road networks, is that applying traditional route choice traffic assignment models to large networks with potentially hundreds of thousands of links can result in very/infeasibly large computation times. This presents a significant challenge for performing rigorous analyses of potential schemes. Multi-region Macroscopic Fundamental Diagram (MFD) traffic equilibrium models have therefore been developed as a computationally efficient alternative to traditional route choice traffic assignment models. Rather than dealing with hundreds of thousands of links as in detailed large-scale network models, the aim is to be able to work with significantly fewer regions to represent the underlying link network. This should dramatically reduce computation times by perhaps several orders of magnitude.

Thus far, multi-region MFD traffic equilibrium models have focused on dynamic traffic assignment, or (as Batista & Leclercq (2019) term) a 'quasi-static' approximation of dynamic traffic assignment. Yildirimoglu & Geroliminis (2014) developed the first such model, where a Multinomial Logit (MNL) regional path choice model is adopted for traffic flow equilibration along with a stochastic network loading procedure to estimate time-dependent regional trip lengths. Batista & Leclercq (2019) later developed a Monte Carlo Simulation (MCS)-based approach for the traffic equilibrium based on monte carlo simulations of trip length / MFD distributions to account for variability in these attributes. Mariotte et al (2020) try two different approaches for dictating the regional path choice: a Wardrop user equilibrium (Wardrop, 1952) variant, as well as optimising the regional path choices to fit MFD production-point data (circumventing the use of choice models). Extending these works, Batista et al (2021) recently developed a heuristic approach for updating the traffic-dependent trip lengths / regional paths during the dynamic traffic assignment, while numerous methods have been proposed for dynamically modelling the transfer of traffic flow at region borders (e.g. Yildirimoglu & Geroliminis, 2014; Mariotte & Leclercq, 2019; Mariotte et al, 2020).

Although the existing dynamic multi-region MFD traffic equilibrium approaches are promising methods for aggregately modelling traffic in urban areas, and improvements continue to be made, there are several potential theoretical and practical limitations. In particular, we highlight two of such limitations.

*Firstly*, there are questions over how well-behaved the proposed approaches are in terms of producing continuous model outputs (crucial for real-life applications of the model). By this we mean that whether for a given grain of the time intervals, regional path choice probability outputs are continuous with varying e.g. model parameters, regional trip lengths, travel demands, MFD functions. In Yildirimoglu & Geroliminis (2014) the random draw element of the stochastic network loading procedure for estimating time-dependent regional trip lengths is likely to mean that outputs are not continuous. And, since the monte-carlo simulation regional path choice model in Batista & Leclercq (2019) operates with discrete distributions, continuity may also not be guaranteed. It is yet to be verified whether either version of the Mariotte et al (2020) approach can produce continuous outputs, though there are other issues with the approach, for example dictating the regional path choices at every time interval based on a single equilibration of average travel times across the entire runtime period.

*Secondly*, an assumption that has been made by all approaches thus far, has been that the regional path choices at each departing time interval are based on the current/instantaneous region travel times at that interval. However, the traffic state will inevitably evolve during a driver's journey, so that the traffic state actually experienced in later regions of the regional path may be considerably different from when the driver departed. With the scale of area considered so far, which has typically been a single city area (e.g. Lyon (Mariotte et al, 2020)), perhaps this issue may not be as

prevalent, but at (sub)national scale where the journey distances are much longer, traffic conditions will undoubtedly change quite significantly throughout a driver's trip, which will likely impact regional path choice.

Addressing these limitations, in this extended abstract we propose a novel *Quasi-Dynamic* Multi-Region MFD Stochastic User Equilibrium (QD-MR-MFD-SUE) model. This new QD-MR-MFD-SUE model is inspired by different quasi-dynamic approaches developed for link-based traffic modelling, such as that developed by Nakayama & Connors (2013). It is not a case, however, of directly transferring one of these approaches to a regional setting, new ideas and concepts are developed to suit the regional setting and stochastic user equilibrium approach. Due to restrictions in the length of this extended abstract, we shall not present a thorough detailing/formulation of the model with mathematical notation. Instead, we shall describe in words the modelling concepts and model features.

The structure of the extended abstract is as follows. In Section 2 we detail the setup of the multi-region MFD system. In Section 3 we introduce the QD-MR-MFD-SUE model and demonstrate its theoretical and behavioural properties. In Section 4 we conclude the extended abstract.

## 2 The multi-region MFD system

The study area is partitioned into a set of regions. The traffic conditions in each region are described by a well-defined speed-MFD function that maps accumulation (number of vehicles in the region) to average speed of the vehicles in the region. As accumulation increases, average MFD speed decreases. There is a set of both internal and external Origin-Destination (OD) movements, i.e. trips originating and destinating in the same region and trips originating in one region and destinating in another, respectively. A regional path is defined as a sequence of regions traversed when travelling an OD movement. The total runtime period of the system is split into a set of equal-sized discrete time-slices. The travel demands for each regional OD movement departing at a given time-slice, are obtained by aggregating travel demands from the actual network ODs over the time-slice between the OD regions. The travel demand is split among the available regional paths according to a regional path choice model, to give regional path flows. Due to the different underlying routes on the actual network typically travelled by drivers taking different regional paths, the region lengths are assumed to be regional-path-specific. Moreover, due to for example drivers taking more detoured routes during more congested peak times, the regional trip lengths may also be time-slice-specific. For a given accumulation level and thereby average MFD speed in a region at a given moment in time, the travel time of a region when travelling a particular regional path at that moment in time, is obtained by dividing the (regional path and time-slice specific) region length by MFD speed.

## 3 The Quasi-Dynamic MFD Stochastic User Equilibrium model

The key features/assumptions of the QD-MR-MFD-SUE model are as follows:

- The regional path choice sets and (regional path and time-slice specific) regional trip lengths are fixed, determined prior to the assignment as inputs to the system.
- The travel demand at each time-slice departs uniformly and continuously.
- The average speed in a region is assumed to be constant during each time-slice, where the average speed measure is based on snapshots of MFD accumulations within the time-slice (akin to the work of Batista & Leclercq (2019)).
- The regional path traffic flow that is due to be residual in a region at the end of a time-slice is assumed to arrive continuously and uniformly to the next region in the regional path during the next time-slice.
- By tracking the journeys of the regional path flow departing at each time-slice, regional path choices are based on realised region travel times.
- Since the travel demand departs at different times throughout the departing time-slice, the regional path flow will experience different region travel times depending on when it departs. The regional path choices at a departing time-slice are therefore based on the average realised travel times experienced by the flow.

The system is inherently full of feedbacks where everything is connected. For example, drivers make regional path decisions based on the region travel times they will experience, but the traffic states (and thereby the region travel times) they will experience depend in turn on the regional path decisions. Moreover, the traffic states experienced are not only dependent on their own regional path decisions, but also the regional path decisions of all drivers travelling all regional paths departing at all time-slices. Consequently, an iterative solution method is needed to identify the equilibrium, to iteratively feedback between a Traffic Propagation Stage and Flow (Regional Path Choice) Updating Stage.

The system at equilibrium can be briefly described as follows. At equilibrium, the region travel times at all time-slices are known to all drivers. At the beginning of time-slice 0 (the initial time-slice) the network is empty (zero accumulation in all regions). Given the average region travel times that will be experienced, drivers departing during time-slice 0 make a regional path choice. The flows travelling each of the regional paths depart uniformly and continuously throughout time-slice 0. Depending on when it departs, the flow will be in different regions of the regional path at the end of time-slice 0. This is the residual flow/accumulation. Thus, given how the regions go from zero

accumulation to the residual accumulation, one can take snapshots of the accumulation level at different points throughout time-slice 0, and for each accumulation snapshot, a corresponding MFD speed can be computed. By averaging the snapshots of MFD speeds one can compute an average speed travelled in the region during the time-slice, and thereby estimate the region travel times at time-slice 0.

Now, at the beginning of time-slice 1 the network is no longer empty, as there are residual accumulations. Moreover, as well as the flows departing during time-slice 1, there are also the residual flows that are continuing their journey to the next region in the regional path, which are assumed to arrive uniformly and continuously to the next region. Given the average region travel times that will be experienced, drivers departing during time-slice 1 make a regional path choice, while the residual flows from time-slice 0 continue along the same regional path. Depending on when the originating flow departs / residual flow arrives and continues, the flow will again be in different regions of the regional path at the end of time-slice 1. Therefore, given how the regions go from the residual accumulation at the end of time-slice 0 to the residual accumulation at the end of time-slice 1, one can take snapshots of accumulation levels throughout time-slice 1, and thereby compute an average MFD speed and estimate region travel times.

This system is repeated for all time-slices. To demonstrate the traffic propagation, consider a 3-region system where there is 1 OD movement and 1 regional path from region 1 to region 3. Suppose that the travel demand / regional path flow departing at time-slices 0 and 1 are 6 and 0, respectively, the time-slice duration is 8 minutes, and that the travel times for regions 1-3 are 4, 2, & 4 at time-slice 0 and 5, 4, & 6 at time-slice 1, respectively. Fig. 1A demonstrates the flow propagation at time-slice 0. As shown, the flow departing after minute 4 into time-slice 0 will not make it out of region 1 by the end of the time-slice and will therefore be residual. Flow departing between minutes 2 and 4 will be residual in region 2, and flow departing before minute 2 will be residual in region 3. Thus, $\frac{4}{8}$ths of the regional path flow will be residual in region 1, $\frac{2}{8}$ths of the flow will be residual in region 2, and $\frac{2}{8}$ths of the flow will be residual in region 3. The residual accumulations for regions 1-3 are therefore $\frac{4}{8} \times 6 = 3$, $\frac{2}{8} \times 6 = 1.5$, and $\frac{2}{8} \times 6 = 1.5$, respectively.

Fig. 1B demonstrates how the residual flow from time-slice 0 propagates at time-slice 1. As per the assumptions of the model, the residual accumulation in region 1 from time-slice 0 arrives uniformly and continuously to region 2 during time-slice 1, the residual accumulation in region 2 arrives uniformly and continuously to region 3 during time-slice 1, and the residual accumulation in region 3 destinates uniformly and continuously. By the end of time-slice 1, all flow will have left region 1 and therefore the residual accumulation is 0. For region 2, the flow arriving before minute 4 into time-slice 1 will be in region 3 at the end of time-slice 1, and the flow arriving after will be residual in region 2. For region 3, the flow arriving before minute 2 into time-slice 1 will destinate, and the flow arriving after will be residual in region 3. Thus, $\frac{4}{8}$ths of the arriving flow to region 2 will be residual in region 2, and the rest will be residual in region 3. $\frac{6}{8}$ths of the arriving flow to region 3 will also be residual in region 3. The total residual accumulations for regions 2 & 3 at the end of time-slice 0 are therefore $\frac{4}{8} \times 3 = 1.5$ and $\frac{4}{8} \times 3 + \frac{6}{8} \times 1.5 = 2.625$, respectively.

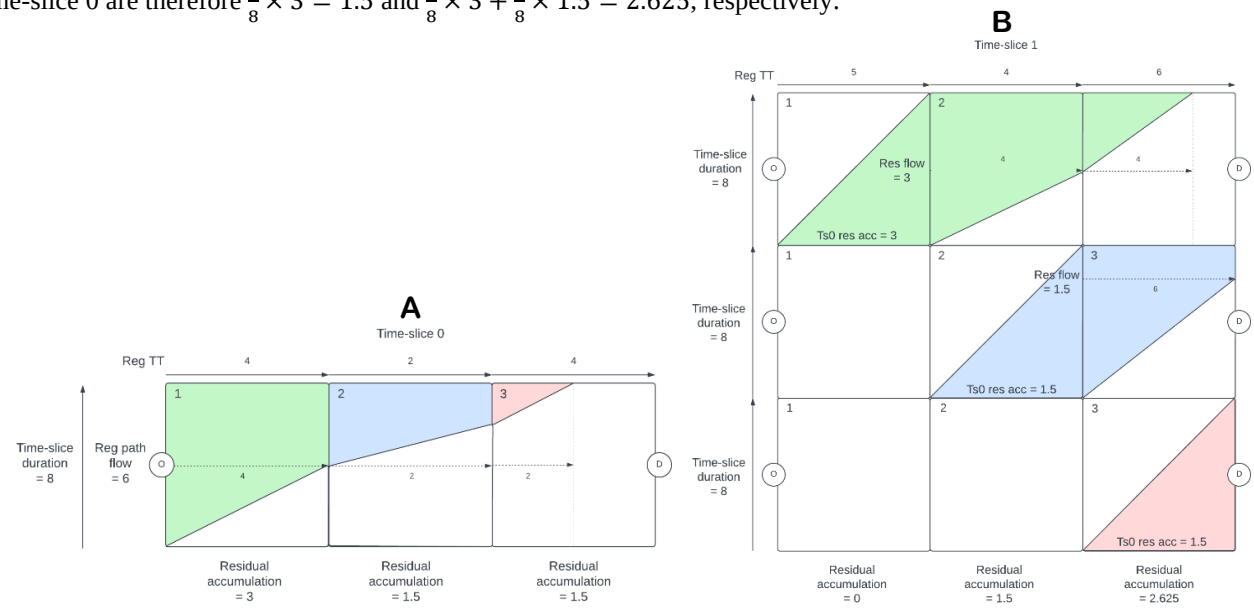



Fig. 1. Demonstrating flow propagation in the 3-region system. **A:** Time-slice 0. **B:** Time-slice 1.

For the average MFD speed measure, one can imagine taking snapshots of the accumulation level at different points throughout time-slices 0/1. For time-slice 0, only the departing regional path flow contributes to the accumulation levels in regions 1-3. For example, taking snapshots at minutes 4, 5, & 6 will yield accumulation levels of 0, 0.75, and 1.5 in region 2, respectively. For time-slice 1, the different bundles of residual flow from time-slice 0 contribute to the

accumulation levels in each of the regions. As shown, the residual flow in region 1 contributes to the accumulation levels in regions 1, 2, & 3 during time-slice 1, the residual flow in region 2 contributes to the accumulation levels in regions 2 & 3, and the residual flow in region 3 contributes to the accumulation levels in region 3.

For the average realised travel times, one tracks the total regional path flow that experiences the travel time of a region at each time-slice, and thereby computes the average travel time of the region experienced by the regional path flow. For example, of the regional path flow of 6 departing at time-slice 0, all 6 experienced the travel time of region 1 during time-slice 0 (4), while 3 experienced the travel time of region 1 during time-slice 1 (5). The average realised travel time of region 1 is therefore $\frac{6}{3+6} \times 4 + \frac{3}{3+6} \times 5 = 4.33$.

A key feature of the QD-MR-MFD-SUE model is that it produces continuous choice probability outputs. Due to the length restrictions of this extended abstract, we cannot fully detail the specifications of the multi-region system, but Fig. 2 displays for a 49-region large-scale $7 \times 7$ grid regional system (e.g. representing a detailed city-area or county/country-area), how the equilibrated choice probabilities of 6 regional paths (for a single OD) vary according the $\theta$ Logit scaling parameter, when using a Multinomial Logit choice model to dictate the regional path choices within QD-MR-MFD-SUE. As shown, the probabilities are continuous with varying $\theta$.

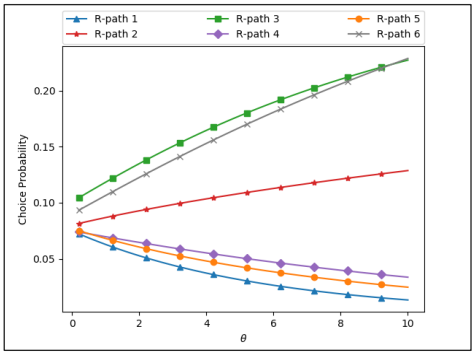



Fig. 2. Demonstrating how QD-MR-MFD-SUE model outputs are continuous, in this case with varying the Logit scaling parameter $\theta$, on a 49-region large-scale $7 \times 7$ grid regional system.

## 4 Conclusions and further research

The extended abstract has proposed a novel QD-MR-MFD-SUE model that addresses some limitations of existing multi-region MFD traffic equilibrium models. Namely, the QD-MR-MFD-SUE model produces continuous model outputs, and can account for regional path choices being based on realised region travel times actually experienced throughout the journey (rather instantaneous travel times at the time of departing). In terms of further research, as shall be elaborated on at the conference, the authors are currently in the process of applying the QD-MR-MFD-SUE model to a real-life large-scale case study of Zealand, Denmark, where: a) there are new types of regions to consider (e.g. urban, rural, and motorway, each requiring different MFD functional forms), b) there are significantly more regions than studied thus far (max being 10 so far in Mariotte et al (2020)), and c) trips span over greater distances (e.g. supporting the need to base regional path choice on 'realised' regional travel times). Furthermore, an estimation procedure is being developed for calibrating the behavioural parameters of underlying regional path choice models, which will then be used to calibrate/validate different QD-MR-MFD-SUE models in the real-life case-study.

## 5 References


Batista S & Leclercq L, (2019). Regional Dynamic Traffic Assignment Framework for Macroscopic Fundamental Diagram Multi-regions Models. Transportation Science, 53(6), p.1563-1590.

Batista S, Leclercq L, & Menendez M, (2021). Dynamic Traffic Assignment for regional networks with traffic-dependent trip lengths and regional paths. Transportation Research Part C: Emerging Technologies, 127.

Geroliminis N & Daganzo C, (2008). Existence of urban-scale macroscopic fundamental diagrams: Some experimental findings. Transportation Research Part B: Methodological, 42(9), p.759-770.

Mariotte G & Leclercq L, (2019). Flow exchanges in multi-reservoir systems with spillbacks. Transportation Research Part B: Methodological, 122, p.327–349.

Mariotte G, Leclercq L, Batista F, Krug J, & Paipuri M, (2020). Calibration and validation of multi-reservoir MFD models: A case study in Lyon. Transportation Research Part B: Methodological, 136, p.62-86.

Nakayama S & Connors R, (2014) A quasi-dynamic assignment model that guarantees unique network equilibrium, Transportmetrica A: Transport Science, 10:7, 669-692.

Wardrop J, (1952). Some theoretical aspects of road traffic research. Proc. Inst. Civil Eng., 1, p.325–378.

Yildirimoglu M & Geroliminis N, (2014). Approximating dynamic equilibrium conditions with macroscopic fundamental diagrams. Transportation Research Part B: Methodological, 70, p.186–200.

# INTEGRATED ACTIVITY-BASED MODELS WITH DYNAMIC MULTIMODAL SIMULATION-ASSIGNMENT WITH MACROSCOPIC CONGESTION FUNCTIONS

**Yuhan Zhou**
Graduate Research Assistant
Transportation Center, Northwestern University
600 Foster St, Evanston, IL 60208
Email: yuhan.zhou@u.northwestern.edu

**Hani Mahmassani, Corresponding Author**
William A. Patterson Distinguished Chair in Transportation,
Transportation Center, Northwestern University
600 Foster St, Evanston, IL 60208
Email: masmah@northwestern.edu
Tel: 847.491.2276

## EXTENDED ABSTRACT

Multiple efforts have been devoted to developing comprehensive frameworks that integrate DTA with agent-based activity-based models (ABM) to capture interactions between the demand and supply sides. Several ABM-DTA implementations have been developed and applied with real-world network data, but several gaps in terms of both theory, methodology and computation remain to be narrowed for large-scale applications. Achieving high degrees of fidelity and internal consistency require considerable computational resources as well as effort required for modeling and calibration of both behavior and performance models, especially for multimodal applications. A single DTA iteration could take several hours, and would require multiple iterations of both ABM and DTA components to reach overall equilibrated (or mutually consistent) status. In order to reduce time and computations required to seek convergence in the ABM-DTA integrated framework, our work explored the possibility to use macroscopic traffic models as a complement to DTA network simulations to varying degrees, and examined different strategies to adopt the macroscopic model at different stages and frequencies of the overall iterative algorithmic process.

Our macroscopic modeling is inspired by the concept of network fundamental diagrams (NFD), or network macroscopic fundamental diagrams (NMFD), which establishes a well-defined relationship between network-level speed, density, and flow. The existence of NMFD in the urban road network has been demonstrated with empirical data, which shows that NMFD for a homogeneous urban network has a low-scatter shape and is not sensitive to demand levels. For large-scale heterogenous real-world networks, previous research has established that they can be partitioned into several homogeneous clusters so that NMFDs can be estimated for each sub-network. Our work adapts this concept for travel time prediction, but instead of the zone-level representation used in NMFDs, we examine the relationship between travel speed and demand at the level of origin-destination (OD) pair, which provides a more direct estimation of trip durations according to different demand patterns in the large-scale network.

In our study, we develop a macroscopic model that predicts the average travel speed for a set of trips characterized by a given combination of origin zone, destination zone, and departure time interval (ODT). It is assumed that average travel speed for a given ODT is the same, regardless of different routes that vehicles may choose. In the integrated ABM-DTA framework, the role of this macroscopic model is to provide the direction of potential changes in travel time as traffic demand patterns are dynamically shifted by ABM, circumventing the needs to run DTA simulations using updated demand inputs from ABM in each iteration.

The OD-level macroscopic congestion model introduced in our work explicitly considers traffic demand of the given ODT, as well as additional spatial and temporal effects. In terms of spatial correlation, when two OD pairs are closely connected with each other and share a large number of road links, vehicle flows from one OD pair have significant impact on the speed of the vehicles from the other OD pair. To determine the range of spatial impacts, OD pairs are partitioned into different clusters based on two criteria: the first one is that, within a cluster, each OD pair has strong spatial correlation with the others, and the average trip speed on one OD pair is significantly influenced by vehicle flows on other OD pairs in the same cluster; the second is that, across clusters, the spatial correlation is low, vehicle loading in one cluster have minimal impact on average speeds in other clusters. As for temporal correlation, the congestion resulting from downstream vehicles can propagate and thus decrease the average speed of the upstream vehicles, especially in urban area when trip distances are relatively short and congestion level is high. To address these effects, the speed prediction model for a given ODT includes multiple input variables of vehicle loading, including the given ODT, the sum of the OD cluster, and the previous departure time intervals. The macroscopic models in our work are calibrated using the vehicle trajectory data produced by the mesoscopic simulation tool DYNASMART, though the procedure could be applied using any source of trajectory data.

The proposed framework has been implemented with the large-scale Chicago metropolitan area multimodal network to study morning commuting hours, including both traffic and transit networks. We simulated an extreme scenario where the major commuter rail service in the area is removed completely (reflecting post-Covid collapse in downtown-oriented daily commuting due to changing hybrid work schedules). In this application, transit network is always simulated by agent-based assignment-simulation tool, while road network estimation can be conducted by both DTA simulation and macroscopic model. To explore strategies to adopt macroscopic traffic model in the integrated ABM-DTA framework, we tested three types of road network estimation approaches: (i) to run simulation-based DTA in every iteration; (ii) to use the macroscopic estimation model in every iteration; and (iii) to use DTA and macroscopic model intermittently, meaning that after every certain number of iterations with macroscopic models applied, one simulation-based DTA is conducted.

The first method the baseline model that generates the most detailed road traffic conditions, but this method could take a few hours to complete just one DTA iteration. The second method uses approximated values and only the travel time by car is updated; therefore, the road network measures from this approach are not completely accurate, but it only takes a few minutes for each iteration. The third method is a hybrid mode: after travel time is updated with the macroscopic model for a few iterations, one DTA simulation is used to update more detailed road network

conditions including trip distance. The total computational time of each method is generally proportional to the number of DTA simulations conducted.

We compare the outcome of both travel behaviors and network performance, along with the convergence speed of the integrated framework. As for the resulting behavior and network conditions, we find that the three methods produce similar mode choice and generalized cost for travelers. In terms of convergence, overall, the three methods are able to reach an equilibrium state, though there is no guarantee of uniqueness. The results show that the first two methods have a similar trend of gap-based convergence, while the third hybrid method experiences more fluctuations each time when DTA simulation is used, since it introduces greater changes in traffic conditions. The results of this application suggest as if planners are primarily interested in behavioral changes and transit performance, the macroscopic network models developed here may be sufficient to reflect the effect of demand changes on roadway congestion and travel times. However, if one is interested in more detailed road network impacts, as well as the impact of operational changes that directly affect the road network, then the more detailed mesoscopic DTA simulation-assignment tool should be used, possibly after reaching initial convergence using the macroscopic model.

## REFRENCES


Cortés, C. E., P. Jara-Moroni, E. Moreno, and C. Pineda. Stochastic Transit Equilibrium. Transportation Research Part B: Methodological, Vol. 51, 2013, pp. 29–44. https://doi.org/10.1016/j.trb.2013.02.001.

Halat, H., A. Zockaie, H. S. Mahmassani, X. Xu, and O. Verbas. Dynamic Network Equilibrium for Daily Activity-Trip Chains of Heterogeneous Travelers: Application to Large-Scale Networks. Transportation, Vol. 43, No. 6, 2016, pp. 1041–1059. https://doi.org/10.1007/s11116-016-9724-4.

Lu, C.-C., H. S. Mahmassani, and X. Zhou. Equivalent Gap Function-Based Reformulation and Solution Algorithm for the Dynamic User Equilibrium Problem. Transportation Research Part B: Methodological, Vol. 43, No. 3, 2009, pp. 345–364.

Verbas, İ. Ö., H. S. Mahmassani, M. F. Hyland, and H. Halat. Integrated Mode Choice and Dynamic Traveler Assignment in Multimodal Transit Networks: Mathematical Formulation, Solution Procedure, and Large-Scale Application. Transportation Research Record, Vol. 2564, No. 1, 2016, pp. 78–88.

Verbas, Ö., H. S. Mahmassani, and M. F. Hyland. Gap-Based Transit Assignment Algorithm with Vehicle Capacity Constraints: Simulation-Based Implementation and Large-Scale Application. Transportation Research Part B: Methodological, Vol. 93, 2016, pp. 1–16. https://doi.org/10.1016/j.trb.2016.07.002.

Xu, X., A. Zockaie, H. S. Mahmassani, H. Halat, Ö. Verbas, M. Hyland, P. Vovsha, and J. Hicks. Schedule Consistency for Daily Activity Chains in Integrated Activity-Based Dynamic Multimodal Network Assignment. Transportation Research Record, Vol. 2664, No. 1, 2017, pp. 11–22. https://doi.org/10.3141/2664-02.

Geroliminis, N., and C. F. Daganzo. Macroscopic Modeling of Traffic in Cities. 2007.

Geroliminis, N., and C. F. Daganzo. Existence of Urban-Scale Macroscopic Fundamental Diagrams: Some Experimental Findings. Transportation Research Part B: Methodological, Vol. 42, No. 9, 2008, pp. 759–770. https://doi.org/10.1016/j.trb.2008.02.002.

Saedi, R., M. Saeedmanesh, A. Zockaie, M. Saberi, N. Geroliminis, and H. S. Mahmassani. Estimating Network Travel Time Reliability with Network Partitioning. Transportation Research Part C: Emerging Technologies, Vol. 112, 2020, pp. 46–61.

Verbas, İ. Ö., H. S. Mahmassani, and M. F. Hyland. Dynamic Assignment-Simulation Methodology for Multimodal Urban Transit Networks. Transportation Research Record, Vol. 2498, No. 1, 2015, pp. 64–74. https://doi.org/10.3141/2498-08.

Wardrop, J. G. Road Paper. Some Theoretical Aspects of Road Traffic Research. Proceedings of the Institution of Civil Engineers, Vol. 1, No. 3, 1952, pp. 325–362. https://doi.org/10.1680/ipeds.1952.11259.

Zhang, L., D. Yang, S. Ghader, C. Carrion, C. Xiong, T. F. Rossi, M. Milkovits, S. Mahapatra, and C. Barber. An Integrated, Validated, and Applied Activity-Based Dynamic Traffic Assignment Model for the Baltimore-Washington Region. Transportation Research Record, Vol. 2672, No. 51, 2018, pp. 45–55. https://doi.org/10.1177/0361198118796397.

Zockaie, A., M. Saberi, H. S. Mahmassani, L. Jiang, A. Frei, and T. Hou. Activity-Based Model with Dynamic Traffic Assignment and Consideration of Heterogeneous User Preferences and Reliability Valuation: Application to Toll Revenue Forecasting in Chicago, Illinois. Transportation Research Record, Vol. 2493, No. 1, 2015, pp. 78–87. https://doi.org/10.3141/2493-09.

# Session 7

Mean-Field Differential Congestion Games Involving Human-Driven Vehicles (HDVs) and Connected-and-Autonomous Vehicles (CAVs)
*Sharon Di, Terry Friesz, Xuegang Ban and Vikash Gayah*

A traffic flow model considering the driving behaviour of connected, cooperative and automated vehicles
*Roberta di Pace, Facundo Storani, Rui Jiang and Stefano de Luca*

Network-level Performance of Dedicated Lanes for Autonomous Vehicles in Large-Scale Transportation Systems
*Ehsan Kamjoo, Fatemeh Fakhrmoosavi, Ali Zockaie and Archak Mittal*

# Mean-Field Differential Congestion Games Involving Human-Driven Vehicles (HDVs) and Connected-and-Autonomous Vehicles (CAVs)

DTA2023 Evanston, Illinois, USA

July 2023

Extended Abstract

Terry L. Friesz

Sharon Di

Jeff Ban

Vikash Gayah

Overview. We propose a paper that discusses possible approaches to the descriptive modeling of mixed-autonomous vehicle flows that consists of human-driven vehicles (HDVs) and connected-and-autonomous vehicles (CAVs). We intend to discuss the key features such models should possess and how they might best be formulated, along with computational methods and an overview of recent preliminary modeling efforts by the authors. Special attention will be given to the notion of many-agent differential games and their approximation as mean field games. The presentation will be purposely presented with limited mathematical detail and primarily in prose to make it accessible to the widest possible audience. Conjectures and intuition will be intrinsic to this "discussion" paper.

Introduction. We are concerned in this paper with CAVs under circumstances we refer to as Level 4.5 of driving automation to signify that control is possible under specific but not all circumstances. However, travel latency, departure time, routing, and maneuvering data at levels of granularity, timeliness, and accuracy allowing us to distinguish between merely connected vehicles and autonomous vehicles will be taken for granted. In particular, we imagine that a merely connected vehicle will typically have only an information system and soft route guidance. Moreover, for Level 4.5, although autonomous vehicles will be able to execute a limited set of maneuvers, like lane changing, without driver involvement, using onboard control systems. As such, there will be no centralized control for any vehicles in the landscape of this paper. This choice of focus is based on our belief that Level 5 autonomous vehicles (AVs) with control under all circumstances will take much longer time and become a reality only if appropriate Federal regulations and safety certifications are implemented.

Enhanced transportation planning is the primary motivation for building a descriptive network model of mixed-autonomous traffic[1]. There will always be a need for such a model, owing to the need to forecast future traffic and design future road infrastructure. Scholars taking up the challenge of creating a descriptive model of mixed-autonomous traffic will be extending a long tradition of analytics in the transportation planning process, a history that dates back to the 1950s and includes the development of key aspects of static and differential applied game theory. In particular, creation of a differential congestion game describing mixed-autonomous traffic poses several significant challenges. These are



[1] We will refer to the combined, interacting flow of human driven vehicles (HDVs) and CAVs on the same right-of-way as *mixed-autonomous traffic*.

discussed below. In our final paper, the individual challenges will be treated in greater depth and more exhaustive citations of prior literature provided.

Vehicle classes and interactions. It is our intent to provide in this section a deeper look at the question of how to model vehicle classes that are mixed in nature, including both flow and particle representations of activity on any given network arc. The vehicle classes most relevant to our focus are HDVs, CAVs, and AVs. For certain roadways, time periods, and technologies, it is possible that vehicles from all three classes will interact with one another. These interactions are not yet fully understood but hypotheses regarding them must necessarily be combined with the limited experimental data that is available in order to model mixed-autonomous traffic.

The conventional traffic engineering perspective is that every added vehicle contributes ever increasingly to total delay, allowing the use of volume delay curves or generalized unit cost functions. A single example serves to illustrate limitations posed by such simplicity: suppose that AVs will likely move in continuously expanding and contracting platoons defined by common travel interests that are nonetheless transient; those same platoons will be surrounded by HDVs. To see the implications of such behavior, imagine that two car-following platoons arrive in proximal time at the entrance to some arc they both need to traverse to ensure the path choices of their respective vehicles are realized. The following are some of the possible consequences: (i) the two platoons occupy different traffic lanes of the intended arc and ignore one another; (ii) one of the platoons slows and attaches to the rear of the other platoon; (iii) the individual vehicles within each platoon merge carefully, decelerating and accelerating as required, and thereby come to occupy new positions relative to some leader within the newly formed grand platoon (i.e., the two platoons merge into one bigger platoon). It is rather obvious that Case (iii) involves a much more complex control strategy and holds the greatest risk for an incident (encroachment or collision). Case (iii) also presents the greatest challenges in terms of formulation, mathematical analysis, and computation. None of the cases works very well with the volume delay curves embraced by the traffic engineer. Modeling the maneuver controls in a way that capture Case (iii) behavior yet is computationally tractable for very large networks is also problematic.

One school of thought, typified by Wang et al. (2019) and Ngoduy et al. (2021), seeks to represent AVs as continuous flows that are carried in the same traffic fluid, while the numbers of HDVs and AVs are calculated ex post, but, unfortunately, those numbers may fail to be integers. Guo et al. (2021) further proposed to apply the differential game approach, in particular, the differential complementarity systems, to model the mixed flow of CAVs and HDVs on a dynamic transportation network, in which both CAV and HDV flows are modeled as continuous variables. In their modeling framework, CAVs are controlled to select routes that benefit the entire network (i.e., under the SO principle), while HDVs follow the UE principle to minimize their own travel costs. By properly selecting the traffic flow models for mixed CAV and HDV flows, they cast the problem as a dynamic bilevel formulation and proposed an MPC-based heuristic method to compute the solutions. One fundamental problem underlies these efforts: the HDVs and CAVs are uniformly mixed in every sample of traffic fluid one examines, and the identities of the AVs are lost. In such a circumstance, one cannot directly study CAV control strategies.

The mix of different classes of continuous traffic flows has sometimes been called *multiclass flow* in the scientific literature. It is difficult to say who first coined this terminology, but certainly Dafermos (1972) used it to describe vehicular flows with distinct cost characteristics, albeit in a static environment. Not surprisingly extension of multiclass flow models from a static to a dynamic setting poses no significant

challenges. However, the multiclass road network carrying both HDVs and AVs presents many obstacles to the conventional style of mathematical modeling employed in the study of dynamic traffic networks, as these are exemplified by Friesz et al. (2019), Ma et al. (2018), Han et al. (2019), and Friesz and Han (2023). Among these obstacles, the most daunting arise from the dichotomy in the types of control variables. HDVs are typically modelled as differential variational inequalities that possess solutions known as dynamic user equilibria (DUE) and are articulated in terms of continuous flow variables. DUE may be thought of as the limit of a differential Nash game as the number of agents becomes uncountable; as such, DUE is sometimes referred to as an atomic differential game. On the other hand, AVs are naturally conceptualized as particles (distinct vehicles). Any marriage of continuous flows and discrete vehicles necessarily requires caution, and the HDV/AV traffic modeling problem is no exception.

Choice of traffic physics and animating dynamics. The behavior of traffic on urban and regional road networks has been studied with considerable intensity since the 1950s. The scientific literature records many important contributions to the theory, forecasting, control, and associated computation of traffic flows on road networks. That literature takes a multiphysics, multi-timescale, multi-spatial scale, and multi-agent perspective in a quest to understand and manage travel on road networks. In fact, there are hydrodynamic, kinetic, car-following, game-theoretic, gravitational, and entropy-based models of transportation and trip distribution corresponding to various levels of vehicle aggregation. These, in turn, involve time lags, transitions from crude spatial heterogeneity to organized networks, and competition among diversely minded users who may be individuals or aggregates of like-minded travelers. Yet no general theory of road traffic has been created, much less widely endorsed.

One is left with relatively little guidance in selecting a paradigm for model formulation. The types of vehicle dynamics, as well as the types of competition among vehicles, turn out to be central to building models of multiple classes of traffic. We expect the full paper will at a minimum review the suitability of the following types of dynamics for mixed-autonomous traffic:

(D1) Demand formation dynamics (flows)
(D2) LWR dynamics and its various continuous/discrete equivalences or approximations (flows)
(D3) Car-following dynamics (particles)
(D4) Dynamics with speed as the control variable (particles)
(D5) Dynamics with acceleration as the control variable (particles)

Gaming and competition among agents. The choice of dynamics is intimately connected with the choice of type of game. Consequently, we expect the discussion of dynamics will be integrated with a discussion of gaming assumptions and formulations that are appropriate to mixed autonomous traffic. These will include:

(G1) Differential Game Theory
(G2) Evolutionary game theory (EGT)
(G3) Mean-field games (MFGs) without diffusion
(G4) Mean-field games (MFGs) with diffusion
(G5) Machine learning (ML) and digital twins (DTs), model-based
(G6) Machine learning (ML) and digital twins (DTs), data-based

EGT is a perspective on learning that is useful for forming demand dynamics; see Friesz et al. (2011). MFGs, ML, and carefully crafted DTs have a clear role to play in extracting approximate dynamics from traffic models/data. We will engage in hypothesizing how DTs can help, not only in approximating dynamics, but also in solving the games that describe HDVs and CAVs. It appears that, the dynamics and/or type of games may vary depending on specific scenarios when dynamics/choices of CAVs and HDVs are considered and especially the interactions of the two; developing the dynamic network modeling framework for mixed-autonomous traffic thus hinges on proper selection of such models under different scenarios and a uniform mathematical paradigm to integrate all of them for analysis and computation.

Very significant is that Huang et al. (2021) form a differential Nash game for AVs in continuous time that addresses user equilibrium. They apply the mean field approximation (MFA) to their model. The MFA of a Nash game replaces, for each agent, knowledge of the other agents with their mean response in order to obtain a more computationally tractable formulation. They report excellent accuracy when the MFA is employed to approximate user equilibria. Background on MFG modeling and ideas about how it might play a role in modeling mixed-autonomous traffic will be presented.

Preliminary Results. The envisioned paper will include prior and preliminary numerical results relevant to mixed-autonomous traffic. We will start with a differential game where the decision making of each agent (an HDV or an AV) is modeled as an optimal control problem. The behavioral difference between HDVs and AVs could lie in various aspects, as depicted by *multiclass flow* models. When a game-theoretic approach is used, both types of vehicles could differ in demand formation dynamics, control strategies, vehicle dynamics, and traffic propagation. Our prior modeling efforts assume that the key difference in the HDV and AV behavior lies in the cost functional an agent seeks to minimize. As a consequence, traffic flow models that are commonly used to describe human driving behavior may be reformulated as MFGs with a carefully specified cost functionals. An AV model may be created using alternative functionals that promote smoother traffic flows. However, physical constraints are needed to assure macroscopic AV traffic dynamics are physically real; for example, AVs must not move backward or be too close to the front vehicle.

Networked agents will, of course, interact with one another via individuals' actions and cost functions, as well as via traffic dynamics represented by density or flux. These agents all play a competitive game, in the absence of centralized control. With individual agents' behavior modeled as an optimal control problem, we need to further understand the macroscopic behavior of the mixed-autonomous flow on a road network. Because of a large number of agents present in the mixed flow, how different types of vehicles interact remains a challenge on a microscopic level, and how such interactions can be scaled up to a macroscopic level remains an unanswered scientific question. In keeping with our preceding remarks, the mechanism for the meeting and merging of two platoons at a junction will impact traffic dynamics on an aggregate level. On an edge, we also would like to derive an aggregate delay function that depicts how the presence of one vehicle class would impact the delays faced by other classes. If neither delay curves nor generalized cost functions are available a priori, the question of how a delay function should be derived from the microscopic interaction between vehicles becomes paramount. The converse is also a challenge.

In our prior work, the MFA is applied to derive aggregate behavior of the vehicle population, before a detailed analysis of agent-level interaction can be modeled. Due to the difference in the cost functionals of AVs and HDVs, a multi-population MFG can be developed to depict the interactions between these

vehicle classes. However, as we have explained in prior work, there are many challenges to be addressed when we start from a differential game and move to a mean-field game. To be more specific, the successful use of a gaming paradigm will necessitate, at a minimum, addressing the following:

(G1) If we are to add diffusion to MFGs, it is reasonable to add acceleration noise in the control dynamics. However, incorporating such a noise when macroscopic traffic dynamics are derived requires caution. Otherwise, we will end up with vehicles appearing or disappearing from the middle of an edge with no reasons.

(G2) If traffic environments are unknown or only partially observable, CAVs will need to learn from data while navigating a road network. In our prior work, if the cost function (or reward function) is unknown, CAVs need to learn this award while interacting with the environment continuously. There are a vast number of model-based approaches from the DTA community that may help derive prior knowledge, while data-based approaches would help with the learning and adaptation processes.

Fig. 1 demonstrates a generic framework of travel choices each agent can make (including departure-time, route, speed, and lane choices), as well as how these choices interact with aggregate traffic dynamics in the dynamic network loading stage. We intend for our paper to describe how the classical DTA framework may be expanded to consider the additional choices depicted in Figure 1. The role of decentralized control in making such additional choices feasible will be discussed in the paper that devolves from this abstract. MFGs will be a major consideration of such decentralized control.

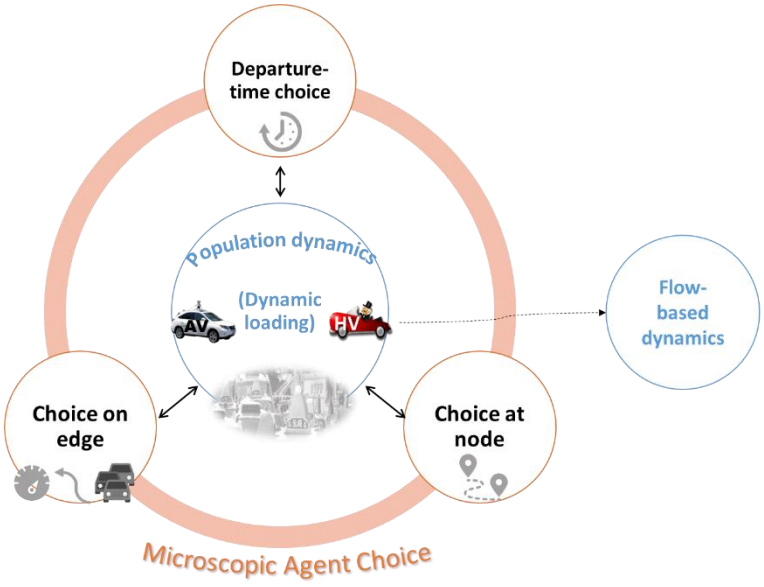



Figure 1: A descriptive network model of mixed-autonomous traffic

We aim to apply MFG to describe both the dynamics and travel choices of individual vehicles/agents (for both HDVs and CAVs). Dynamics include lane/speed choices of CAVs, car-following of HDVs, and platoon forming, merging, and diverging of mixed HDVs and CAVs; travel choices include departure-time and route choices of individual agents, which can be formulated as certain type of games among the agents. As discussed above, the *first key research challenge* is how to select the proper traffic dynamics/interactions

and the best type of games of the mixed flow to construct the MFG for analysis and computation. The aggregate of all individuals' dynamics and travel choices will lead to population-wise, or flow-based, dynamics or patterns, such as equilibrium patterns. It is the *second key research challenge* to derive, understand, and analyze such flow-based dynamics/patterns, emerged from MFG. As proposed in Guo et al. (2021), one may directly simplify the modeling of mixed-autonomous traffic by pure flow-based approaches (i.e., multi-class models). It would be interesting to investigate whether the MFG-based approach we just proposed will lead to flow-based dynamics/patterns that are consistent with (or different from) those by the flow-based models. Formal investigation of the comparison and contrast of the two will be the third key research challenge.

References

Friesz, T. L., Kim, T., Kwon, C., & Rigdon, M. A. (2011). Approximate network loading and dual-time-scale dynamic user equilibrium. *Transportation Research Part B: Methodological*, *45*(1), 176-207.

Friesz, T. L. and Han, K. (2023) *Dynamic User Equilibrium*, Springer, 363 pages (ISBN-13: 978-3031255625)

Guo, Q., Ban, X., Aziz, H., 2021. Mixed traffic flow of human driven vehicles and connected/automated vehicles on a dynamic transportation network. Transportation Research Part C 128, 103159.

Han, K., Eve, G., & Friesz, T. L. (2019). Computing dynamic user equilibria on large-scale networks with software implementation. *Networks and Spatial Economics*, *19*(3), 869-902.

Huang, K., Chen, X., Di, X., & Du, Q. (2021). Dynamic driving and routing games for autonomous vehicles on networks: A mean field game approach. *Transportation Research Part C: Emerging Technologies*, *128*, 103189.

Ma, R., Ban, X., Pang, J.S., 2018. A link-based dynamic complementarity system formulation for continuous-time dynamic user equilibria with queue spillbacks. *Transportation Science 52(3),* 564-592.

# A traffic flow model considering the driving behaviour of connected, cooperative and automated vehicles

R. Di Pace[a1], F. Storani[a], R. Jiang[b], S. de Luca[a]

[a] *Department of Civil Engineering, University of Salerno, Fisciano (Salerno) 84084, Italy*

[b] *Laboratory of Transport Industry of Big Data Application Technologies for Comprehensive Transport, Ministry of Transport, Beijing Jiaotong*

*University, Beijing 100044, China*

## Abstract

This paper proposes an enhanced traffic flow model that can support the presence of human-driven and connected vehicles in a mixed traffic flow condition. It is based on the hybrid model H – CA&CTM; Hybrid Cellular Automata Cell Transmission Model of Storani et al. (2022), which combines the meso-microscopic Cellular Automata model (CA; Nagel and Schreckenberg, 1992) and the macroscopic Cell Transmission Model (CTM; Daganzo, 1994). The whole model can represent: i) the Human-driven vehicles mode, ii) the Adaptive cruise control mode, and iii) the Cooperative adaptive cruise control mode. In more detail: the considered CTM was adapted from Yao et al. (2022), while the CA model was adapted from Jiang et al. (2021). The model was tested on a ring-shaped arc.

*Keywords: traffic flow model; Cell Transmission Model; Cellular Automata; connected and automated vehicles*

## Introduction and motivation

Recent advances in vehicle-to-X communication and faster communication protocols have opened new and challenging scenarios to be investigated. Specific traffic flow models are required to support a suitable representation of vehicle (driver) behaviour, utilizing the information provided by connected vehicles. Collecting and processing information from vehicles approaching junctions and providing them with information requires microscopic models that can deal with each vehicle separately. However, microscopic models are highly detailed, require multiple parameters to set, and are computationally demanding. This problem grows exponentially as network size increases. An alternative approach based on hybrid models can overcome these issues. Traffic flows can be represented by coupling different models, even with a variable aggregation level, switching between a macroscopic, mesoscopic, or microscopic approach. Hybrid traffic flow models are not new in the literature and have been extensively studied for human-driven vehicles, used for traffic analysis at both local and network levels (Jin and Zhang, 2003; Lebacque and Khoshyaran, 2005; Burghout et al., 2005; Yang and Morgan, 2006; Leclercq, 2007; Joueiai et al., 2014).

This paper proposes an enhanced traffic flow model capable of supporting the presence of human-driven and connected vehicles in a mixed traffic flow condition. The proposed model is an enhanced version of the hybrid model H – CA&CTM; Hybrid Cellular Automata Cell Transmission Model proposed by Storani et al. (2022), which combines a meso-microscopic Cellular Automata model (CA; Nagel and Schreckenberg, 1992) for node representation and a macroscopic Cell Transmission Model (CTM; Daganzo, 1994) for link representation. The proposed model in this paper can represent different traffic classes mixed with connected automated vehicles and regular human-driven vehicles, including i) the human-driven vehicle mode, ii) the adaptive cruise control mode, and iii) the cooperative adaptive cruise control mode. The CTM was adapted from Yao et al. (2022), while the CA model was adapted from Jiang et al. (2021).

## Methodology

The proposed hybrid model's general architecture consists of a combination of two sub-models for each link, as depicted in Figure 1. The initial section uses the microscopic Cellular Automata (CA), adapted from Jiang et al. (2021) to model the acceleration of vehicles from the upstream junction. Then, vehicles approach the first transition and are converted into a macroscopic flow modelled with the

([1]) Corresponding author (rdipace@unisa.it)

CTM sub-model adapted from Yao et al. (2022). Later, they are discretized again in the second transition with a microscopic CA approach to model the traffic phenomena upstream of a junction, such as horizontal queue formation and discharge, congestion propagation and acceleration, speed optimization, etc.

The transitions from CA to CTM and vice versa are based on the introduction of spatially extended transition zones comprising both sub-models. Each sub-model is discrete in time and space, and they model the traffic flow using discrete cells that vary their state at each time step. The simulation time step they adopted makes it possible to obtain a consistent queuing formation and backward propagation of congestion.

**Figure 1.** Example of a link with a hybrid approach


*Source:* own elaborations

*From CA to CTM - Transition 1:* This transition receives vehicles from the Cellular Automata upstream, to then insert a flow into the Cell Transmission Model downstream. The transition zone has two subsections, one for each model: the macroscopic CTM section has the same properties as the cells of the CTM model, while the meso-microscopic CA section is an extension of the upstream CA link, with the same properties, parameters, and rules as the CA model.

The length of the transition depends on the minimum outflow capacity of the modified CTM model. If the reaction time of human-driven vehicles is higher than the connected autonomous vehicles, then the minimum outflow capacity is calculated as:

$$Q_{c,HDV} = V_f \cdot k_{c,HDV} = V_f/(V_f \cdot \tau_{HDV} + 1/k_j)$$

Considering a human-driven vehicle, it takes $1/Q_{c,HDV}$ seconds per vehicle to completely cross a cell on a free-flow condition. Then, the transition is composed of $k = 1/Q_{c,HDV} \cdot \Delta t$ cells of CTM model. When a human driven vehicle travelling on the CA model enters on the transition section, the density of the first $k$ CTM cells are updated as follows:

$$k_{HDV,i}(t) = k_{HDV,i}(t) + Q_{c,HDV} \cdot \Delta t \qquad \text{for cells } i = 1 \rightarrow k - 1$$

$$k_{HDV,k}(t) = k_{HDV,k}(t) + 1 - \left((k - 1) \cdot Q_{c,HDV} \cdot \Delta t\right) \qquad \text{for cell } k$$

A connected autonomous vehicle takes $1/Q_{c,CAV}$ seconds per vehicle to completely cross a cell on a free-flow condition. Then, it is necessary to update the first $l = 1/(Q_{c,CA} \cdot \Delta t)$ cells of CTM model. When a connected autonomous vehicle enters on the transition section, the density of the first $l$ CTM cells are updated as follows:

$$Q_{c,CAV} = V_f \cdot k_{c,CAV} = V_f/(V_f\, \tau_{CAV} + 1/k_j)$$

$$k_{CAV,i}(t) = k_{CAV,i}(t) + Q_{c,CAV} \cdot \Delta t \qquad \text{for cells } i = 1 \rightarrow l - 1$$

$$k_{CAV,l}(t) = k_{CAV,l}(t) + 1 - \left((l - 1) \cdot Q_{c,CAV} \cdot \Delta t\right) \qquad \text{for cell } l$$

After updating the densities of the CTM models, the flow between cells is calculated using the CTM procedure. The vehicle on the CA takes the speed of the CTM cells to update its position for the next time step. Once that the transition is completely emptied, it is cancelled from the CA model. This procedure allows upstream vehicles to adjust their speed in accordance to the congestion of the transition.

*From CTM to CA – Transition 2:* This transition receives flow from the Cell Transmission Model and converts it to discrete vehicles in the Cellular Automata downstream. It has a similar layout to the first transition, as it is composed of two subsections with the same rules and procedures as the CTM and CA models. In this case, the cells on the CTM subsection act as accumulation cells. Once the sum of their densities equals that of a vehicle (either CAV or human), a vehicle with those characteristics is created on the CA sub-model, and the densities on the CTM are updated.

## Findings

The proposed model was tested using several numerical applications based on both human-driven vehicles and connected and automated vehicles. Hybrid traffic flow models need to be analysed in terms of local and global (in)consistencies (Joueiai et al., 2013, 2014). Local (in)consistency describes the conservation of vehicles through a local interface, while global (in)consistency refers to the impact of aggregating data (such as average flow, speed, and density) from a microscopic representation into its equivalent macroscopic representation. The main layout considered for testing the proposed model was a ring-shaped link of 5000 m. The ring was used to: i) calibrate and validate each of the two enhanced sub-models, ii) verify the (in)consistencies of the hybrid model, and iii) test the effects of capacity/demand variations on the CA, on the CTM, and on the hybrid model (H - CA&CTM). The CA and CTM models have a free flow speed of 15 m/s and a jam density of 200 veh/km. The CA has a cell length of 0.01vm and a time step of 0.1 s, while the CTM has a cell length of 15m and a time step of 1 s, that verifies the Courant-Friedrichs-Lévy condition.

The analysis of local (in)consistency is presented heuristically for the two transition zones: from the meso-microscopic CA model to the macroscopic CTM and vice versa (see Figures 2 and 3). The columns of the figures represent space, while the rows represent time steps. The CA and CTM sections of the transitions are synchronized. For transition 1, this means that the distance that a vehicle can travel in the CA section depends on the number of vehicles in the transition (the numbers in the CA cells represent the speeds of the vehicles), which affects the vehicles upstream of the transition zone. For transition 2, the supply of the cell of the CTM section depends on the total number of vehicles in this transition zone, affecting the incoming flow from the last cell of the CTM model. As detailed above, there is a spatial overlap between the CA section and the CTM section in both transition zones.

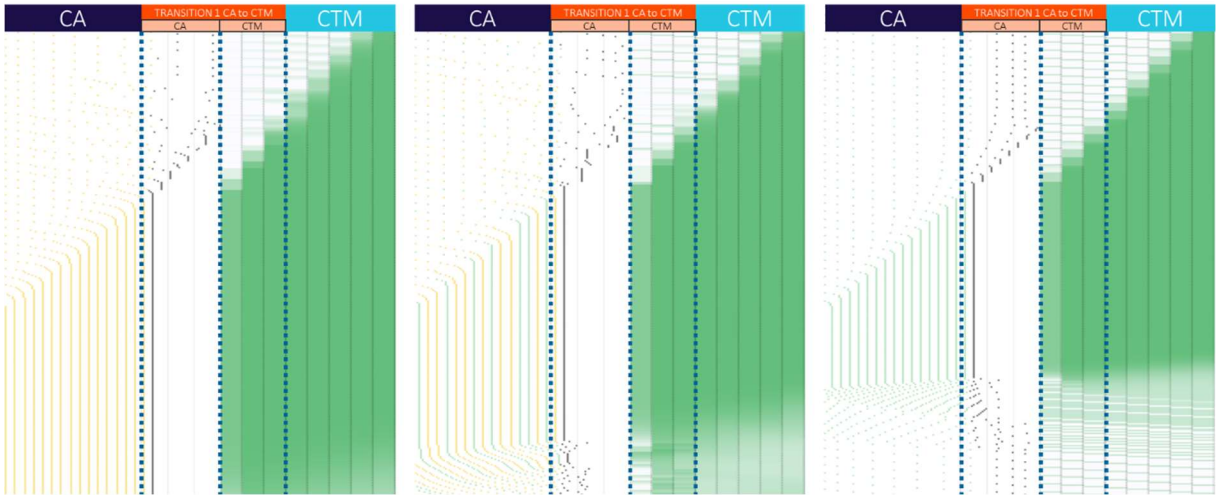
This application has also been used to verify the global (in)consistency of the model by evaluating the conservation of the drivers' behaviour with respect to human-driven and connected and automated (CA) vehicles.

**Table 1.** Vehicle parameters.

| **Parameters** | **Connected Autonomous** | **Human Driven** |
|---|---|---|
| Max speed [m/s] | 15 | 15 |
| Acceleration [m/s$^2$] | 2 | 2 |
| Max deceleration [m/s$^2$] | 5 | 5 |
| Reaction time [s] | 0.6 | 2 |
| Safe distance CACC [m] | 0.50 | - |
| Random deceleration [m/s$^2$] | - | 3 |
| Dawdling probability | - | 0.20 |

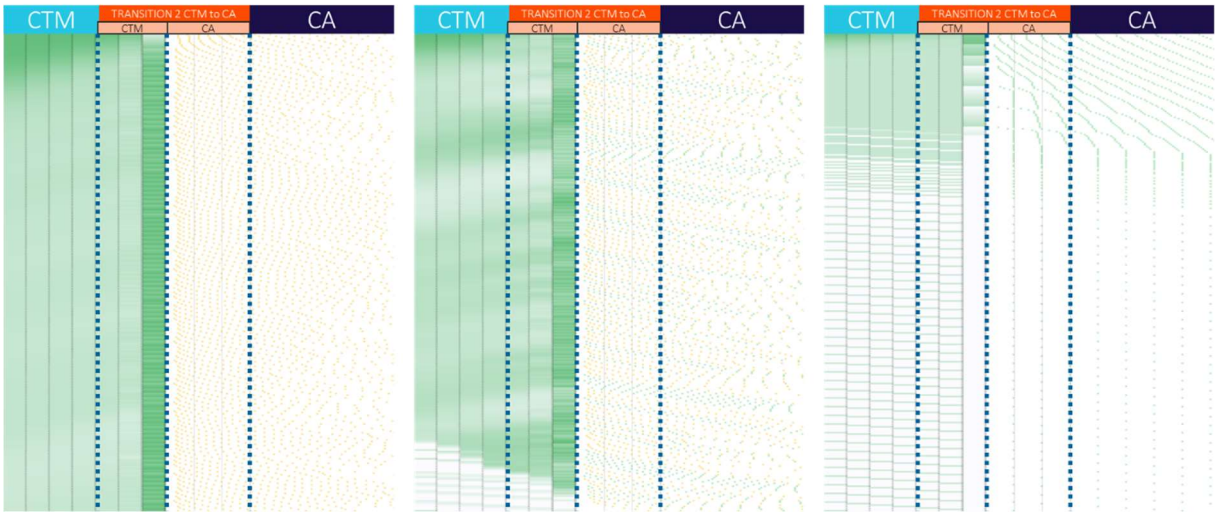
*Source:* Own elaborations.

**Figure 2.** Transition 1 from CA to CTM considering a) 0% CAV, b) 50% CAV, and c) 100% CAV




*Source:* own elaborations

**Figure 3.** Transition 2 from CTM to CA considering a) 0% CAV, b) 50% CAV, and c) 100% CAV




*Source:* own elaborations

## Conclusions and future works

This paper proposes an enhanced traffic flow model capable of supporting the presence of human-driven and connected vehicles in mixed traffic flow conditions. The considered model is an enhanced version of the H-CA&CTM from Storani et al. (2022). The entire model can represent different classes of traffic mixed with connected automated vehicles and regular human-piloted vehicles. It is able to represent: i) the human-driven vehicles mode, ii) the adaptive cruise control mode, and iii) the cooperative adaptive cruise control mode. The considered CTM was adapted from Yao et al. (2022), while the CA model was adapted from Jiang et al. (2021).

The ring layout allowed to test and verify the consistency of the model, to i) calibrate/validate each one of the two enhanced sub-models, ii) verify the (in)consistencies of the hybrid model, and iii) test the effects of the capacity/demand variations on the CA, on the CTM, and on the H-CA&CTM.

In terms of future research, two main perspectives will be investigated: i) the proposed model will be further developed to consider lane-changing, and ii) the model will be integrated with enhanced control strategies (such as vehicle platooning).

## References

Burghout, W., Koutsopoulos, H., Andreasson, I., 2005. Hybrid Mesoscopic-Microscopic Traffic Simulation. Transportation Research Record: Journal of the Transportation Research Board 1934, 218–255

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, 269–287

Jiang, Y., Wang, S., Yao, Z., Zhao, B., Wang, Y., 2021. A cellular automata model for mixed traffic flow considering the driving behavior of connected automated vehicle platoons. P*hysica A: Statistical Mechanics and its Applications* 582, 126262

Jin, W., Zhang, H., 2003. On the distribution schemes for determining flows through a merge. Transportation Research Part B: Methodological 37, 521–540

Joueiai, M., Van Lint, H., Hoogendoorn, S., 2013. Generic solutions for consistency problems in multi-scale traffic flow models - Analysis and preliminary results, in: IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC, IEEE. pp. 310–315

Joueiai, M., van Lint, H., Hoogendoom, S.P., 2014. Multiscale Traffic Flow Modeling in Mixed Networks. Transportation Research Record: Journal of the Transportation Research Board 2421, 142–150

Lebacque, J.P., Khoshyaran, M.M., 2005. First-order macroscopic traffic flow models: Intersection modeling, network modeling, in: Transportation and Traffic Theory. Flow, Dynamics and Human Interaction. 16th International Symposium on Transportation and Traffic Theory University of Maryland, College Park, pp. 365–386.

Leclercq, L., 2007. Hybrid approaches to the solutions of the "Lighthill–Whitham–Richards" model. Transportation Research Part B: Methodological 41, 701–709

Nagel, K., Schreckenberg, M., 1992. A cellular automaton model for freeway traffic. Journal de Physique I 2, 2221–2229

Storani, F., Di Pace, R., de Luca, S., 2022. A hybrid traffic flow model for traffic management with human-driven and connected vehicles. Transportmetrica B: Transport Dynamics 10, 1151–1183

Yang, Q., Morgan, D., 2006. Hybrid traffic simulation model, in: Transportation Research Board 85th Annual Meeting, Washington DC, United States, pp. 06–2582.

Yao, Z., Jin, Y., Jiang, H., Hu, L., & Jiang, Y. (2022). CTM-based traffic signal optimization of mixed traffic flow with connected automated vehicles and human-driven vehicles. *Physica A: Statistical Mechanics and Its Applications*, 603, 127708,

# Network-level Performance of Dedicated Lanes for Autonomous Vehicles in Large-Scale Transportation Systems

Ehsan Kamjoo[a], Fatemeh Fakhrmoosavi[b], Ali Zockaie[a,*], Archak Mittal[c]

[a]*Michigan State University, 428 S. Shaw Ln., East Lansing, Michigan, United States, 48824*
[b]*University of Connecticut, Castleman Building, Storrs, Connecticut, United States, 06269*
[c]*Ford Motor Company, 2101 Village Road, Dearborn, Michigan, United States, 48124*

**Corresponding author. Email Address: zockaiea@egr.msu.edu*

Extended Abstract

*Keywords:* Autonomous Vehicles, Traffic Simulation, Dedicated Lanes, Market Penetration Rate

**ABSTRACT**

Autonomous vehicle (AV) dedicated lanes are introduced as an effective strategy to improve mobility in heterogeneous traffic consisting of AVs and heterogeneous drivers. There are several approaches for implementation of these lanes, such as adding a new lane or dedicating an existing lane to AVs. Previous studies investigated the impacts of AV dedicated lanes on throughput at the segment level and/or assumed a fixed route choice. However, AV dedicated lanes affect traffic distribution over the network due to changes route choice behavior. Therefore, this study explores the impacts of AV dedicated lanes on traffic performance of large-scale networks by conducting a network-level cost-benefit analysis on the implementation of AV dedicated lanes under different demand levels, AV Market Penetration Rate (MPR) scenarios, and also locations of their implementation. Various factors including changes in aggregate flow-density relationships, throughput, and average travel times are explored. In this regard, to simulate the mixed traffic, DYNASMART-P is updated to consider AV dedicated lanes. In this study, traffic simulation analyses are performed on the large-scale network of Chicago to explore the impacts of AV dedicated lane implementation on the performance at the corridor level and network level under different AV MPRs, demand levels, AV dedicated lane implementation approaches, and the locations of the dedicated lanes. The results of this study helps transportation planners in using the AV dedicated lanes to improve the efficiency of these vehicles at the early stages of their deployment.

## INTRODUCTION

Autonomous Vehicles (AVs) are shown to have the potential to improve the performance of transportation

systems. The market penetration rate (MPR) of AVs is not expected to increase significantly at the early stages of their deployment. One of the main concerns in this regard is the heterogeneous traffic flow with both AVs and human-driven vehicles (HDVs), which will inevitably exist before AVs completely replace the HDVs in the transportation systems (1). In effect, the cooperation between AVs will be extremely challenging with the low MPRs of AVs in traffic. In practice, HDVs can interrupt the AV platoons, which makes the cooperation between AVs less feasible (2). Therefore, efforts have been directed toward improving the efficiency of AVs at the early stages of their deployment to the extent possible.

Enlightened from the managed lane strategies, AV dedicated lanes have drawn significant attention from policymakers as a practical and effective solution to the abovementioned challenges. AV dedicated lanes attract high AV densities along the road segment, which leads to lower interactions between AVs and HDVs and facilitates forming AV platoons. Therefore, these lanes can amplify the benefits of AVs due to the higher speed and throughput on these lanes. AV dedicated lanes can either be added to the roadway or be implemented by preventing HDVs from using one of the existing lanes and dedicating those lanes to AVs. The former is a costly approach and can be infeasible in some urban areas. The latter requires less infrastructure adaptation but may increase congestion and cause shockwaves near the entry and exit areas of these lanes, which may deteriorate the overall throughput (3).

Several studies in the literature have investigated the effects of AV dedicated lanes on improving congestion, throughput, and traffic performance at the corridor level (4). These studies have concluded that several factors are effective in the operation of AV dedicated lanes, such as the AV MPR, AV technology specifications, road type, travel demand level. Investigating the impacts of dedicated lanes on traffic performance at the network level is essential to making better plans for the implementation of these lanes. Notably, Ivanchev et al. (5) explored the impacts of dedicated lanes for AVs on the network by using an agent-based macroscopic simulation. They concluded that AV dedicated lanes do not improve the average travel time of vehicles at all AV MPRs, since the HDVs would lose their access to a large proportion of the network due to the implementation of AV dedicated lanes on all of the highway segments. This calls for identifying optimal locations for implementing these dedicated lanes to maximize their benefits regarding the travel time of the users. Studies have shown that the locations of dedicated lanes significantly affect the performance of the network (6). However, these studies do not investigate the impacts of AV dedicated lanes for large-scale realistic networks and mostly ignore the infrastructure cost of AV dedicated lanes to economically evaluate their deployment. Also most of these studies have used static assignment model.

To this end, the present study aims to observe the impacts of AV dedicated lanes on the traffic performance of large-scale realistic networks. Note that this study considers level 5 AVs based on the SAE J3016 definitions for the levels of driving automation. In this regard, various factors need to be explored, including changes in the aggregate flow-density relationships, throughput, and the average travel times of AVs and HDVs over the entire network and on the freeway links with AV dedicated lanes, and the location of the dedicated lanes. This study also considers different travel demand levels and AV MPRs in the network and deployment scenarios of AV dedicated lanes, including adding a dedicated lane to the roadway, converting an existing lane to an AV dedicated lane, and setting a higher speed limit on the AV dedicated lanes. To do so, an updated DYNASMART-P, as a mesoscopic traffic simulation tool, is used, which considers the interactions between different vehicle types (i.e., HDVs and AVs) considering various car-following behaviors for these vehicles. Finally, the benefits in terms of travel time savings and the implementation cost of dedicated AV lanes in each scenario are considered to estimate the economic benefit of AV dedicated lanes deployment under various scenarios with different demand levels and AV MPRs. The main contributions of this study are as follows.

- Providing a framework to observe the impacts of AV dedicated lanes on the traffic performance of large-scale networks
- Investigating the impacts of different travel demand levels, MPRs of AVs, and the locations of dedicated lanes on the efficiency of AV dedicated lanes at the network level, in addition to the corridor level
- Incorporating the interactions between AVs and HDVs as well as the changes in the route choice of different vehicles after implementing AV dedicated lanes in a large-scale network

- Performing network-level cost-benefit analyses of deploying AV dedicated lanes under various implementation scenarios

In the remainder, first, the problem statement is provided. Then, the simulation framework of the study is explained. Finally, the case study plan of this research is presented.

## PROBLEM STATEMENT

Investigating the impacts of AV dedicated lanes on traffic congestion and optimizing the number and location of dedicated lanes considering only a specific segment or corridor have multiple caveats. First, the positive/negative impacts of AV dedicated lanes on traffic of the segment can propagate to other links in the network and change the system-level traffic distribution. Second, dedicating an existing lane to AVs or adding a new lane for these vehicles changes the attracted demand to the dedicated lanes and their adjacent lanes. As a result of potential changes in the users' route choice (both AVs and HDVs users), due to the implementation of AV dedicated lanes, the density of AVs and HDVs on all network links is subject to change. Thus, such an adjustment at a corridor-level is likely to change the distribution of AVs and HDVs over the entire network. Therefore, it is important to analyze the impacts of AV dedicated lanes, either by dedicating an existing lane to AVs or adding a new lane, at the network level in addition to the corridor level. This study focuses on showing the importance of conducting cost-benefit analyses of such lanes at the network level.

Thus, this study considers a large-scale general network, $G(N, A)$, where $N$ is the number of nodes and $A$ is the number of links. A subset of links that are freeway, $F \subseteq A$, is considered as candidate links for implementing dedicated lanes. Two scenarios are considered for AV dedicated lane implementation: dedicating an existing lane on freeway links to AVs and adding a new lane to freeway links as AV dedicated lanes. The costs associated with adding a new lane and the technology cost of building an AV lane are considered in the cost-benefit analysis of these scenarios.

Besides, in this study, AVs are assumed to have access to the V2V and V2I information, so they are expected to have en-route information, that enables them to change their route in the middle of their trips once they find a route with a lower travel time than their initially selected routes. Therefore, this study presents a mesoscopic simulation-based analysis to explore the impacts of AV dedicated lanes on traffic flow at the network level, considering the differences in the driving behavior of AVs and HDVs, including their route choice behavior. Also, the effects of different travel demand levels as well as locations of dedicated lanes in the network on the performance of these lanes are investigated. The current study also considers a non-uniform distribution of AVs over large-scale networks.

## SIMULATION FRAMEWORK

A recently updated version of DYNASMART-P is used in this study to simulate the mixed traffic of AVs and HDVs on large-scale networks (7). This traffic simulator accounts for the presence of HDVs and AVs at the network level by considering adaptive fundamental diagrams due to the non-uniform distribution of HDVs with heterogeneous drivers and AVs over the network. To consider the heterogeneity of HDV drivers and the difference in the driving behavior of HDVs and AVs, different microsimulation frameworks are considered for different vehicle types. These microsimulation models are then used to develop spacing-speed relationships for different vehicle types. The spacing-speed relationship of HDVs is then calibrated using real vehicle trajectories from NGSIM data. This tool also incorporates the capacity variations at intersections due to the presence of AVs. Another feature of this simulation tool is to provide en-route information to the AV users and a portion of HDVs. In this regard, users switch their route based on boundedly rational rule (8). Please refer to (9) for more information on the recently developed simulation tool.

The stochastic acceleration model of Hamdar et al. (10) is used in this study as the microscopic model of HDVs. For the acceleration behavior of AVs, the microscopic model of Talebpour and Mahmassani (11) is utilized. The relationships are then used in mesoscopic simulation tool to update the speed values for each link and time interval given the proportion of heterogeneous drivers on the link. In the mesoscopic traffic simulation tool, the proportion of each vehicle type and HDV driver class on each link is tracked in the traffic propagation process at each time interval. Using this proportion, a non-linear equation is solved using a golden section method to obtain the current speed of the link, satisfying the spacing values of all vehicles traversing it.

The NGSIM data represents traffic movements along a highway. To update the traffic flow models of arterials, it is assumed that AVs have similar impacts on the traffic flow of freeways and arterials. Speed values

relative to different density levels are then estimated when AVs occupy a segment with a unit length. Adjustment arterial factors for each density value are estimated by dividing these speed values by the ones when the segment is occupied only by HDVs. The speed values for arterial links are updated at each time interval in the mesoscopic simulation via multiplying a weighted average of adjustment factors associated with each vehicle/driver type by estimated speed from the base traffic flow models for arterials. Finally, adjustment factors are used to increase or decrease the capacity of movements for each intersection at each time interval, once each vehicle type with heterogeneous drivers passes the intersection. The simulation tool is further adapted in this study to model dedicated lanes for AVs. In this regard, different reaction times are considered for AVs on dedicated lanes and regular lanes. In addition to modifying the network structure to include dedicated lanes (either new ones or switching existing lanes), the simulation tool is modified to prevent HDVs from entering the dedicated lanes. Figure 1 briefly illustrates the framework to simulate AVs and HDVs over a large-scale network.

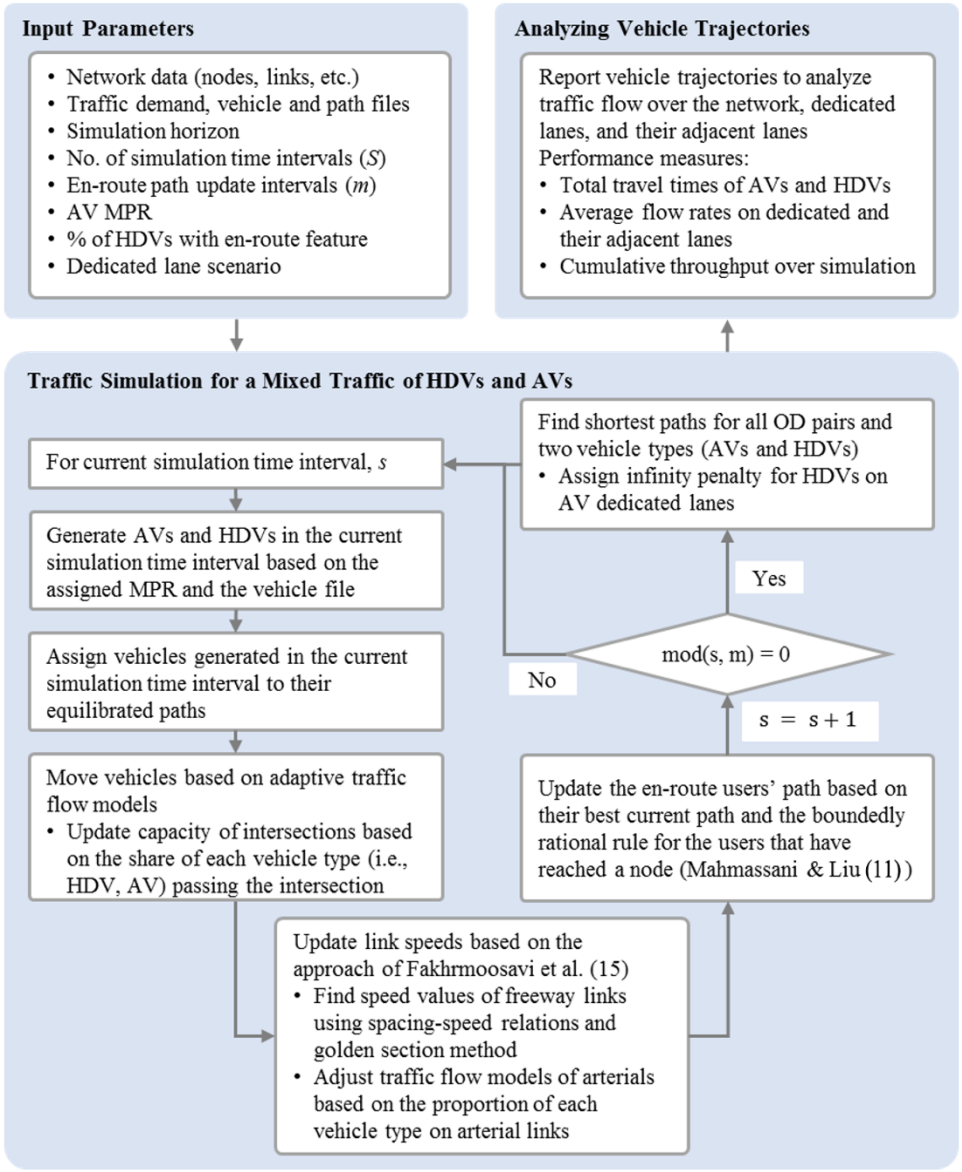



**FIGURE 1 Simulation framework for mixed traffic of AVs and HDVs considering AV dedicated lanes**

## CASE STUDY

## Network Specifications

To evaluate the impacts of AV dedicated lanes on traffic flow at the corridor and network levels, this paper considers the city of Chicago network as the case study. This network contains 1578 nodes, 4805 links, and 218 traffic analysis zones. The network includes 150 freeway links, which are considered candidates for implementing AV dedicated lanes.

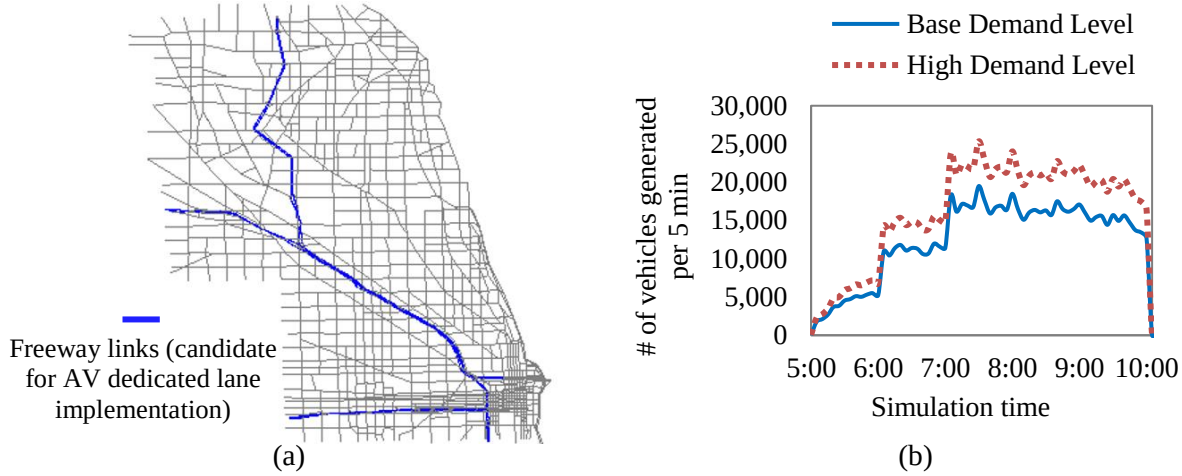



**FIGURE 2 (a) Schematic illustration of the Chicago downtown network; (b) Base and high demand levels for AM peak period in the city of Chicago network**

## Dedicated Lane Scenarios

Four scenarios are investigated in terms of dedicated lanes, which are listed below.

- *Without dedicated lane:* This scenario is the base scenario in which no AV dedicated lane is implemented
- *Making an existing lane dedicated to AVs:* This scenario dedicates one existing lane of each freeway link to AVs. In this scenario, the reaction time of AVs on dedicated lanes is assumed to be half of the one on the regular lanes.
- *AV dedicated speed lane on an existing lane:* In this scenario, one existing lane of freeway links is reserved for AVs. The reaction time of AVs is also similar to the second scenario. The speed limits of the AV dedicated lanes are increased by 10 mph, as AVs can operate efficiently on these lanes.
- *Adding one lane as AV dedicated lane:* This scenario adds a new lane to the existing lanes of freeway links as AV dedicated lanes. The reaction time and speed limit of this scenario are similar to the second scenario.

Different scenarios for locations of AV dedicated lane implementation are also investigated.

## Demand Scenarios and Simulation Settings

The traffic is simulated over AM peak period from 5:00 AM to 10:00 AM using a time-dependent origin-destination demand table. An extra two hours of simulation is considered to unload the network. Two demand levels are investigated in this paper: The "base demand level", for which the number of vehicles simulated in the network during the AM peak period is about 760,000. The second scenario is the "high demand" scenario, in which the origin-destination demand values of the first scenario are increased by 30%. This scenario is considered as AVs are expected to increase vehicle-miles-traveled by providing the driving option for non-drivers and inducing some extra trips in the system due to reduced travel time. The demand profiles for base demand level and high demand level scenarios are illustrated in Figure 2b.

Different MPRs of AVs are considered in this paper, and traffic is simulated with a non-uniform distribution of AVs and HDVs, with 10 different driver classes, over the network. The reaction time of AVs is assumed to be 1 second on regular lanes. It is found that the reaction time of AVs when moving among AV platoons could be lower, compared to when they are in mixed traffic. Hence, the reaction time of AVs on AV dedicated lanes is assumed to be 0.5 s on AV dedicated lanes. Note that in order to thoroughly account for the uncertainty of this parameter, a sensitivity analysis is conducted for different reaction times of AVs on their dedicated lanes. Also, only 25% of HDVs have access to real-time information through GPS devices. The

number of simulation time intervals ($S$) is considered to be 4200. Shortest paths are recalculated for each 10 simulation time intervals (i.e., $m = 10$). The indifference band and the threshold bound are assumed to be 20% and 1 min for the en-route users. Once vehicles are simulated, vehicle trajectories are reported as the output of the simulation for further analyses.

The shape of NFDs, the average the flow rates on the freeway lanes, the cumulative throughput of the dedicated and non-dedicated lanes, the average travel time of HDVs and AVs over the simulation horizon, and the total benefit/loss of the system are compared between different scenarios developed in this study.


## ACKNOWLEDGEMENT

The authors received no financial support for this research.


## REFERENCES


1. Bansal P, Kockelman KM. Forecasting Americans' long-term adoption of connected and autonomous vehicle technologies. Transp Res Part A Policy Pract. 2017;95:49–63.
2. Kockelman KM, Avery P, Bansal P, Boyles SD, Bujanovic P, Choudhary T, et al. Implications of connected and automated vehicles on the safety and operations of roadway networks: a final report. 2016.
3. Talebpour A, Mahmassani HS, Elfar A. Investigating the effects of reserved lanes for autonomous vehicles on congestion and travel time reliability. Transp Res Rec. 2017;2622(1):1–12.
4. Ye L, Yamamoto T. Impact of dedicated lanes for connected and autonomous vehicle on traffic flow throughput. Phys A Stat Mech its Appl. 2018;512:588–97.
5. Ivanchev J, Knoll A, Zehe D, Nair S, Eckhoff D. Potentials and implications of dedicated highway lanes for autonomous vehicles. arXiv Prepr arXiv170907658. 2017;
6. Chen X, Lin X, He F, Li M. Modeling and control of automated vehicle access on dedicated bus rapid transit lanes. Transp Res Part C Emerg Technol. 2020;120:102795.
7. Mahmassani HS, Saberi M, Zockaie A. Urban network gridlock: Theory, characteristics, and dynamics. Procedia-Social Behav Sci. 2013;80:79–98.
8. Mahmassani HS, Liu Y-H. Models of user pre-trip and en-route switching decisions in response to real-time information. IFAC Proc Vol. 1997;30(8):1363–8.
9. Fakhrmoosavi F, Saedi R, Zockaie A, Talebpour A. Impacts of Connected and Autonomous Vehicles on Traffic Flow with Heterogeneous Drivers Spatially Distributed over Large-Scale Networks. Transp Res Rec. 2020;
10. Hamdar SH, Treiber M, Mahmassani HS. Calibration of a Stochastic Car-Following Model Using Trajectory Data: Exploration and Model Properties. In: Transportation Research Board 88th Annual Meeting. 2009.
11. Talebpour A, Mahmassani HS. Influence of connected and autonomous vehicles on traffic flow stability and throughput. Transp Res Part C Emerg Technol. 2016;71:143–63.

# Session 8

Inter-operator competition in a regulated dockless bikesharing market with price-sensitive travelers
*Hongyu Zheng, Kenan Zhang and Marco Nie*

Scalable Routing of Multiple Vacant Taxis in Large Networks
*Guocheng Jiang, Xinlian Yu and Song Gao*

Equitable Collaborative Routing using Incentive Mechanisms
*Chaojie Wang and Srinivas Peeta*

# Inter-operator competition in a regulated dockless bikesharing market with price-sensitive travelers.

Hongyu Zheng[1], Kenan Zhang[2], and Yu (Marco) Nie [*1]

[1]Department of Civil and Environmental Engineering, Northwestern University
2145 Sheridan Road, Evanston, IL 60208, USA
[2]Automatic Control Laboratory, ETH Zurich, Physikstrasse 3, Zurich, 8006, Switzerland

February 1, 2023

**Extended Abstract**
for presentation at the 9th International Symposium on Dynamic Traffic Assignment

***Keywords:*** Market equilibrium, regulation, dockless bikesharing, social welfare.

## 1 Introduction

The first dockless bikesharing (DLB) system appeared on a Chinese university campus in 2015 (Shen et al., 2018). The system features GPS-enabled bikes that can be be unlocked and locked with smartphones, and thus can be easily located and tracked in real time. Soon after the launch, DLB began to grow at a staggering pace. The fleet of DLB bikes in China increased 10,000 folds in two and half years, from 2000 bikes in 2015 to over 23 million at the end of 2017, and the total DLB ridership reached 70 million per day in 2018 (Ecns.cn, 2017). As a healthy, environment friendly, and affordable mode, bikesharing enriches travel choice (Shaheen et al., 2010), and strengthens the resilience of the transportation system, as evidenced during the COVID-19 pandemic.[1] The sheer scale of DLB adds another decisive benefit: convenience. Indeed, the near ubiquity of shared bikes in many large Chinese cities has made DLB a competitive mode for short commute/school trips and a desirable first/last mile solution to long transit trips (Jiang et al., 2020). However, the explosive growth of DLB in China was not without downsides. Driven by the "winner-take-all" mentality, DLB startups locked themselves in fierce competitions that appeared to prioritize market share over everything else. Nasty pricing wars and massive oversupply soon ensured. While one is inclined to think such competitions should benefit travelers, if only temporarily, the reality is many were also negatively affected by the diminishing shared

[*]Corresponding author, E-mail: y-nie@northwestern.edu; Phone: 1-847-467-0502.

[1]In the US, for example, although bikesharing lost about 44% of its pre-pandemic ridership, it significantly outperformed transit, which was hit by a 75% loss; see Bureau of Transportation Statistics (2020).

public space—sidewalks, bike paths, and even entrances/exits of transit stations—that are taken away by bikes. Burdened by excessive supply, DLB operators struggled to properly maintain their fleet, leaving many damaged or disabled bikes to litter the streets (Spinney and Lin, 2018). The operating budget also swelled—probably an unpleasant surprise—with the need for rebalancing, i.e., regularly moving bikes to places where the demand is high. As a result, most operators caught in the DLB gold rush quickly failed. Even ofo, once an icon of the industry, did not survive. Another giant Mobike was acquired by Meituan in 2018 (Kubota, 2018).

The above episode indicates more is not always better when it comes to DLB systems. Moreover, unregulated competition may ruin, rather than strengthen, a DLB market. Inspired by these observations, Zheng et al. (2021) analyzed fleet sizing and pricing decisions by a monopoly operator who must compete with other modes, including walk and a generic motorized mode. Using a parsimonious model calibrated from emprical data collected in Chengdu, China, they showed that the level of service of a DLB system is subject to rapidly diminishing returns to the investment on the fleet. As a result, they observed the fleet cap currently implemented in Chengdu is still three times the fleet size needed for social optimum under monopoly. The analysis by Zheng et al. (2021), however, left out inter-operator competition, a key ingredient in the events that led to the massive correction to the DLB industry in China. Specifically, the fact that their model predicts a much smaller optimal fleet size may be an artifact of the monopoly assumption. It would be interesting to see whether and how their model would replicate the kind of irrational growth manifested between 2015 and 2018. Also, revealing the mechanisms driving the market failures would help the regulators understand how to prevent them in the future.

To the above ends, we build and analyze an oligopoly version of Zheng et al. (2021)'s model, in which the competition between different DLB operators is treated as a Nash-Cournot game with full information. Each DLB operator decides its fleet size and fare schedule according its own objective. The performance of the DLB market, e.g., the level of service, bike ridership, are affected by all operators' decisions. Correspondingly, each operator's market share is determined by its own decisions, as well as those of the competitors. Based on the new model, we explore how an operator may choose the price and fleet size to maximize its profit or ridership when peer competition exists. We use social welfare as a criterion to gauge the effectiveness of three regulation strategies: operator cap, fleet cap, and price cap. The outcome of the oligopoly competition game is a general Nash equilibrium, since both operators' objective and decisions interact with those of their competitors. To solve such game, we first transform it into a set of partial derivative equations, and then solve it using a Bi-level Dual Gradient Descent (AD-BDGD) algorithm. The model will be calibrated against baseline conditions, using a large-scale DLB operational data set collected in Chengdu, China. The calibrated model will then be employed to explain observed market evolution and to compare different regulations.

The next section of this abstract sketches the modeling framework, followed by a brief discussion of anticipated results.

## 2 Model

Suppose there are $I$ DLB operators (denoted as set $\mathcal{O} = \{1, 2, ...I\}$) operating in a city with an area $A$. Each operator $i \in \mathcal{O}$ decides its fleet size $B^i$ and fare rate $f_b^i$. On the demand side, passenger's

travel choice is modeled as follows (see Figure 1). First, passengers choose from three modes in the mobility market: walking, biking, and a composite motorized mode that encompasses taxi, public transit and private auto (see Figure 2). In this step, they minimize the travel cost based on the bike trip fare *averaged across all operators* and the *overall* level of DLB service in the market. The level of DLB service depends on the access time, which is in turn determined by the number of idle shared bike. The reader is referred to Zheng et al. (2021) for the characterization of the relationship between the expected access time and the idle bike density.

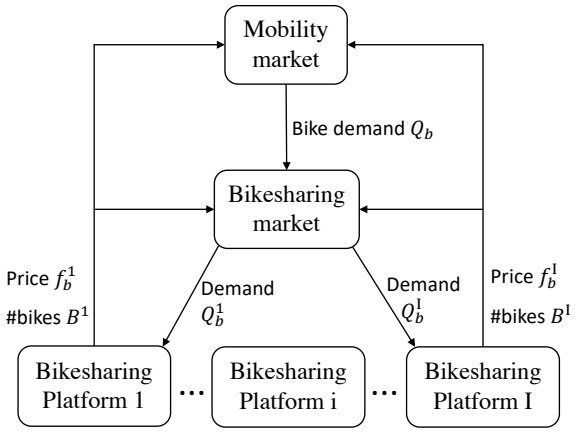



Figure 1: Relationship between shareholders in a mobility market with $I$ bikesharing operators.

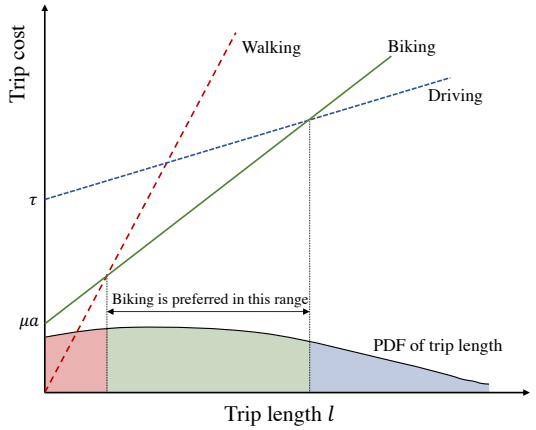



Figure 2: Illustration of mode split in the DLB market.

In the next step, the total DLB demand for bikesharing, $Q_b$, is distributed to each operator. We use $i$ and $-i \in \mathcal{O}$ to denote an operator $i$ and all of its competitors, respectively. The share of platform $i$ in the total bikesharing demand depends on its own fleet size, as well as those of its competitors. More specifically, Operator $i$'s bikesharing demand is calculated as follows:

$$Q_b^i = \frac{n^i}{\sum_{j\in\mathcal{O}} n^j} Q_b - k \sum_{j\in -i} (f_b^i - f_b^j), \tag{1}$$

where $k$ is coefficient with respect to the difference of fare rate between Operator $i$ and $-i$. The first term $\frac{n^i}{\sum_{i\in\mathcal{O}} n^i} Q_b$ is Operator $i$'s bikesharing demand if all operators in the market have the same fare rate and DLB users randomly choose shared bikes to ride. The second term $-k(f_b^i - f_b^{-i})$ reflects that Operator $i$'s market share is also influenced by the difference in fare between itself and its components. Under this setting, a lower price in Operator $i$ will attract more users, while a higher price leads to a smaller market share. We employ a linear price function in Eq. (1) because (i) it enhance the model's tractability; (ii) it easily guarantees the conservation of bikesharing demand regardless of the cross-operator shift, i.e., $\sum_{j\in\mathcal{O}} Q_b^j = Q_b$; and (iii) it has been widely used to capture price elasticity under competition in transportation systems (Charnes et al., 1972; Hurley and Petersen, 1994; Kurata et al., 2007; Zhou and Lee, 2009; Ülkü and Bookbinder, 2012; Zheng et al., 2017; Choi et al., 2020).

Following Zheng et al. (2021), we assume each operator regularly rebalance its bike fleet to maintain the level of service. Accordingly, the conservation of total bike time always holds, i.e., the total bike operating time should equal to the summation of three components: (i) the total

parking time of idle bikes; (ii) the total trip duration of occupied bikes; and (iii) the rebalancing time. Details of rebalancing and each operator's objective functions will be provided in the full presentation.

## 3 Anticipated results

Using the model developed above, we will first analyze under simple settings how each operator's strategy influences its ridership, before developing the AD-BDGD algorithm for the general oligopoly competition game. We will then calibrate the model using a real-world DLB trip data set, and solve it numerically to analyze the following counterfactual scenarios: (i) a duopoly in which both operators keep fixed strategy, maximize profit, and maximize ridership without running a deficit, (ii) a duopoly with asymmetric operators, (iii) an oligopoly. Finally, different regulations will be compared in each of the above scenarios. Detailed results will be provided in the full presentation.

## Acknowledgment

This research is funded by the U.S. National Science Foundation (award # CMMI 2127678).

## References

Bureau of Transportation Statistics. Bikeshare Ridership Down 44% During COVID-19, 2020. URL https://www.bts.gov/newsroom/bikeshare-ridership-down-44-during-covid-19.

A. Charnes, W. W. Cooper, M. J. L. Kirby, and W. Raike. Regulatory Models for Pricing and Evaluation of Transport Services. *Transportation Science*, 6(1):15–31, February 1972. ISSN 0041-1655, 1526-5447. doi: 10.1287/trsc.6.1.15.

Tsan-Ming Choi, Sai-Ho Chung, and Xiaopo Zhuo. Pricing with risk sensitive competing container shipping lines: Will risk seeking do more good than harm? *Transportation Research Part B: Methodological*, 133:210–229, March 2020. ISSN 01912615. doi: 10.1016/j.trb.2020.01.003.

Ecns.cn. Beijing gears up technical, parking rules for shared bikes, 2017. URL https://www.chinadaily.com.cn/china/2017-09/20/content{_}32238345.htm.

W. J. Hurley and E. R. Petersen. Nonlinear Tariffs and Freight Network Equilibrium. *Transportation Science*, 28(3): 236–245, August 1994. ISSN 0041-1655, 1526-5447. doi: 10.1287/trsc.28.3.236.

Hui Jiang, Su Song, Xuan Zou, and Lu Lu. How Dockless Bike Sharing Changes Lives: An Analysis of Chinese Cities, 2020. URL https://www.wri.org.cn/sites/default/files/DBS-EN{_}0.pdf.

Yoko Kubota. Bike-Share Startup in China Veers Off Course - WSJ, 2018. URL https://www.wsj.com/articles/bike-share-startup-in-china-veers-off-course-11545400157.

Hisashi Kurata, Dong-Qing Yao, and John J. Liu. Pricing policies under direct vs. indirect channel competition and national vs. store brand competition. *European Journal of Operational Research*, 180(1):262–281, July 2007. ISSN 03772217. doi: 10.1016/j.ejor.2006.04.002.

Susan Shaheen, Stacey Guzman, and Hua Zhang. Bikesharing in Europe, the Americas, and Asia. *Transportation Research Record*, (2143):159–167, 2010. ISSN 03611981. doi: 10.3141/2143-20.

Yu Shen, Xiaohu Zhang, and Jinhua Zhao. Understanding the usage of dockless bike sharing in Singapore. *International Journal of Sustainable Transportation*, 12(9):686–700, 2018. ISSN 15568334. doi: 10.1080/15568318.2018.1429696. URL `https://doi.org/10.1080/15568318.2018.1429696`.

Justin Spinney and Wen-I Lin. Are you being shared? Mobility, data and social relations in Shanghai's Public Bike Sharing 2.0 sector. *Applied Mobilities*, 3(1):66–83, 2018. ISSN 2380-0127. doi: 10.1080/23800127.2018.1437656. URL `http://doi.org/10.1080/23800127.2018.1437656`.

M. Ali Ülkü and James H. Bookbinder. Optimal quoting of delivery time by a third party logistics provider: The impact of shipment consolidation and temporal pricing schemes. *European Journal of Operational Research*, 221(1): 110–117, August 2012. ISSN 03772217. doi: 10.1016/j.ejor.2012.03.021.

Hongyu Zheng, Kenan Zhang, Marco Nie, Pengyu Yan, and Yuan Qu. How Many Are Too Many? Analyzing Dockless Bikesharing Systems with a Parsimonious Model. *SSRN Electronic Journal*, 2021. doi: http://dx.doi.org/10.2139/ssrn.3919914.

Wei Zheng, Bo Li, and Dong-Ping Song. Effects of risk-aversion on competing shipping lines' pricing strategies with uncertain demands. *Transportation Research Part B: Methodological*, 104:337–356, October 2017. ISSN 01912615. doi: 10.1016/j.trb.2017.08.004.

Wei-Hua Zhou and Chung-Yee Lee. Pricing and competition in a transportation market with empty equipment repositioning. *Transportation Research Part B: Methodological*, 43(6):677–691, July 2009. ISSN 01912615. doi: 10.1016/j.trb.2008.12.001.

# Scalable Routing of Multiple Vacant Taxis in Large Networks

Guocheng Jiang
Department of Civil and Environmental Engineering
University of Massachusetts Amherst, USA
E-mail: guochengjian@umass.edu

Xinlian Yu
School of Transportation
Southeast University, China
Email: xinlianyu@seu.edu.cn

Song Gao
Department of Civil and Environmental Engineering
University of Massachusetts Amherst, USA
E-mail: sgao@umass.edu

January 6, 2023

# 1 Introduction

Ride-sourcing and traditional taxis provide on-demand, point-to-point mobility services in cities. The problem of routing a single vacant taxi or multiple vacant taxis aims to provide optimal decisions regarding their spatial movements, and is crucial in improving operational performance of the ride-sourcing market as well as the benefits to the society.

This study focuses on the optimal routing of multiple vacant taxis. Each vacant taxi needs to decide where to move to after dropping off the previous passenger to maximize profit. General principles apply to both single and multiple vacant taxi routing, including moving to areas with high demand with long prospective trip length. The multi-taxi problem, however, needs to consider competition among taxis, not unlike the congestion effect in a traffic assignment problem. The studies in the literature on multiple vacant taxi routing problem all use a Markov Decision Process (MDP) approach, where states are locations of taxis, and actions are movement to adjacent spatial units (zone or node). Almost all studies of multiple taxi routing use a data-driven reinforcement learning (RL) approach, with the exception of preliminary work by the co-authors Yu et al. (2021). In all cases the combinatorial nature of the state and action space due to multiple agents is dealt with by 1) ignoring vehicle interactions (Oda and Joe-Wong, 2018), 2) using a representative learning agent with some measures to approximate vehicle interactions such as mean-field (Shou and Di, 2020; Ren et al., 2020), context-aware adjustments (Lin et al., 2018; Zhang et al., 2020; Chaudhari et al., 2020; Jiao et al., 2021) and analytical competition function (Yu et al., 2021), or 3) working with multiple taxis sequentially (Liu et al., 2022).

In this study, an analytical framework is developed with an emphasis on the scalability of the method, which is achieved in two ways: 1) the analytical formulation that avoids the need for expensive simulation in data-driven RL approaches; and 2) implementation techniques that make best use of parallelism. Case studies in both a simplified network and real large network are conducted.

# 2 Problem Formulation

A fleet of $M$ taxis travel in a network $G = (N, A)$. $N$ is the set of nodes and $A$ the set of links. There is at most one directional link, $a$, from the source node $i$ to sink node $j$. A vacant taxi searches for passengers during a planning horizon. Vacant taxis' routing decisions are modeled as an infinite-horizon Markov Decision Process (MDP).

**State** The state of a taxi, $s$, is described by node $i \in N$. All taxis are assumed to follow a uniform, probabilistic policy so that the state $s$ is defined just for one taxi at node $i$. A probabilistic policy avoids crowding and allows for optimization over the spreading of taxis at each node.

**Action** The action set for state $s = i$ is $A(s)$, the set of downstream links of $i$. A parameterized probabilistic policy, $\pi_\vartheta$, defines $\pi(a|s, \vartheta)$, the probability that action $a = (i, j)$ is chosen given parameters valued at $\vartheta = \{\theta^1, \theta^2, \ldots, \theta^K\}$ with a feature vector of size $K$. An exponential softmax distribution is adopted such that

$$\pi(a|s;\vartheta) = \frac{\exp\left(\sum_{k=1}^{K} \theta^k x^k_{(a|s)}\right)}{\sum_{a' \in A(s)} \exp\left(\sum_{k=1}^{K} \theta^k x^k_{(a'|s)}\right)}, \tag{1}$$

where $x^k_{(a|s)}$ is the $k^{th}$ feature of action $a$ out of state $s$. The linear combination of features is chosen for its interpretability Murdoch et al. (2019).

**State Transition Probabilities** Passengers are assumed to arrive at node $h$ following a temporal Poisson process with a rate of $\lambda_h$ per hour. The nearest vacant taxi to a passenger gets matched to that passenger. Vacant taxis around node $h$ follow a two-dimensional spatial Poisson distribution with density $\gamma_h$. The set of potential pick-up nodes is assumed to be within a certain distance to the node that the vacant taxi is traveling to, node $j$. Let $N(j)$ denotes such a set. The combined process over all nodes in $N(j)$ is also a Poisson process with an arrival rate $\lambda_{N(j)} = \sum_{j \in N(j)} \lambda_j$.

The vacant taxi gets matched with a passenger at node $h$ when 1) a passenger arrives at a node in $N(j)$ during the traversal time $\tau_a$, and 2) the arrival at node $h$ is earlier than arrivals at all other nodes in the candidate set $N(j)$, for which the taxi in question is the nearest. The probability of the second event can be approximated by the product of probabilities of two decoupled events: 2a) the arrival at node $h$ is earlier than arrivals at all other nodes in the candidate

set $N(j)$, and 2b) the taxi in question is the nearest vacant taxi to node $h$. The adequacy of the approximation should be checked *ex post*, and adjustments should be made if needed.

Therefore, $p_{a,h}$, the probability that the taxi is matched with a passenger at node $h$ over the traversal of link $a = (i, j)$ is

$$p_{a,h} = \frac{\lambda_h}{\lambda_{N(j)}} \left(1 - \exp(-\lambda_{N(j)}\tau_a)\right) \exp(-2\gamma_h \mathcal{L}^2_{h\to a}), \tag{2}$$

where $\mathcal{L}_{h\to a}$ is the right-angle distance from node $h$ to link $a$, approximated by the distance to the middle point of link $a$.

The probability of a passenger picked up at node $h$ having node $k$ as the destination, $p_{h\to k}$, is approximated by the observed fraction of passengers picked up at node $h$ going to $k$.

For a given node $i$ and action $a \in A(i)$ with sink node $j$, the transition probability to $i'$ is defined as follows:

$$p(i'|i, a) = \begin{cases} 1 - \sum_{h\in N(j)} p_{a,h} + \sum_{h\in N(j)} p_{a,h} p_{h\to j}, & \text{if } i' = j \\ \sum_{h\in N(j)} p_{a,h}\, p_{h\to i'}, & \text{if } i' \neq j \end{cases}, \tag{3}$$

In the first case when the next state is the sink node $j$, two scenarios could happen: 1) no passenger is matched during the traversal of link $a$, and 2) the taxi is matched with a passenger at node $h$ whose destination happens to be $j$. In the second case when the next state is not the sink node $j$, the taxi is matched with a passenger at node $h$ whose destination is not node $j$.

**Steady-State Distribution** Let $\mu(s; \vartheta)$ be the stationary distribution of the Markov chain under policy $\pi(\vartheta)$. The system of equations to solve for $\mu(s; \vartheta)$ is

$$\mu(s; \vartheta) = \sum_{s'} \mu(s'; \vartheta) \sum_{a\in A(s')} \pi(a|s', \vartheta) p(s|s', a; \vartheta), \forall s, \tag{4}$$

where the transition probability, $p(s|s', a; \vartheta)$, is explicitly written as a function of $\vartheta$ due to its dependence on the vacant taxi densities and thus the policy parameters that regulate how to spread taxis.

**Endogenous Vacant Taxi Density** At a potential passenger location node $h$, the vacant taxi density, $\gamma_h(\vartheta)$, is the sum of the within-fleet vacant taxi density, $M\Gamma(\vartheta)\mu(h; \vartheta)\beta_h$, and background vacant taxi density, $\omega_h$, that is,

$$\gamma_h(\vartheta) = M\Gamma(\vartheta)\mu(h; \vartheta)\beta_h + \omega_h. \tag{5}$$

$M$ is the fleet size, and $\Gamma(\vartheta)$ the within-fleet vacancy rate, further developed in Eq. (6) below. $\beta_h$ is the average number of network nodes per area unit around $h$ to obtain the number of within-fleet vacant taxis per area unit around $h$. The background vacant taxi density, $\omega_h$, is assumed in this study to be exogenous to focus on a single fleet's decision making.

The within-fleet vacancy rate is represented by a proxy quantity, the expected proportion of time a taxi is vacant per transition, that is,

$$\Gamma(\vartheta) = \sum_{s} \mu(s; \vartheta) \sum_{a=(i,j)\in A(s)} \pi(a|s; \vartheta) \frac{\tau_a}{\tau_a + \sum_{h\in N} p_{ah}(\vartheta) \sum_{i'\in N} p_{h\to i'} (\mathcal{T}_{j\to h} + \mathcal{T}_{h\to i'})}, \tag{6}$$

where $\mathcal{T}_{\bullet\to\bullet}$ is the shortest path travel time between two nodes.

**Reward** The reward from the chosen action is 1) the operating cost of traversing link $a$ if the taxi is not matched with a passenger, or 2) the taxi fare from the matched trip minus the associated operating cost if the taxi is matched with a passenger.

**Optimal Policy** The state does not include a time stamp, and taxis continue searching for the next passenger after dropping off the previous one. Therefore, the process is an infinite-horizon MDP. To resolve the infinite-return issue, we maximize the average reward per vehicle per time unit, $r_\pi(\vartheta)$. Under the assumption of a steady-state distribution with a given policy $\pi(\vartheta)$, the average reward $r_\pi(\vartheta)$ does not depend on the starting state (Sutton and Barto, 2017).

Solving the system of equations (4) and (5) utilizing definitions in (2), (3) and (6), gives the stationary distribution $\mu(s;\vartheta), \forall s$ and state transition probability $p(s'|s,a;\vartheta), \forall s, s', a$. The average reward per time unit then can be calculated as:

$$r_\pi(\vartheta) = \frac{\sum_s \mu_\vartheta(s;\vartheta) \sum_{a\in A(s)} \pi(a|s;\vartheta) \sum_{s',r} p(s',r|s,a;\vartheta) r}{\sum_s \mu(s;\vartheta) \sum_{a=(i,j)\in A(s)} \pi(a|s;\vartheta) \Big[\tau_a + \sum_{h\in N} p_{ah}(\vartheta) \sum_{i'\in N} p_{h\to i'} (\mathcal{T}_{j\to h} + \mathcal{T}_{h\to i'})\Big]}, \tag{7}$$

where the enumerator is the average reward per transition (see Sutton and Barto, 2017, Section 10.3), and the denominator the expected duration per transition. $r$ is the immediate reward of the transition from $s$ to $s'$ given action $a$, and can be calculated following Yu et al. (2019).

# 3 Computational Tests

We first conducts systematic investigation of the optimal solutions in hypothetical and simplified network settings, and then applies it to a real case study.

## 3.1 Hypothetical and Simplified Network

In the hypothetical and simplified network, we construct a network with two nodes to allow for the study of the spatial dimension of the multi-taxi problem, shown in Figure 1. Each node represents a zone (the two words will be used interchangeably), for example, a neighborhood in an urban area, a major transportation hub such an airport or train station.. Two actions are available at each node: stay or go (to the other node), representing, e.g., cruising in the current neighborhood or traveling a relatively long distance to the airport.

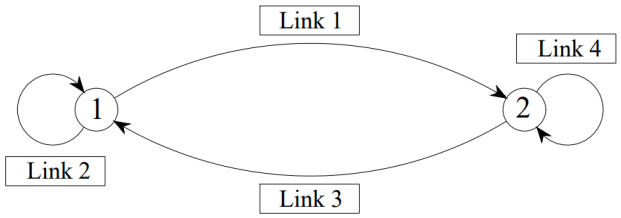



**Fig. 1** : The two-node network

The optimal policies are then compared with three heuristics: 1) Random walk: each action out of a given state has an equal probability of 0.5; 2) 0/1 hotspot: at any state, 100% of taxi goes to the node with higher demand, and 0% to the node with lower demand; break the tie arbitrarily when the two nodes have the same demand; 3) Proportional hotspot: set the policy probability ratio to be the same as the demand ratio.

Sensitivity analysis analyses are carried out with respect to a number of input parameters: demand spatial distribution, competition spatial distribution, and taxi group size.

## 3.2 The Real Case Study

A case study is conducted in the the roadway network of Shanghai, the most populous city in China. We obtain the taxi GPS trajectories data in April, 2015 from one major taxi company with around 25 percent market share. Trajectories from 10,421 taxis in 5:30-11:30 am on a weekday are used to obtain necessary parameters in this study, e.g., passenger arrival rate at each node and passenger destination probability.

The travel time on each link is computed based on the link length and corresponding speed limit. The trip fare is calculated based on the shortest travel distance between origins and destinations. The demand rate of a link is the number of pick-ups per hour that occur along the link. The background vacant taxi density at each node is assumed to be the same as the vacant taxi density from the trajectory data. The optimization is carried out for a fleet with a varying size $M \in \{1, 5000, 10000\}$ to investigate the effect of fleet size on the running time and optimal solution.

The taxi fleet problem is equivalent to the single taxi problem when the size is 1. The matching candidate set of a link consists of the nodes within 1km radius of the midpoint of link.

The probabilistic policy in Eq. (1) is parameterized with two exogenous features of action $a$ out of state $s$ with parameters $\vartheta = \{\theta^1, \theta^2\}$: (1) link demand rate divided by link length; (2) vacant taxi density on link. The baseline model of random walk is thus obtained by setting $\vartheta = 0$.

The BFGS algorithm is used to solve for the optimal policy. Table 1 shows the computational performance of various implementations with various devices, number of cores, and computational methods in terms of the running time per inner iteration. All implementations are tested on a computer cluster at the Massachusetts Green High Performance Computer Center (https://www.mghpcc.org/). The running time reduces significantly after using matrix operation method, parallel computing over multiple nodes and GPU implementation.

| Type | Device | Number of Nodes | Cores per Node | Method | Running Time per Inner Iteration (sec) |
|---|---|---|---|---|---|
| CPU | 2.4GHz Intel Xeon Gold 6148 CPU | 1 | 15 | For-loops | 1934 |
| | 2.4GHz Intel Xeon Gold 6148 CPU | 1 | 15 | Matrix Operation | 132 |
| | 2.4GHz Intel Xeon Gold 6148 CPU | 1 | 15 | Sparse Matrix | 57 |
| | 2.4GHz AMD Opteron Processor 6278 | 1 | 15 | Sparse Matrix | 154 |
| | 2.4GHz Intel Xeon Gold 6148 CPU | 1 | 40 | Sparse Matrix | 31 |
| | 2.4GHz Intel Xeon Gold 6148 CPU | 2 | 40 | Sparse Matrix with MPI | 24 |
| GPU | 4x Tesla K80 devices per node | 1 | 20 | Spare Matrix | 22 |

**Table 1** : Computational Performance of Various Implementations

| Strategy | Fleet Size | $\theta_1$ | $\theta_2$ | Unit profit (CNY/hour) |
|---|---|---|---|---|
| Optimal fleet policy (parameterized) | 1 | 5.96 | -3.51 | 83.1 |
| | 5,000 | 5.25 | -3.72 | 81.1 |
| | 10,000 | 4.92 | -3.80 | 80.2 |
| Random walk | 10,000 | 0 | 0 | 69.4 |
| Single taxi policy (tabular) | 1 | - | - | 84.8 |
| | 5,000 | - | - | 76.4 |
| | 10,000 | - | - | 73.1 |

**Table 2** : Optimal coefficients and average reward of different strategies

Table 2 summarizes the coefficients of optimal policy, random walk policy, and single taxi policy. The single taxi policy is derived based on Yu et al. (2019). It is in a tabular form; it ignores taxi competition and guide all vehicles on the best action. When the fleet size $M = 1$, the single taxi policy outperforms the optimal fleet policy, because it is not parameterized and more flexible. The optimal policy outperforms the single taxi policy when the fleet size = 5000 and the advantage of optimal policy over single taxi policy becomes increasingly larger when the fleet size keep increasing from 5000 to 10000, as the single taxi policy does not account for competition.

# 4 Conclusions

This paper formulates the routing problem of a fleet of vacant taxis as an MDP. A stochastic and uniform policy is adopted to reduce the state space, and a system of equations is built to solve for the endogenous vacant taxi density, state transition probabilities and the resulting average reward. The BFGS algorithm is used to solve for the optimal policy. Sparse matrix operations and GPUs are used to shorten the running time in a very large network.

The optimal policy is compared with the naive routing strategy, random walk. The average reward for the optimal policy is 15.6% higher than that of random walk. The optimal average reward decreases slightly when the fleet size increase, reflecting certain loss of profit caused by the higher competition. In terms of the running time, the algorithm converges faster when the fleet size decreases because the endogenous vacant taxi density does not change much when the fleet size is small. The running time however increases only slightly with much larger fleet size due to the method's analytical nature. The stationary distribution derived from the optimal policy shows that vacant taxis tend to stay around or go to high-demand areas including downtown and major transportation hubs.

# References


Chaudhari, H. A., Byers, J. W. and Terzi, E. (2020). Learn to earn: enabling coordination within a ride-hailing fleet, *IEEE International Conference on Big Data*, Atlanta, GA.

Jiao, Y., Tang, X., Qin, Z., Li, S., Zhang, F., Zhu, H. and Ye, J. (2021). Real-world ride-hailing vehicle repositioning using deep reinforcement learning, *Transportation Research Part C* **130**: 103289.

Lin, K., Zhao, R., Xu, Z. and Zhou, J. (2018). Efficient large-scale fleet management via multi-agent deep reinforcement learning, *KDD'18: The 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, London, UK.

Liu, Z., Li, J. and Wu, K. (2022). Context-aware taxi dispatching at city-scale using deep reinforcement learning, *IEEE Transactions on Intelligent Transportation Systems* **23**(3): 1996–2009.

Murdoch, W. J., Singh, C., kumbier, K., Abbasi-Asi, R. and Yu, B. (2019). Definitions, methods, and applications in interpretable machine learning, *PNAS* **116**(44): 22071–22080.

Oda, T. and Joe-Wong, C. (2018). MOVI: A model-free approach to dynamic fleet management, *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, Honolulu, HI.

Ren, S., Li, Q., Zhang, L., Qin, Z. and Yang, B. (2020). Optimizing stochastic routing for taxi fleets with model enhanced reinforcement learning. arXiv:2010.11738.
**URL:** *https://arxiv.org/abs/2010.11738*

Shou, Z. and Di, X. (2020). Reward design for driver repositioning using multi-agent reinforcement learning, *Transportation Research Part C* **119**: 102738.

Sutton, R. S. and Barto, A. G. (2017). *Reinforcement Learning: An Introduction*, 2nd edn, The MIT Press.

Yu, X., Gao, S., Hu, X. and Park, H. (2019). A markov decision process approach to vacant taxi routing with e-hailing, *Transportation Research Part B* **121**: 114–134.

Yu, X., Gao, S. and Khani, A. (2021). Optimal routing of multiple vacant taxis: A policy gradient method with endogenous state transition probabilities, *The 100th Annual Meeting of Transportation Research Board*.

Zhang, W., Wang, Q., Li, J. and Xu, C. (2020). Dynamic fleet management with rewriting deep reinforcement learning, *IEEE Access* **8**: 143333–143341.

# Equitable Collaborative Routing using Incentive Mechanisms

*Chaojie Wang*, cwang717@gatech.edu
School of Civil and Environmental Engineering, Georgia Institute of Technology

*Srinivas Peeta*, peeta@gatech.edu∗
School of Civil and Environmental Engineering, Georgia Institute of Technology
H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology

Keyword: Accessibility Equity, Inclusion Equity, Utility Equity, Incentive Mechanism

## BACKGROUND AND MOTIVATION

Traffic assignment models are essential for traffic planning and network operations. Different traffic assignment models provide/predict different flow patterns based on the origin-destination (O-D) traffic demand in transportation networks, depending on the assumptions regarding system-level objectives and/or individual user behavior/objectives. Also, various behavior interventions, such as tolls and incentive mechanisms, have been proposed to enable collaborative routing and nudge the user equilibrium (UE) flows, which are more indicative of the real-world situation, closer to the system optimal (SO) conditions in order to improve mobility efficiency in transportation systems. Other metrics that are considered extensively in the literature are environmental sustainability and safety. However, under SO, some users from the same origins and heading to the same destinations face much longer trip times than others, highlighting increased equity-related concerns in collaborative routing (Jahn et al., 2005). A few studies have emphasized the significance of mobility equity in traffic assignment by limiting travel cost disparities among users with identical O-D pairs (Jahn et al., 2005; Angelelli et al., 2018, 2021; Jalota et al., 2021). Thereby, the current perspective on equity issues in traffic assignment is largely restricted to mobility equity, which assesses differences in travel cost among travelers with same O-D pairs.

This study proposes that it is more pertinent to consider the following three types of equity in behavior intervention-enabled collaborative routing: (i) accessibility equity, which prioritizes equal access to jobs, services, and education/training opportunities for all travelers; (ii) inclusion equity, which implies that behavioral interventions should not impede users' access to mobility options; and (iii) utility equity, which focuses on reducing disparities in utilities instead of travel cost differences among users. The current literature is lacking in terms of considering equity comprehensively related to accessibility, inclusion, and utility, which is more reflective of the true transportation-related barriers faced by users. Further, existing equity modeling in collaborative routing has the same drawbacks as more general traffic assignment models and pricing strategies, such as the inability to model user heterogeneity and the computational complexity of solution algorithms. Specifically, the computational burden increases not just with the size of the network, but also with the number of users if models are to capture individual characteristics. Consequently, modeling and solution design are subject to restrictive constraints when dealing with computational challenges of this magnitude.

This study proposes an incentive-enabled strategy to promote equity in collaborative routing. Figure 1 presents the conceptual framework of the proposed collaborative routing strategy,

illustrating how it promotes accessibility equity, inclusion equity, and utility equity using individualized incentive mechanisms. The traffic operator and the users both pursue mobility efficiency and equity and consider the effects of incentives, but from different perspectives. The traffic operator focuses on system mobility efficiency, which is quantified in this study as total system travel time, whereas users aim to reduce their individual travel times. Regarding equity, the traffic operator seeks to promote accessibility equity and inclusion equity, while users are primarily concerned with utility equity, as utility theory provides a reasonable basis for their expected behaviors. Providing incentives to users can enhance their benefits but increases the investment costs of the traffic operator. As shown in Figure 1, the objective function of the traffic operator comprises of measurements for system mobility efficiency, incentive costs, and accessibility equity. The users' utility functions consists of their individual mobility efficiency and incentive benefits, while utility equity and inclusion equity entail additional behavior constraints on each user.

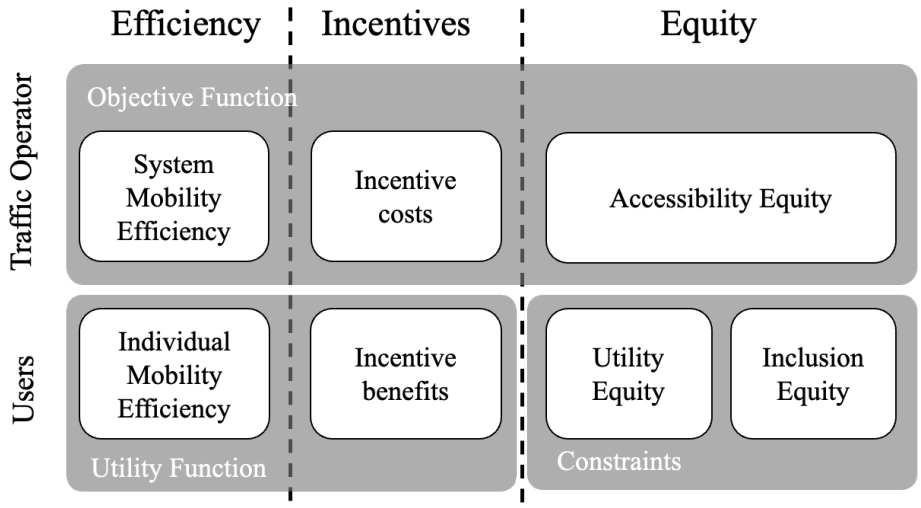



**Figure 1. Conceptual framework of the collaborative routing strategy promoting accessibility equity, inclusion equity, and utility equity**

Inspired by the Gini index and the mean absolute difference, we design a measure of inequity based on the travel times to several types of places (e.g., medical service sites, grocery stores) that entail equitable access. A bi-level optimization model is developed as a benchmark following conventional transportation network design. The upper-level problem optimizes the objective function of the traffic operator, while the lower-level problem models the users' route choice equilibrium given their utility functions. The formulation for network design cannot be solved in a computationally efficient way. Therefore, we propose a utility-and-incentive constrained traffic assignment (BICTA) model, which integrates the utility maximization principle and an envy-free compensation technique as model constraints for traffic assignment. We then show that it is equivalent to the network design formulation. We further create a consensus optimization model (Nedic et al., 2010) to decompose the BICTA model into a series of local optimization problems, that enables the use of a decentralized algorithm based on the Alternating Direction Method of Multipliers (ADMM) method (Boyd et al., 2010). We perform theoretical assessments of model equivalence, equilibrium uniqueness, and algorithm convergence, while also theoretically and quantitatively validating utility equity. Numerical examples illustrate the effectiveness and

performance of our proposed models and strategies, revealing how they enhance the three dimensions of equity and how the hyperparameters of the model impact the tradeoffs between multiple system objectives. In addition, the computational efficiency of the decentralized algorithm is illustrated by comparing it to centralized algorithms.

## REFERENCES


Angelelli, E., Morandi, V., Savelsbergh, M., & Speranza, M. G. (2021). System optimal routing of traffic flows with user constraints using linear programming. *European Journal of Operational Research*, *293*(3), 863–879. https://doi.org/10.1016/J.EJOR.2020.12.043

Angelelli, E., Morandi, V., & Speranza, M. G. (2018). Congestion avoiding heuristic path generation for the proactive route guidance. *Computers & Operations Research*, *99*, 234–248. https://doi.org/10.1016/J.COR.2018.07.009

Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2010). Distributed optimization and statistical learning via the alternating direction method of multipliers. In *Foundations and Trends in Machine Learning* (Vol. 3, Issue 1). https://doi.org/10.1561/2200000016

Jahn, O., Möhring, R. H., Schulz, A. S., & Stier-Moses, N. E. (2005). System-Optimal Routing of Traffic Flows with User Constraints in Networks with Congestion. *Https://Doi.Org/10.1287/Opre.1040.0197*, *53*(4), 600–616. https://doi.org/10.1287/OPRE.1040.0197

Jalota, D., Solovey, K., Tsao, M., Zoepf, S., & Pavone, M. (2021). Balancing Fairness and Efficiency in Traffic Routing via Interpolated Traffic Assignment. *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems, AAMAS*, *2*, 678–686. https://doi.org/10.48550/arxiv.2104.00098

Nedic, A., Ozdaglar, A., & Parrilo, P. A. (2010). Constrained consensus and optimization in multi-agent networks. *IEEE Transactions on Automatic Control*, *55*(4). https://doi.org/10.1109/TAC.2010.2041686

Wang, C., Peeta, S., & Wang, J. (2021). Incentive-based decentralized routing for connected and autonomous vehicles using information propagation. *Transportation Research Part B: Methodological*, *149*, 138–161. https://doi.org/10.1016/J.TRB.2021.05.004

# Session 9

Proactive Defense-based Cyberattack Mitigation for Connected and Autonomous Vehicles using Moving Target Defense
*Yangjiao Chen and Srinivas Peeta*

Dynamic Evacuation Management: Shelter Allocation and Traffic Assignment with Telecommunication Networks
*Hassan Idoudi, Mostafa Ameli, Cyril Nguyen Van Phu, Abderrezak Rachedi and Mahdi Zargayouna*

Improving Air Traffic Flow Management with Macroscopic Traffic Flow Modelling: Model Formulation, Simulation and Comparison
*Christopher Cummings, Vasileios Volakakis and Hani Mahmassani*

# Proactive Defense-based Cyberattack Mitigation for Connected and Autonomous Vehicles using Moving Target Defense

Yangjiao Chen, ychen3254@gatech.edu
School of Civil and Environmental Engineering, Georgia Institute of Technology

Srinivas Peeta, peeta@gatech.edu*
School of Civil and Environmental Engineering, Georgia Institute of Technology
School of Industrial and Systems Engineering, Georgia Institute of Technology

## BACKGROUND AND MOTIVATION

Emerging connectivity and automation technologies enabling connected and autonomous vehicles (CAVs) are expected to revolutionize the transportation system. Vehicle-to-vehicle (V2V)-based communication enables efficient information sharing between vehicles while automation allows vehicles to react and respond instantaneously to the information received. CAVs utilize various proprioceptive and exteroceptive sensors as well as communication networks to collect traffic and other critical information, and further use the information for vehicle decision-making. Time-dependent knowledge on vehicular states and traffic conditions can be disseminated among vehicles to improve vehicle actuation decisions with enhanced situational awareness and help to collaboratively enhance the traffic system performance. Hence, these emerging technologies and rich information highlight the potential to develop safer and more efficient vehicles with new travel behaviors and emergent traffic dynamics. However, while CAVs hold promise for enhanced safety, mobility, and efficiency, cyberattack risks are a key barrier to their large-scale implementation, and can diminish their putative benefits or even deteriorate the traffic performance.

Past studies have highlighted many vulnerabilities in CAVs facing cyberattacks (Eiza and Ni, 2017; Petit and Shladover, 2015). V2V communications enable CAVs to obtain additional information from surrounding vehicles. However, though this information is intended to assist vehicle decision-making, it may not be reliable, or may even be harmful due to cyberattack risks. Attackers can target the vehicle's software, hardware, communication channel or sensors to introduce misleading information into the traffic network to confuse vehicle perception. Many studies have sought to understand the impacts of cyberattacks on transportation systems. For example, Wang et al. (2020) propose a framework to investigate the effects of cyberattacks on vehicle dynamics under cooperative vehicle behavior. However, there are two critical gaps related to understanding and addressing cyberattacks in the transportation context. First, while there are some efforts in the literature related to descriptive models of cyberattack impacts, prescriptive models to mitigate the effects of cyberattacks are sparse and entail significant challenges. Second, even those studies focus on defense measures at the vehicle hardware or software level. However, it is difficult to identify effective defense measures related to traffic performance due to the lack of appropriate models to analyze their impacts on vehicle dynamics.

In general, system security depends on the outcome of competition between the attack side and the defense side. Reactive defense and proactive defense are two common defense strategies. While reactive defense mechanisms, such as various intrusion detection systems, have significantly enabled efforts to secure systems/networks for the last several decades, they have limitations due to their inability to prevent potential attackers in advance as well as the difficulty

of finding an appropriate anomaly threshold capable of detecting various types of attacks. With the proliferation of persistent, advanced, intelligent attacks in the cybersecurity domain (Chen et al., 2014), defenders are often way behind attackers in taking appropriate actions to thwart potential attackers. In this context, the concept of Moving Target Defense (MTD) has emerged as a proactive defense mechanism that aims to prevent attacks and has gained popularity in protecting various cyber-physical systems (CPSs) (Griffioen et al., 2021; Kanellopoulos and Vamvoudakis, 2020). The MTD mechanism utilizes a stochastic switching structure to alter the parameters of the system dynamically and continuously, resulting in periodic changes to the system matrices. The time-varying dynamics act as a moving target, hindering a potential adaptive adversary from performing system identification, which limits the attacker's understanding of the system model and increases the unpredictability of the system behavior for attackers to conduct successful attacks. Hence, incorporating proactive defense-in-design principles into CAV design is promising to ensure the security and robustness of the CAV-based transportation system. V2V communications may provide additional capabilities to address cybersecurity concerns. It is possible to leverage the diversity in information flow topologies to develop proactive defense-based strategies for CAVs to increase system unpredictability, thereby reducing the predictability asymmetry between attackers and defenders. Several studies (Gong et al., 2019; Wang et al., 2019; Zhou et al., 2020) have illustrated how platoon performance can be enhanced by leveraging information flow topology dynamics in CAV design.

Despite the potential to leverage MTD techniques to enable reliability and security of the CAV system, several implementation challenges arise due to problem complexity across multiple dimensions. First, a capability to generalize the effectiveness of the attack mitigation strategy across various attack types is lacking. As a complex CPS, there are many ways to initiate cyberattacks in CAVs to modify the information exchanged between vehicles that is used for vehicular decision-making. The defense mechanisms in most previous studies are attack specific as the attack categorization is limited to the low-level layers of the CAV system, while their linkages to the traffic layer are not considered. However, attacks at the lower layers typically propagate upward to the traffic layer in terms of incorrect perception or information, leading to inaccurate vehicle prediction and motion planning. This motivates the need for generalizable high-level defense strategies that can account for the heterogeneity of lower-layer attacks.

Second, the interactions between the attack and defense mechanisms need to be captured as they influence the information flow topology. However, past studies have focused primarily on how information propagation impacts the traffic and can be leveraged for enhancing performance. Kim and Peeta (2019) investigate how information flow propagation affects traffic flow dynamics, and further leverage V2V information to develop decentralized real-time routing strategies to improve system performance. Wang et al. (2021) develop incentive-based decentralized routing strategies for CAVs using V2V information. CAV technologies necessitate the development of new vehicle dynamics models, which can be leveraged to target micro vehicle-level decisions to enhance system performance. Some studies have proposed CAV longitudinal dynamics models that use V2V-based information to foster string stability (Wang et al., 2019), improve riding comfort ((Wang et al., 2019; Zhou et al., 2020), or enable cooperative behavior (Wang et al., 2020). However, malicious adversaries can implement various types of attacks at different levels of the sensing, communication, and computation layers to modify expected vehicle behaviors that undermine the reliability of these vehicle dynamics models under adversarial conditions. Hence, there is a critical need to factor cyberattack risks and develop CAV designs that proactively mitigate the impacts of cyberattacks.

## MAIN IDEA

This study proposes a proactive defense-based vehicle dynamics model that leverages moving target defense strategies to mitigate the impacts of cyberattacks for the CAV system. It uses active switching strategies to dynamically change communication topologies to increase the unpredictability of CAV behavior, thereby counteracting attackers from achieving their desired goals. With connectivity capabilities, a CAV can receive information from multiple preceding CAVs within communication range. If CAVs adopt a fixed communication scheme, attackers can decipher the cooperative relationship between vehicles easily and influence vehicle behavior by simply modifying some information. To address this issue, we propose to switch the cooperation schemes dynamically to reduce predictability while ensuring the benefits of vehicular cooperation. In critical contrast to current cooperative vehicle dynamics models where the V2V information on surrounding vehicle states and/or traffic information is fully trusted and adopted, the proposed strategy is robust to attacks generating malicious information.

Figure 1 illustrates the conceptual flow chart of the proposed MTD-based defense framework that consists of three interacting components: an information propagation model, a CAV dynamics model, and a switching-based MTD scheme. The CAV model leverages the information from the information propagation model to elicit benefits from cooperation between vehicles. In addition, it integrates a MTD scheme to maximize the unpredictability of CAV behavior for attackers and mitigate the attack effects. The information propagation model, which consists of a dissemination scheme and an update scheme, is developed to track information on the vehicle and traffic states. After the CAV receives this information, the possible vehicle cooperation schemes are considered and adopted sequentially based on the MTD scheme. The MTD scheme consists of an optimization model and a Markov model. The unpredictability maximization based on an information entropy measure generates a probabilistic distribution for the switching-based cooperation schemes. The Markov model describes the switching logic between possible cooperation schemes as time progresses. In contrast to existing models that adopt constant cooperation schemes, the MTD-based defense strategy proactively switches its cooperation schemes which increases unpredictability while ensuring the benefits of cooperation.

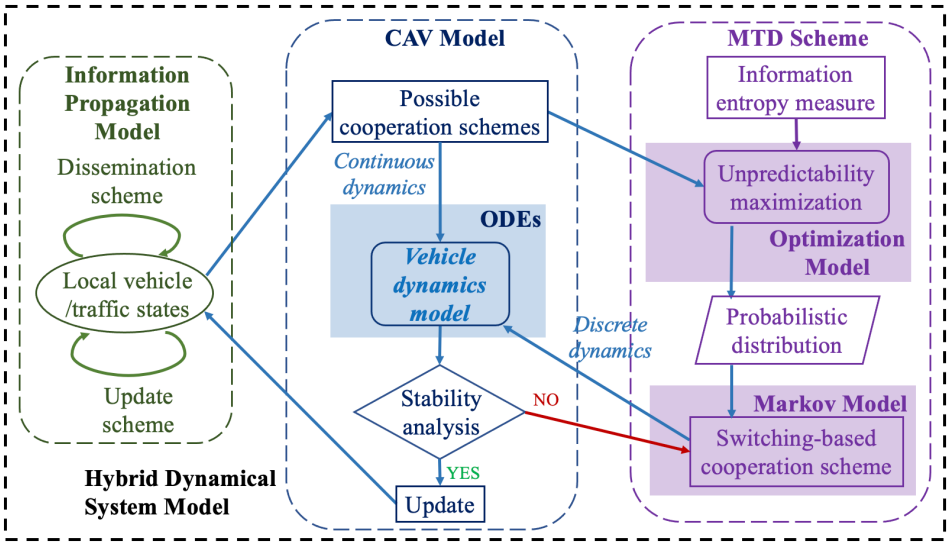



Figure 1. Conceptual flow chart of MTD-based defense framework

Figure 2 further conceptually illustrates the switching-based cooperation scheme in a time interval $T$. Suppose a CAV cooperates with two preceding vehicles, including the immediate predecessor vehicle and another preceding vehicle. For example, if vehicle $i$ can potentially receive information on four preceding vehicles' states, it can result in three cooperation schemes. Then, following the MTD scheme, at time instant $t = 0$, it could utilize the velocity and location information from only preceding vehicles $i-1$ and $i-2$. At the next time instant $t = t_1$, vehicle $i$ may switch to utilize information from vehicles $i-1$ and $i-4$. Vehicle $i$ will keep switching the cooperation scheme until $T$ is reached. The implementation distribution of these three cooperation schemes would follow the probabilistic distribution determined by the unpredictability maximization model. Hence, for example, if the attackers modify the information from vehicle $i-2$, their impacts can be mitigated compared to the no-switching scenarios. Thus, the proposed cooperation strategy leveraging MTD is stochastic, precluding attackers from launching effective strategies to achieve their malicious goals. From a modeling perspective, the vehicle dynamics is described by a set of ordinary differential equations (ODEs) with continuous dynamics. Besides, each switch would form a new set of ODEs and the CAV dynamics is governed by all these sets of ODEs. Hence, the switching of cooperation schemes introduces discrete dynamics to the vehicle dynamics model besides the continuous vehicle dynamics, leading to a hybrid dynamical system for the CAV model. The stability of the hybrid system model needs to be verified before implementing the switching strategy based on MTD. If stability conditions cannot be satisfied, the probability distribution for switching should be adjusted. When stability conditions are satisfied, the outcomes of the cooperation scheme switching in $T$ will update the vehicle and traffic states. The real-time information on vehicle/traffic states is then disseminated in the local area. Due to the communication dynamics, possible vehicle cooperation schemes differ at different times. Thus, they need to be updated repeatedly using the MTD framework. This framework is repeated for the next time interval and so on, until the vehicle reaches its destination. The framework is applied for every CAV vehicle.

To the best of our knowledge, this study is the first to propose security-in-design principles for CAVs, incorporate the consideration of information reliability and security concerns into the CAV dynamics modeling process, and propose an analytical proactive defense model to mitigate cyberattack impacts with theoretical analysis and validations.

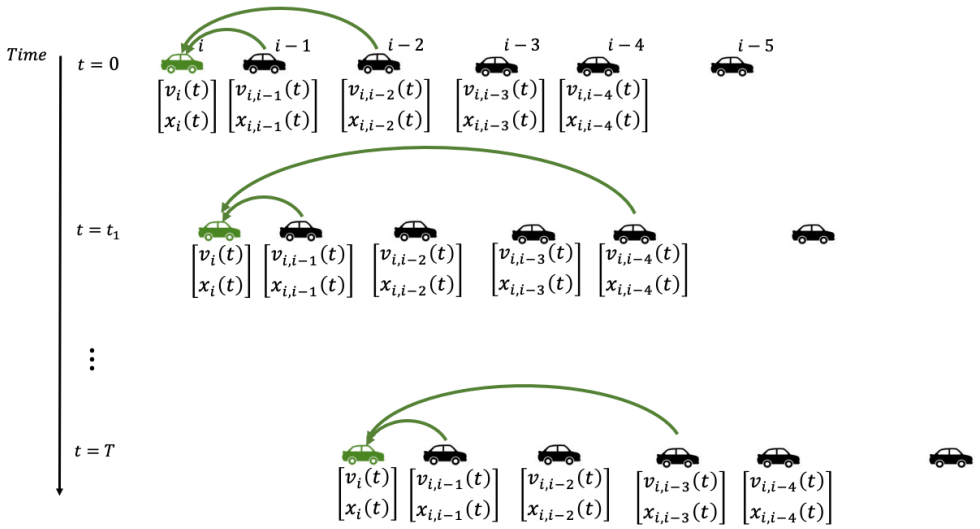



Figure 2. Illustration of the communication topology switching scheme

In the full paper, we will provide details of the entropy-based unpredictability maximization model, the switching-based cooperation schemes, the resulting hybrid dynamical system model properties, and the stability analysis algorithm. The model behavior under different cooperative scheme assumptions for the proposed mitigation strategy will be analyzed. Theoretical analysis will be conducted for sensitivity, stability, and aggregated effects of the mitigation strategy at a system level. We will prove the properties of the hybrid dynamical system with the designed verification algorithm and validate the properties revealed through numerical experiments. Finally, the effectiveness and performance of the proposed model and the mitigation strategies under different attack types will be illustrated through multiple simulation experiments.

## REFERENCES


Chen, P., Desmet, L., Huygens, C., 2014. A Study on Advanced Persistent Threats, in: De Decker, B., Zúquete, A. (Eds.), Communications and Multimedia Security, Lecture Notes in Computer Science. Springer, Berlin, Heidelberg, pp. 63-72.

Eiza, M.H., Ni, Q., 2017. Driving with Sharks: Rethinking Connected Vehicles with Vehicle Cybersecurity. IEEE Vehicular Technology Magazine 12, 45-51.

Gong, S., Zhou, A., Peeta, S., 2019. Cooperative Adaptive Cruise Control for a Platoon of Connected and Autonomous Vehicles considering Dynamic Information Flow Topology. Transportation Research Record 2673, 185-198.

Griffioen, P., Weerakkody, S., Sinopoli, B., 2021. A Moving Target Defense for Securing Cyber-Physical Systems. IEEE Transactions on Automatic Control 66, 2016-2031.

Kanellopoulos, A., Vamvoudakis, K.G., 2020. A Moving Target Defense Control Framework for Cyber-Physical Systems. IEEE Transactions on Automatic Control 65, 1029-1043.

Kim, Y-H. and Peeta, S. 2019. Evaluation of decentralized vehicle-to-vehicle communications based advanced traveler information system, Under Review.

Petit, J., Shladover, S.E., 2015. Potential Cyberattacks on Automated Vehicles. IEEE Transactions on Intelligent Transportation Systems 16, 546-556.

Wang, C., Gong, S., Zhou, A., Li, T., Peeta, S., 2020. Cooperative Adaptive Cruise Control for Connected Autonomous Vehicles by Factoring Communication-Related Constraints. Transportation Research Part C: Emerging Technologies 113, 124-145.

Wang, C., Peeta, S., Wang, J., 2021. Incentive-based decentralized routing for connected and autonomous vehicles using information propagation. Transportation Research Part B: Methodological 149, 138-161.

Wang, P., Wu, X., He, X., 2020. Modeling and analyzing cyberattack effects on connected automated vehicular platoons. Transportation Research Part C: Emerging Technologies 115, 102625.

Zhou, A., Gong, S., Wang, C., Peeta, S., 2020. Smooth-Switching Control-Based Cooperative Adaptive Cruise Control by Considering Dynamic Information Flow Topology. Transportation Research Record 2674, 444-458.

**Extended ABSTRACT for 9th International Symposium on Dynamic Traffic Assignment (DTA2023)**

# DYNAMIC EVACUATION MANAGEMENT: SHELTER ALLOCATION AND TRAFFIC ASSIGNMENT WITH TELECOMMUNICATION NETWORKS

Hassan Idoudi[a, b], Mostafa Ameli[a, *], Cyril Nguyen Van Phu[a], Abderrezak Rachedi [b] and Mahdi Zargayouna[a]
[a] *University Gustave Eiffel, COSYS, GRETTIA, Paris, France*
[b] *University Gustave Eiffel, LIGM, Paris, France*
[*] *Corresponding author, Email: mostafa.ameli@univ-eiffel.fr*

**Keywords:** Network evacuation; online disaster management; dynamic traffic assignment; VANET.

## 1 INTRODUCTION

To mitigate and decrease losses caused by disasters, we need to evacuate people from the affected areas to safe areas, i.e., shelters. Evacuation orders are then vital and should be effective in evacuating people safely. Evacuation orders must dynamically change with the hazard evolution and the evacuees' requirements for real-time guidance. This real-time aspect of evacuation orders could only be targeted with telecommunication technology (Pan *et al.*, 2016). In other words, dynamic population evacuation (DPE) could be studied by new emerging technologies of vehicle-to-everything (V2X) communication.

In this context, the success of an evacuation plan depends mainly on two choices of evacuees: the locations of shelters and the evacuation route toward the selected shelter. Mathematically speaking, finding the best shelter and route considering the dynamic relationships between paths, time, and network characteristics is known as solving dynamic shelter allocation problem (SAP) and dynamic traffic assignment (DTA) problems. The shelter and route choice models of evacuees used in the literature are based on three principles: user equilibrium (UE), known as Nash Equilibrium, system optimum (SO), and the nearest allocation (NA) approach. The difference between these models relies on the evacuees' behaviors. In UE models, each traveler aims to minimize his individual travel time (Wardrop, 1952). From the system point of view, the ultimate goal is to minimize the total evacuation time or network clearance time (i.e., the arrival time of the last evacuee to its shelter). Under the SO principle, travelers may not be assigned to the fastest route for the benefit of the overall system, which could be difficult to accept by evacuees. The NA model assigns evacuees the shortest path based on the distance between the hazardous zone and the shelters. Such a model cannot provide acceptable results from evacuees' and system points of view.

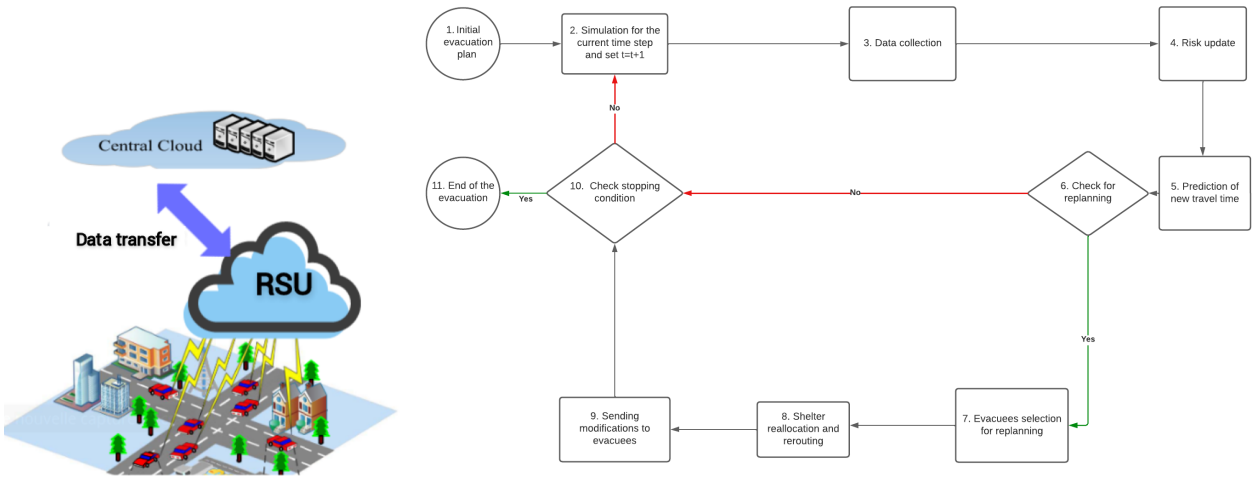
The reactive nature of both SAP and DTA limits their effectiveness during the evacuation process, i.e., they are more contributing to the planning phase compared to online management (Pan *et al.*, 2016). In this context, adding telecommunication technologies moves one step forward by providing effective methods for proactive rerouting when an emergency is predicted based on real-time traffic information.

In the literature, two main categories of studies focus on evacuation models. The first group solves both shelter allocation or/and traffic assignment to minimize users' or/and system costs. The second group focused on vehicular communication techniques and protocols to handle the DPE. We carry out a literature review of these two groups and highlight the existing research gaps. Few studies considered both SAP and traffic assignment in the dynamic context. Hsu & Peeta (2014) considered the evacuation planning process, including the DTA problem with a fixed shelter allocation schema. In our previous study, (Idoudi *et al.*, 2022), we tackled both dynamic problems sequentially, minimizing the total travel time in SAP, and calculating UE for DTA. However, we did not consider telecommunication networks in the methodology and also clearance time as the objective function for SAP. To the best of our knowledge, no study in the literature addresses the DPE problem considering both the planning phase (optimization methodology for SAP and DTA) and the online management phase (taking into account vehicle communication). However, the DPE problem is addressed with telecommunication technologies in the literature independently of the planning phase.

This study proposed a new framework to solve DPE problems, considering the planning phase and vehicular communication during the online evacuation management phase. In the planning phase, an initial evacuation plan is obtained by dynamically solving the SAP for destination choice and DTA for path choice toward the chosen destinations. The DPE process then starts with the initial plan, and thanks to the vehicular ad hoc network (VANET), evacuees communicate during the evacuation. We use VANET under the vehicular cloud computing (VCC) architecture to update vehicle decisions chosen by the initial plan, considering the dynamic evolution of the hazard and traffic congestion levels.

## 2 METHODOLOGY

We design an online evacuation management framework wherein we represent the evacuation process over time. The goal is to respond to unexpected events and dynamically modify the initial plan during the evacuation by considering vehicles' communication capacity that permits evacuees to receive and send new information and update their routes. Our online evacuation guidance system is composed of three components, as depicted by Figure 1(a): The first component is a centralized traffic monitoring and rerouting service representing the cloud server (which can physically be distributed across several servers). The second one is the set of roadside units (RSU) that represent the network infrastructure that should be existent to ensure communication between the cloud server and vehicles. The third layer is represented by vehicles equipped with onboard units (OBU). Through this communication network, we can re-plan the evacuation route and shelter during the evacuation process and provide them with instructions in real-time. The proposed methodology of this study is presented in Figure 1(b). The steps of the framework are detailed as follows:




(a) Vehicular cloud computing.

(b) Flowchart of the solving the DPE problem.

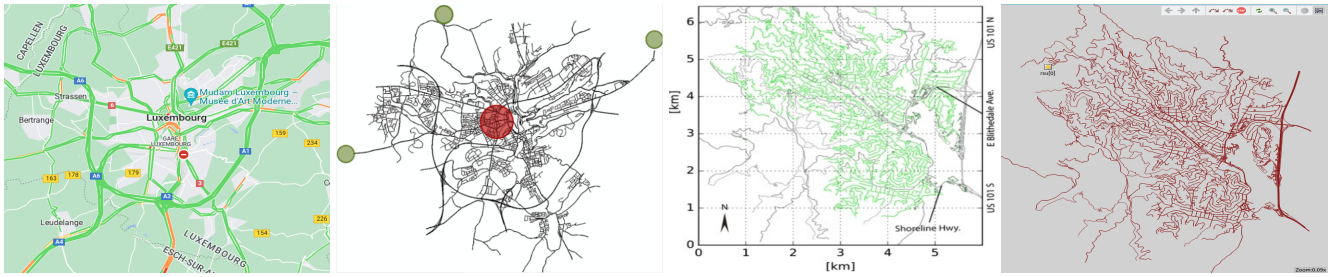
Figure 1: Proposed methodology for DPE problem.

**Step 1: Initial evacuation plan:** Solving the multi-level DTA and SAP to generate an evacuation plan.
**Step 2: Simulation for the current time step and set t=t+1:** Simulating the evacuation process that could be the same as proposed by the plan, or new events could occur due to several decisions made by evacuees. We have also to increment the simulation time index. Note that any trip-based dynamic simulator can be used in this step.
**Step 3: Data collection:** Each vehicle (node) broadcasts data messages, using their OBU, to RSUs that send it to the cloud server.
**Step 4: Risk update:** Update the risk based on data from step 3. The considered risk consists of two main components: the vehicle's distance from a hazardous area and the congestion levels of the vehicle's location.
**Step 5: Prediction of new travel times:** The travel time of edges might change according to the risk and congestion evolving by Step 4. In this step, we use a prediction model to predict new travel times. Here we use the prediction model embedded in the used simulator.

**Step 6: Check for replanning:** Deciding whether a user *i* is concerned by the rerouting process or not. For user *i* we estimate edge density, including the road speed and traffic density based on the Greenshield model (Pan *et al.*, 2016). The user *i* is considered to be in congestion if his current edge density is above a certain threshold.
**Step 7: Evacuees selection for replanning:** Selecting vehicles that to go to another safe destination or have to be rerouted before getting inside a congested edge (road). For shelter reallocation, we select vehicles if there is congestion in front of their original destinations, and the server asks them to go to a less congested destination. For rerouting, we select all users with the congested edge as the next planned edge of their initial journey.
**Step 8: Shelter reallocation and rerouting:** Prepare a message to the targeted users.
**Step 9: Sending notification to evacuees:** Cloud server sends its decisions to RSUs that forward the results to vehicles to react accordingly.
**Step 10: Check stopping condition:** Check if all the demand is evacuated or not. If yes, we had to end the process. Otherwise, we go to simulate the next step.
**Step 11: End of the simulation:** We end the evacuation process and calculate different performance measures.

# 3 NUMERICAL EXPERIMENT

We implement our proposed solution in order to model the DPE problem on two realistic networks. The first network is the LuST scenario, which represents Luxembourg city (Figures 2(a) and 2(b)), and the second is the network of Mill Valley city in California (Figures 2(c) and 2(d)). We employed a solution method using the simulation-based DTA by performing all simulations by SUMO simulator. We calculated the C-logit model and the prediction of travel time by SUMO (Lopez *et al.*, 2018). We used ILOG CPLEX version 12.9 to implement the SAP model and solve it. To simulate the scenario taking into account vehicular communication, we used the Veins/Omnet++ simulator and a cloud computing architecture based on works done in Wang *et al.* (2020).




(a) Luxembourg mapping data (b) SUMO city network (c) Mill Valley mapping data (d) OMNET city network

Figure 2: Evacuation network map of Luxembourg city

# 4 PRIMARY RESULTS AND DISCUSSION

We use the evaluation metrics presented in Idoudi *et al.* (2022) to measure the quality of the solution provided by our methodology. Here we present the primary results of the four following scenarios for the Luxembourg network:

- **Scenario P+C: Scenario with both planning and vehicular communication:** This scenario follows the proposed framework (demonstrated in figure 1(b)).
- **Scenario P: Scenario with the initial plan only:** This scenario illustrates the case of just planning for evacuation without any communication between vehicles or vehicles to RSUs.
- **Scenario C: Scenario with vehicular communication only:** This scenario is the same as figure 1(b) except in step 1, where evacuees consider the nearest shelter and choose their routes following the SUE.
- **Scenario N: Naive scenario:** This scenario represents the case where the system operators do not provide guidelines for evacuees. It means that the evacuees choose the nearest shelter and choose their routes

following the SUE without vehicular communication.

The results in Table 1 show a significant improvement in the quality of the final solution obtained by scenario P+C wherein we used both planning and online guidance models. For instance, the reduction of more than 18 minutes (39%) in the network clearance time compared to the naive scenario. Also, there is an improvement of more than 2 minutes (10%) between scenario P+C, and scenario P. Results show that scenario P represents the second-best solution. The comparison between scenarios P+C and P proves that new orders handling new (not expected) events in planning create a more successful evacuation operation.

Table 1: Performance metrics

| **Metrics / Scenario** | **P+C** | **P** | **C** | **N** |
|---|---|---|---|---|
| Network clearance time(s) | 1775.00 | 1980.00 | 2765.00 | 2835.0 |
| Mean evacuation time(s) | 1071.54 | 1093.70 | 1407.92 | 1447.61 |
| Average travel delay (ATD) | 205.47 | 220.62 | 341.63 | 349.78 |
| Average evacuation delay (AED) | 241.32 | 366.65 | 366.65 | 392.12 |

Besides, scenario C provides a better solution compared to scenario N, meaning that using the telecommunication network can improve the evacuation solution, even without any planning phase. This observation could prove the effectiveness of online communication and highlights the importance of giving new orders to evacuees to revise their route choice during the evacuation process. Inspecting the result for scenario P and scenario C, we can observe that planning contributes more than telecommunication during the evacuation operation. One of the reasons behind this observation is that in scenario C the shelter allocation was done without considering the congestion level. We have monitored scenario C to have a better view and understand more of the effect of online evacuation guidance. We observe that allocating all users to the same nearest shelters in all evacuation operation generate congestion that could not be escaped even by using online vehicle rerouting. That is why different shelters, like in scenario P in each state, will ensure that we assign evacuees to the closest destinations in terms of time-dependent shortest path and not distance measure.

In addition to these results, we applied our methodology to the large-scale network of Mill Valley city in California (Figures 2. We plan to analyze the performance of the proposed framework by changing the penetration rate of connected vehicles in VANET. Besides, we modified the SAP formulation by transforming it from the p-median model to the p-center model to minimize the clearance time and total travel time together and investigate which objective function is more suitable for the DPE problem.

# References


Hsu, Yu-Ting, & Peeta, Srinivas. 2014. Risk-based spatial zone determination problem for stage-based evacuation operations. *Transportation research part C: emerging technologies*, **41**, 73–89.

Idoudi, Hassan, Ameli, Mostafa, Van Phu, Cyril Nguyen, Zargayouna, Mahdi, & Rachedi, Abderrezak. 2022. An Agent-Based Dynamic Framework for Population Evacuation Management. *IEEE Access*, **10**, 88606–88620.

Lopez, Pablo Alvarez, Behrisch, Michael, Bieker-Walz, Laura, Erdmann, Jakob, Flötteröd, Yun-Pang, Hilbrich, Robert, Lücken, Leonhard, Rummel, Johannes, Wagner, Peter, & Wießner, Evamarie. 2018. Microscopic Traffic Simulation using SUMO. *In: The 21st IEEE International Conference on ITS*. IEEE.

Pan, Juan, Popa, Iulian Sandu, & Borcea, Cristian. 2016. Divert: A distributed vehicular traffic re-routing system for congestion avoidance. *IEEE Transactions on Mobile Computing*, **16**(1), 58–72.

Wang, Zhen, Zheng, Sifa, Ge, Qiang, & Li, Keqiang. 2020. Online offloading scheduling and resource allocation algorithms for vehicular edge computing system. *IEEE Access*, **8**, 52428–52442.

Wardrop, John Glen. 1952. Road paper. some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, **1**(3), 325–362.

# Improving Air Traffic Flow Management with Macroscopic Traffic Flow Modelling: Model Formulation, Simulation and Comparison

**Christopher Cummings**
Graduate Research Assistant
Transportation Center
Northwestern University
600 Foster Street
Evanston, IL 60208, USA
Email: christopher.cummings@northwestern.edu

**Vasileios Volakakis**
Graduate Research Assistant
Transportation Center
Northwestern University
600 Foster Street
Evanston, IL 60208, USA
Email: vasileios.volakakis@northwestern.edu

**Hani S. Mahmassani***
William A. Patterson Distinguished Chair in Transportation
Director, Transportation Center
Northwestern University
600 Foster Street
Evanston, IL 60208, USA
Email: masmah@northwestern.edu
Tel: 847.491.2276

Submission Date:

* Corresponding Author

**ABSTRACT**

Advanced air mobility (AAM) and air taxi services may add hundreds or even thousands of aircraft to the airspace. The resultant high density of aircraft will cause congestion issues, motivating the formulation of an efficient air traffic flow management (ATFM) system. The main goal of this work is to investigate and present how macroscopic air traffic flow models can be utilized to improve ATFM across a given network. An ATFM optimization problem was formulated using macroscopic air traffic flow models to track and predict congestion levels and travel times. The formulation's objective is to efficiently assign aircraft departure times and route choices such as to minimize the total travel time of the network. The travel times of every link/sector are directly connected to the corresponding density in each link/sector, providing a more accurate accounting for congestion within the network. A dynamic flow management approach is taken, which determines aircraft departure times and routes on a rolling basis. The introduced algorithm is used to manage simulated air traffic flows in a microscopic air traffic simulator with two different network structures. Comparisons are made with two other optimization approaches from the literature. The results demonstrate that the macroscopic traffic flow models provide a more accurate method for tracking and predicting congestion in the network than other methods, ultimately improving the efficiency of the ATFM algorithm. The algorithm provides a basis for an ATFM system that can efficiently manage hundreds or thousands of aircraft in a limited airspace.