UNIVERSITY OF LEEDS

This is a repository copy of *Multiclass dynamic system optimum solution for mixed traffic of human-driven and automated vehicles considering physical queues*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/169409/

Version: Accepted Version

**Article:**
Ngoduy, D, Hoang, NH, Vu, HL et al. (1 more author) (2021) Multiclass dynamic system optimum solution for mixed traffic of human-driven and automated vehicles considering physical queues. Transportation Research Part B: Methodological, 145. pp. 56-79. ISSN 0191-2615

https://doi.org/10.1016/j.trb.2020.12.008

© 2021 Elsevier Ltd. All rights reserved. This manuscript version is made available under the CC-BY-NC-ND 4.0 license http://creativecommons.org/licenses/by-nc-nd/4.0/.

**Reuse**
This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

# Multiclass dynamic system optimum solution for mixed traffic of human-driven and automated vehicles considering physical queues

D. Ngoduy[a,*], N.H. Hoang[a], H.L. Vu[a], D. Watling[b]

[a] *Institute of Transport Studies, Monash University, Australia*
[b] *Institute for Transport Studies, University of Leeds, UK*

**Abstract**

Dynamic traffic assignment (DTA) is an important method in the long term transportation planning and management processes. However, in most existing system optimum dynamic traffic assignment (SO-DTA), no side constraints are used to describe the dynamic link capacities in a network which is shared by multiple vehicle types. Our motivation is based on the possibility for dynamic system optimum (DSO) to have multiple solutions, which differ in where queues are formed and dissipated in the network. To this end, this paper proposes a novel DSO formulation for the multi-class DTA problem containing both human driven and automated vehicles in single origin-destination networks. The proposed method uses the concept of link based approach to develop a multi-class DTA model that equally distributes the total physical queues over the links while considering explicitly the variations in capacity and backward wave speeds due to class proportions. In the model, the DSO is formulated as an optimization problem considering linear vehicle composition constraints representing the dynamics of the link capacities. Numerical examples are set up to provide some insights into the effects of automated vehicles on the queue distribution as well as the total system travel times.

*Keywords:* multi-class two regime transmission model, dynamic system optimum, physical queues, mixed traffic networks, automated vehicles

## 1. Introduction

Advances in communication technologies have resulted in new vehicle generations, such as automated vehicles (AVs). In general, an AV can obtain neighbouring information via vehicle-to-vehicle (V2V) communication and/or vehicle-to-infrastructure (V2I) communication, and then adopt a suitable control law to achieve a certain objective, such as maintaining a short constant headway within vehicles or smooth driving patterns. Recent studies have shown that the use of AVs have the potential to improve road safety, traffic efficiency, and environmental sustainability (Jia and Ngoduy, 2016a,b, Levin and Boyles, 2016a,b, Ngoduy, 2013b,c, Shladover and Lu, 2012, Zhao et al., 2018). Nevertheless, there is still a long way to go before full deployment of AVs vehicles on roads is viable. It has been predicted that by 2030 AVs will make up a significant share of the vehicle market, which indicates a long lifespan for mixed traffic flow consisting of both human-driven vehicles (HVs) and and AVs. Modelling and managing the dynamics of such mixed traffic flow have received a great attention over the last few years. The majority of the research focuses on the short-term planning, devoted to the understanding of the complex (asymmetric) interactions between HVs and AVs in traffic flow dynamics (Delis et al., 2015, Kesting et al., 2010, Levin and Boyles, 2016a, Ngoduy, 2012, 2013a, Ye and Yamamoto, 2018). In general, the AVs will increase link capacity because of the smaller headway. For example, Levin and Boyles (2016b) extended the Cell Transmission Model-CTM

*Corresponding author.
*Email address:* dong.ngoduy@monash.edu (D. Ngoduy)

(Daganzo, 1995) to capture variations in capacity and backwards wave speed in response to class proportions (e.g. the penetration of AVs) within each cell. Recently, Melson et al. (2018) have adopted a different shape of the fundamental diagram for AVs and integrated it with the HVs in the Link Transmission Model-LTM, originally proposed by Yperman et al. (2005).

The Link Tranmission Model (LTM) in either discrete form (Yperman et al., 2005) or continuous form (Han et al., 2015, Jin, 2015) has been developed using the Newells theory which is a special formulation of Lax-Hopf (LH) formula, where the state of the whole link (i.e. either free-flow or congested) will be determined by the entry and exit flow. More specifically, the flow propagation in LTM is based on Newell's cumulative flow curves applied at the entry/exit of each link, with node models used to calculate the transition flows, which are based on conservation of flow between the incoming and outgoing flows. Sending/receiving flows, together with transition flows and other flow constraints, form the basis for updating the cumulative flows at the link boundaries. Osorio and Flotterod (2014), Osorio et al. (2011) have then developed a stochastic version of LTM, which is a so-called Double Queue Model (DQM). In the DQM, the link is treated as a set of two queues, referred to as the upstream queue and the downstream queue. Both LTM and DQM can properly capture the free-flow travel time delay when the link is in a free-flow state and the backward shoc-kwave time delay when the link is in a congested state, which make them possible to capture queue spillbacks. The DQM was used in Ma et al. (2014) to find a free-flow DSO solution where spillback is tracked by the traffic state at the link entrance (in free flowing) or at the link exit (in congested) accounting for some time shift. Nevertheless, either LTM or DQM does not determine explicitly the propagation of the front shocks within a link and thus is unsuitable for providing the detailed traffic state within the link. To fulfill our objective (i.e. equal distribution of the queues over links to reduce the heterogeneity of congestion), we must track the time and space evolution of the queue lengths in the DSO problem. While LTM or DQM cannot describe such evolution of the queue lengths, CTM can do it with high computational cost, especially for the DSO problem. We will address this problem by applying a recently proposed Two-regime Transmission Model (TTM) in Ngoduy et al. (2016) for our DSO problem. The TTM principle is more desirable for our DSO-related problem than using the CTM, LTM or DQM as, on the one hand, it utilizes the entry and exit flows to describe the link state similarly to the LTM or DQM, and, on the other hand, it provides the time and space evolution of the queue lengths.

Basically, the TTM assumes that the link is homogeneous (e.g. in terms of link capacity) and the actual bottleneck is at the downstream node. This is highly justified in the urban traffic context as the majority of traffic congestion occurs at the downstream node and the tail of such congestion may propagate either upstream or downstream of traffic flow depending on the traffic incoming from the upstream node and exiting at the downstream node. Furthermore, the TTM also assumes that traffic along a link is characterised by two regimes (i.e. either non-congested where the density is below a critical density or congested where the density is above the critical density). Depending on how traffic flows into and flows out of the link, the time-varying length of each regime and the evolution of traffic density in time and space are determined. With these assumptions, we are only interested in the congestion formation and the propagation of the tail of the congestion under a certain boundary condition (when and where a traffic congestion occurs).

It has been shown recently by Shen and Zhang (2014) that DSO may have multiple solutions which share the same DSO objective (i.e. the total travel times) but have different queue lengths at nodes in the network. Our aim is to find a DSO solution that optimally distributes the congestion over links inside the network which essentially reduces queue spillbacks. In order to do so, we require more information than the LTM can offer, but wish to avoid the computational burden of CTM for DSO. We thus adopt an extension of the TTM in Ngoduy et al. (2016) to capture the effects of AVs in the context of dynamic traffic assignment (DTA) problems, particularly how various initial penetrations of the AVs affect the queue distribution at nodes in the dynamic system optimum-DSO principle. To this end, this paper focuses on a multi-class dynamic traffic assignment-system optimum problem, which equally distributes the (horizontal) queue over links under mixed AVs and HVs operations. We expect that such equal distribution of the queue over links will lead to better traffic management inside the network (e.g. reduced queue spillbacks) such as road allocations and priority control for specific vehicle classes. Therefore, there is a need to explicitly determine the time and space evolution of the queues in such multi-class traffic networks. Particularly, how the variations in capacity and backwards wave speed affect the distribution of the physical queues at nodes.

Multi-class static traffic assignment has been initially introduced in Dafermos (1972). van Vuren and Watling (1991) have investigated the concept of multi-class traffic assignment further where one user class (i.e. guided one) is aiming to minimise marginal travel costs (as would be the case in a pure system optimum (SO) pattern) and the other class is aiming to minimise actual travel costs (as would be the case in a pure user equilibrium (UE) UE pattern). In a similar line, Bagloee et al. (2017) argued that the AVs are seeking the SO principles, while the remainder (i.e. the HVs) continue to pursue a UE pattern. This idea is then formulated using the static nonlinear complementarity problem (NCP) in which a mixture of SO and UE representing HVs is enforced. Chen et al. (2016) proposed a mathematical procedure to optimally deploy AV lanes considering the endogenous AV market penetration. Given AV lanes deployed in a general road network, the flow distributions of both HVs and AVs are captured by a multi-class network equilibrium model. Recemtly, Wang et al. (2019) proposed a multi-class traffic assignment model, where HV users and AV users follow different route choice principles, characterized by the cross-nested logit (CNL) model and UE model, respectively. Their proposed multi-class model can capture the characteristics of mixed traffic flow such as the difference in value of time between HVs and AVs and the asymmetry in their driving interactions, thereby enhancing behavioral realism in the modeling.

Nevertheless, all of these strategies have been studied in a stead-state (or static) framework, while the representation of traffic dynamics and driver behavior (e.g. different reaction times and backwards speeds) has not been yet taken into account. Especially, there has been a discernible trend in the literature toward those that explicitly integrate the dynamics of flow predicted by a traffic model with DTA/DSO problems, such as CTM (Lo and Szeto, 2002), LTM or DQM (Gentile et al., 2007, Han et al., 2016, Long et al., 2018, Osorio et al., 2011) and Two-regime Transmission Model-TTM (Ngoduy et al., 2016). This dynamic approach has then been extended to capture the route choices of multi-class traffic flow by Szeto et al. (2011) using a multi-class CTM or Jiang et al. (2016) using a multi-class LTM. Given the background described above, we aim to formulate a DSO problem with a KWM as linear constraints. To fulfill our objective (i.e. equal distribution of the queues over links), we must track the time and space evolution of the queue lengths in the DSO problem. We will address this problem by applying the TTM in Ngoduy et al. (2016) for our DSO problem of single O-D traffic networks containing mixed AVs and HVs. Note that in the scope of this paper, we are not dealing with lane changing behavior or more sophisticated maneuver close to intersection, which is a challenging problem itself in modelling mixed traffic flow dynamics. A microscopic model for control of car-following and lane-changing has been developed in Wang et al. (2015). Xie et al. (2019), Zhang et al. (2019) proposed deep learning methods for calibrating microscopic models of lane-changing and combined car-following and lane-changing in systems of HVs, respectively. Mohajerpoor and Ramezani (2019) estimated mean and variance of headway in a system of mixed AV/HV traffic and the impact of various lane allocation policies. Ghiasi et al. (2019) proposed speed trajectory controls for a mixed system of AVs and HVs, and considered its extension to multi-lane scenarios. Others considered safety implications of mixed AV/HV operations with a multi-lane discrete simulation (Ye and Yamamoto, 2019).

In this paper, we restrict our analysis to a (lane-aggregated) macroscopic model for mixed AV/HV operations which is simple enough to be embedded into a DTA framework as a set of linear constraints. More specially, we will illustrate that modelling explicitly the queue length dynamics along the link in the multi-class SO-DTA will help reduce the spillback problems in the congested network. The main contributions of this paper are given below:

- We develop a method to model mixed AVs and HVs in DTA. We derive a new optimisation framework to find a multi-class DSO solution which can equally distribute the average queue lengths over links so that the spillback due to different classes can be reduced.

- We set out the theoretical background necessary for the formulation of the (network-level) multi-class TTM as a set of side constraints, in which the (dynamic) link based share factor of the AVs will be modelled as an extra set of variables.

The organization of this paper follows. We set out some preliminaries needed for the model development in Section 2. We propose a link aggregate fundamental diagram for the mixed traffic of automated and human-driven vehicles in Section 3. Section 4 formulates the multi-class kinematic wave model (Lighthill

and Whitham, 1955) as a set of constraints for a multi-class DSO traffic assignment problem in which the queue lengths are calculated explicitly. Section 5 presents a new optimization framework which finds DSO solutions equally distributing the total queue lengths over links in the network under various penetrations of the AVs, using the constraints formulated in Section 4. We discuss the complexity of the optimization problem in Section 6. Section 7 illustrates our numerical studies for a small and a reasonably large network in order to support the advantages of our approach. Finally, we conclude our paper in Section 8.

## 2. Preliminaries

A traffic network is considered a directed graph, which consists of a set of links connected via a set of nodes. Two rules for source and sink nodes:

- There is only one link from a source node, called source link.
- There is only one link to a sink node, called sink link.

The notation below will define a traffic network being considered in this paper.

- $\mathbb{T}$: a set of discrete time slots, $\mathbb{T} \subset \mathbb{I}$ (integer). For continuous time domain, we use notation $\mathbb{T}_{\mathbb{R}}$ ($\mathbb{T}_{\mathbb{R}} \subset \mathbb{R}$). We also define $T = |\mathbb{T}|$ as the number of time slots.
- $\mathbb{V}$: Set of nodes. There are two subsets: set of source nodes $\mathbb{V}_R$ and set of sink nodes $\mathbb{V}_S$, such that: $(\mathbb{V}_R \cup \mathbb{V}_S) \subseteq \mathbb{V}$ and, $\mathbb{V}_R \cap \mathbb{V}_S = \varnothing$.
- $\mathbb{M}$: Set of vehicle classes.
- $\mathbb{E}$: Set of directed links, combined by any two nodes in $\mathbb{V}$, e.g. if $(e_u, e_d) \in \mathbb{E}$ then we call $e_u$ the upstream node, and $e_d$ the downstream node of this link. There are three subsets of links: normal, source ($\mathbb{E}_R$) and sink ($\mathbb{E}_S$) links. The last two types are the specialized and virtual links to provide features of source nodes and sink nodes. For each link $a \in \mathbb{E}$ and vehicle class $m \in \mathbb{M}$ (e.g. $m = 1$ for HVs and $m = 2$ for AVs):
  - $l_a$: length of link $a$.
  - $\Upsilon^-_{am}$: set of inflow links to link $a$ of vehicles of class $m$.
  - $\Upsilon^+_{am}$: set of outflow links from link $a$ of vehicles of class $m$.
  - $S_{am}(t)$: upstream capacity (supply) of vehicles of class $m$ on link $a$ at time $t$.
  - $D_{am}(t)$: downstream capacity (demand) of vehicles of class $m$ on link $a$ at time $t$.
  - $u_{am}(t)$: incoming traffic flow to link $a$ of vehicles of class $m$ at time $t$.
  - $v_{am}(t)$: outgoing traffic flow from link $a$ of vehicles of class $m$ at time $t$.
  - $n_{am}(x,t)$: number of vehicles of class $m$ on link $a$ at location $x$ and time $t$.
  - $N_{am}(t)$: spatial average number of vehicles of class $m$ on link $a$ at time $t$.
  - $\rho_{am}(x,t)$: density of vehicles of class $m$ on link $a$ at location $x$ and time $t$.
  - $q_{am}(x,t)$: flow of vehicles of class $m$ on link $a$ at location $x$ and time $t$.
  - $\rho_a(x,t)$: total density on link $a$ at location $x$ and time $t$, $\rho_a(x,t) = \sum_{m \in \mathbb{M}} \rho_{am}(x,t)$.
  - $\alpha_{am}(x,t)$: share factor of vehicles of class $m$ on link $a$ at location $x$ and time $t$.
  - $\hat{\alpha}_{am}(t)$: link aggregate share factor of vehicles of class $m$ on link $a$ at time $t$.
  - $l^c_{am}(t)$: length of congested regime of vehicles of class $m$ on link $a$ at time $t$.
  - $l^f_{am}(t)$: length of non-congested regime of vehicles of class $m$ on link $a$ at time $t$.

∗ Note that: $u_{am}(t)$, $v_{am}(t)$, $S_{am}(t)$ and $D_{am}(t)$ are continuous in time domain $\mathbb{T}_{\mathbb{R}}$.

Also for each node $n \in \mathbb{V}$:

- $\Gamma^{-}_{nm}$: set of inflow links to node $n$ of vehicle class $m$.
- $\Gamma^{+}_{nm}$: set of outflow links from node $n$ of vehicle class $m$.

$$\begin{aligned}(\Gamma^{-}_{nm} \cup \Gamma^{+}_{nm}) &\subseteq \mathbb{E} \\ \Gamma^{-}_{nm} \cap \Gamma^{+}_{nm} &= \varnothing \\ \Upsilon^{-}_{am} &= \{e | a \in \Gamma^{+}_{nm}; e \in \Gamma^{-}_{nm}\} \\ \Upsilon^{+}_{am} &= \{e | a \in \Gamma^{-}_{nm}; e \in \Gamma^{+}_{nm}\}\end{aligned}$$

- $f_{abm}(t)$: upstream traffic at link $b$, coming from downstream traffic at link $a$ of vehicle class $m$.

$$\begin{aligned}a, b &\in \mathbb{E} \\ a \cap b &\neq \varnothing\end{aligned}$$

## 3. Multi-class fundamental diagram

In this section, we formulate the aggregate fundamental diagram for mixed AVs and HVs traffic flow for given spatial aggregate share factor of AVs using microscopic car-following approach. Let us start with a general continuous car-following model in the following form:

$$\frac{dv_j}{dt} = \Phi\left(s_j, v_j, \Delta v_j\right) \tag{1}$$

where $v_j$ and $s_j$ denote, respectively, the speed and distance gap of vehicle $j$. $\Delta v_j = v_j - v_{j-1}$ is the relative speed w.r.t. the leading vehicle $j-1$. The non-linear function $\Phi(.)$ denotes any elementary car-following models in the literature (for example the optimal velocity model-OVM, the full velocity diference model-FVDM, the intelligent driver model-IDM or the general motors model-GMM). Then at the steady-state condition (i.e. $s_j = s_e, v_j = v_e \ \forall j$ ):

$$\Phi\left(s_e, v_e, 0\right) = 0 \tag{2}$$

Solving the above equation will lead to the speed-density relation $v_e = v(\rho)$ where $\rho$ is the density. In other words, the car-following and macroscopic models share the same speed-density relationship and, therefore, the same fundamental diagram. We can always use any existing continuous car-following models (e.g. OVM, FVDM or IDM) to derive the fundamental diagram. However, we have chosen the IDM because it contains explicitly the safe time headway which is a most important factor reflecting the difference between the automated and human-driven vehicles. Note that some advanced car-following models for AVs (Jia and Ngoduy, 2016b, Liu et al., 2017) and mixed HVs and AVs (Jia et al., 2019) traffic dynamics can be used to derive the equilibrium fundamental diagram of mixed AVs and HVs traffic flow. However, they will likely result in non-linear flow-density relationships, which are not relevant for the derivation of the constraints of the DSO problem in the ensuing paper. For the sake of simplicity, we drop the link index in this section.

The adopted IDM Treiber et al. (2006) has been used rather extensively in the literature to model the dynamics of both human-driven and automated vehicles (Kesting et al., 2010, Ngoduy, 2013a, Sharma et al., 2019). The IDM for our mixed traffic problem can be described as below:

$$\frac{dv_j}{dt} = A_j \left[1 - \left(\frac{v_j}{V}\right)^{\delta} - \left(\frac{s^*_j}{s_j}\right)^2\right] \tag{3}$$

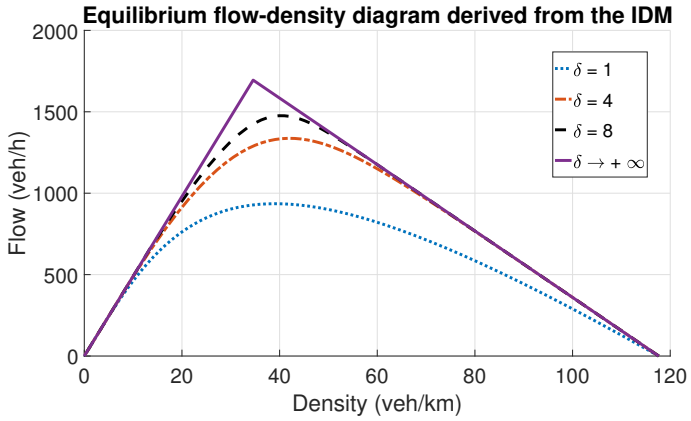


Figure 1: Impact of the exponent coefficient $\delta$ on the shape of the fundamental diagram

where $\delta$ is the acceleration exponent coefficient which specifies how the acceleration decreases when approaching the free-flow speed. $V$ is free-flow speed. The effective desired gap $s_j^*$ is defined as:

$$s_j^* = s_0 + v_j T_j + \frac{v_j \Delta v_j}{2\sqrt{A_j B_j}} \tag{4}$$

where $s_0$ is clearance gap (which is the same for all vehicles in our case), $T_j$ is a (constant) desired time gap of vehicle $j$. $B_j$ is the maximum deceleration of vehicle $j$. It is worth noticing that the last term in the right hand side of equation (4) is only active in non-stationary traffic corresponding to situations in which $\Delta v_j \neq 0$.

Under steady state conditions, where the acceleration $dv_j/dt = 0$, the follower and the leader travel at the same speed $v_j = v_{j-1} = v_e$, by using equations (3)-(4) we obtain:

$$s_j = \frac{s_0 + T_j v_e}{\sqrt{1 - \left(\frac{v_e}{V}\right)^{\delta}}} \tag{5}$$

Let us discuss here the impact of $\delta$ on the shape of the equilibrium space gap. When $\delta = 1$ we obtain an exponential relaxation to the free-flow speed $V$ with a finite relaxation time, while $\delta \to +\infty$ corresponds to approaching $V$ with a constant acceleration. The latter case conforms to the linear fundamental diagram which is used in the ensuing paper. For decreasing $\delta$ it becomes smoother and smoother, and various non-linear fundamental diagrams are obtained. However, such non-linear diagrams are not in the scope of our paper. Fig. 1 shows an example of various fundamental diagrams derived from the IDM with different values of $\delta$ for homogeneous traffic.

Let us assume that all vehicles regardless of the class travel with the same free-flow speed $V$ until they reach an aggregate critical distance headway $H_c$ (i.e. corresponding to the aggregate critical density) where the gap control policy in the AVs will be activated. Therefore, the approximation of equation (5) reads:

$$v_j = V \ \text{ if } \ s_j \geq H_c \ \text{ (free-flow regime)} \tag{6}$$

$$s_j = s_0 + T_j v_e \ \text{ if } \ s_j < H_c \ \text{ (congested regime)} \tag{7}$$

Obviously, the distance gap in the control mode (i.e. in the congested regime) depends on the desired time gap $T_i$. It thus depends on the type of car-following behavior of different vehicle class. In the randomly mixed traffic flow, there are four types of car-following:

- Automated vehicle following automated vehicle (A-A type). In this case, $T_j = T_{AA}$.
- Automated vehicle following human-driven vehicle (A-H type). In this case, $T_j = T_{AH}$.

| Car-following type | A-A type | A-H type | H-A or H-H type |
|---|---|---|---|
| Probability | $\hat{\alpha}_1^2$ | $\hat{\alpha}_1\hat{\alpha}_2$ | $\hat{\alpha}_2$ |

Table 1: Probability of car-following type

- Human-driven vehicle following automated vehicle (H-A type). In this case, $T_j = T_{HA}$.
- Human-driven vehicle following human-driven vehicle (H-H type). In this case, $T_j = T_{HH}$

where $T_{AA}$, $T_{AH}$, $T_{HA}$, and $T_{HH}$ denote, respectively, the desired time gap for A-A, A-H, H-A, and H-H car-following pair. In terms of driving behavior, the human-driven vehicles do not receive information from the leading vehicle, regardless of the class so the reaction time or desired time gap is identical for the H-A and H-H car-following type (i.e. $T_{HA} = T_{HH} = T_H$). Whereas, the A-H car-following type will have larger desired time gap than the A-A car-following type since the former does not receive information from the leading vehicle. However, the automated vehicle can react better to the front stimulus than the human so it is reasonable to assume that $T_{AA} < T_{AH} < T_H$. Under the equilibrium state, given the link aggregate share factor for a vehicle class $m$, $\hat{\alpha}_m$, (here $m = 1$ denotes AVs, $m = 2$ denotes HVs), the probability for each car-following type to happen is determined as in Table 1. Note that in Table 1: $\hat{\alpha}_1^2 + \hat{\alpha}_1\hat{\alpha}_2 + \hat{\alpha}_2$ $= \hat{\alpha}_1^2 + \hat{\alpha}_1(1 - \hat{\alpha}_1) + \hat{\alpha}_2 = \hat{\alpha}_1 + \hat{\alpha}_2 = 1$.

From table 1, in the congested conditions, the expected aggregate equilibrium distance gap is determined as:

$$\hat{s} = s_0 + v_e \left[\hat{\alpha}_1^2 T_{AA} + \hat{\alpha}_1\hat{\alpha}_2 T_{AH} + \hat{\alpha}_2 T_H\right] \tag{8}$$

which leads to:

$$v_e(\hat{s}) = \frac{\hat{s} - s_0}{\hat{\alpha}_1^2 T_{AA} + \hat{\alpha}_1\hat{\alpha}_2 T_{AH} + \hat{\alpha}_2 T_H} = \frac{\hat{s} - s_0}{\hat{T}} \tag{9}$$

where

$$\hat{T}(\hat{\alpha}_1, \hat{\alpha}_2) = \hat{\alpha}_1^2 T_{AA} + \hat{\alpha}_1\hat{\alpha}_2 T_{AH} + \hat{\alpha}_2 T_H \tag{10}$$

denotes the expected link aggregate desired time gap of the mixed AVs and HVs traffic.

At macroscopic scale, the the distance gap is the inverse of the density, i.e. $\hat{s} = \dfrac{1}{\rho} - l_v$ where $l_v$ is the average vehicle length, from equations (6)-(9) we obtain the aggregate fundamental diagram for mixed AV-HV traffic flow as below:

$$q\left(\rho, \hat{\alpha}_1, \hat{\alpha}_2\right) = \rho V \;\text{ if }\; \rho \le C(\hat{\alpha}_1, \hat{\alpha}_2) \;\text{ (free-flow regime)} \tag{11}$$

$$q\left(\rho, \hat{\alpha}_1, \hat{\alpha}_2\right) = \rho v_e(\hat{\alpha}_1, \hat{\alpha}_2) = (K - \rho)W(\hat{\alpha}_1, \hat{\alpha}_2) \;\text{ if }\; \rho > C(\hat{\alpha}_1, \hat{\alpha}_2) \;\text{ (congested regime)} \tag{12}$$

where $q$ is the expected aggregate flow and $\rho$ is the total density. $K$ denotes the jam density, which is identical for all vehicles and is calculated from the safe distance gap as $s_0 = 1/K - l_v$. The aggregate shock-wave speed, $W(\hat{\alpha}_1, \hat{\alpha}_2)$, is defined as:

$$W(\hat{\alpha}_1, \hat{\alpha}_2) = \frac{1}{K\hat{T}(\hat{\alpha}_1, \hat{\alpha}_2)} \tag{13}$$

The aggregate equilibrium flow-density equations (11)-(12) conform to a triangular flow-density fundamental diagram with (dynamic) aggregate critical density $C(\hat{\alpha}_1, \hat{\alpha}_2)$ and (dynami) aggregate capacity $Q(\hat{\alpha}_1, \hat{\alpha}_2)$ defined as below:

$$\begin{aligned} C(\hat{\alpha}_1, \hat{\alpha}_2) &= \frac{KW(\hat{\alpha}_1, \hat{\alpha}_2)}{V + W(\hat{\alpha}_1, \hat{\alpha}_2)} \\ Q(\hat{\alpha}_1, \hat{\alpha}_2) &= \frac{KVW(\hat{\alpha}_1, \hat{\alpha}_2)}{V + W(\hat{\alpha}_1, \hat{\alpha}_2)} \end{aligned} \tag{14}$$

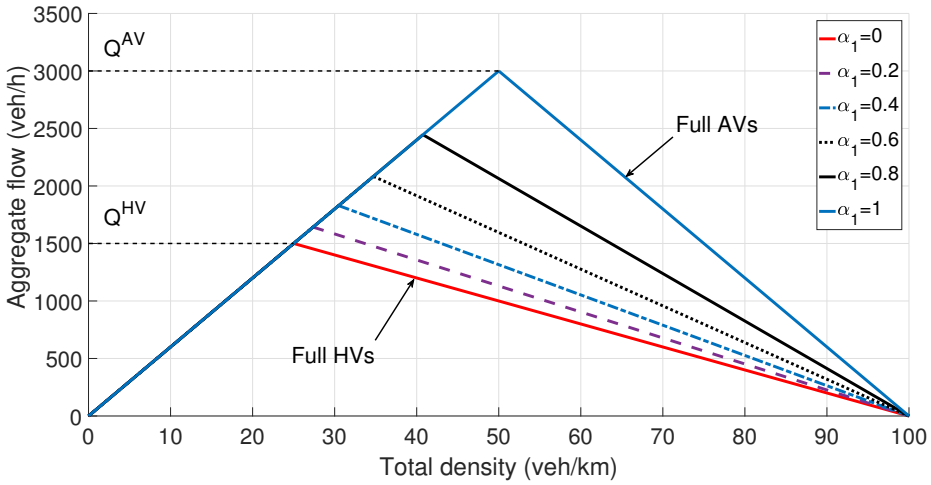


**Figure 2: Aggregate fundamental diagram for mixed AVs and HVs.**

The aggregate fundamental diagram for mixed AVs and HVs for various penetration of the AVs is depicted in Figure 2. In this fundamental diagram, $Q(\hat{\alpha}_1, \hat{\alpha}_2) = Q^{AV}$ when $\hat{\alpha}_1 = 1$ (i.e. the capacity of full AVs traffic) and $Q(\hat{\alpha}_1, \hat{\alpha}_2) = Q^{HV}$ when $\hat{\alpha}_1 = 0$ (i.e. the capacity of full HVs traffic). The link aggregate capacity is a function of the aggregate penetration of the AVs (i.e. $\hat{\alpha}_1$) over time, which will be included in our DSO solution in the ensuing sections.

## 4. Multiclass Two regime Transmission Model

This section describes our proposed model to be embedded in the dynamic system optimum problem in Section 5. More specifically, we extend the two traffic regime concept in Ngoduy et al. (2016) to account for the multiclass traffic network in this paper, which is latter called the multiclass two regime transmission model (MTTM). The following assumption is made for our model development.

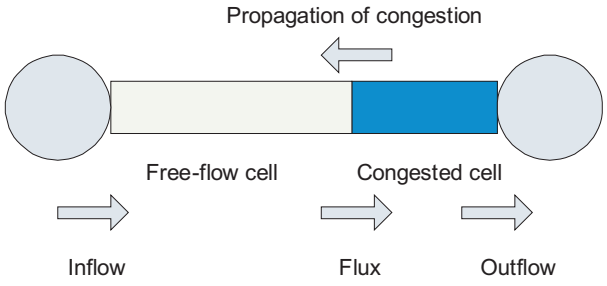


**Figure 3: Illustration of the two regime based model.**

**Assumption 1.** Traffic queue only happens at the downstream bottleneck of the link. As such, the link is divided into two regimes: free-flow and congested area (Figure 3). The length of each regime is time varying, subject to the class specific in-and out-flow.

It is worthy to mention that Assumption 1 actually has a rigorous proof, see Bretti et al. (2006). This assumption is true when the link of interest is empty at the beginning of the analysis period. It does not hold

when the initial density condition is specified arbitrarily (e.g. incidents) and congestion may start appearing anywhere on the link. Nevertheless, these cases are excluded from our model for the DSO problem in this paper.

*4.1. Basic model equations*

For the sake of simplicity, we drop out the link index in this section but the derivation holds for every link. The extension of LWR model to a number of vehicle classes has been undergone recently. In the ensuing section, we will recall the concept of the multiclass LWR model. According to the conservation law, each vehicle class should satisfy the following equation:

$$\frac{\partial \rho_m}{\partial t} + \frac{\partial q_m}{\partial x} = 0 \qquad \forall m \in \mathbb{M}. \tag{15}$$

To follow the model of Newell (1993) for each vehicle class, the dynamics of the number of vehicles each regime are given below:

$$\frac{dn_m(x,t)}{dt} = \begin{cases} u_m(t) - u_m\left(t - \dfrac{x}{V}\right) & \text{if } x < l_m^f(t) \text{ (in free-flow regime)} \\ v_m\left(t - \dfrac{l-x}{W(t)}\right) - v_m(t) & \text{if } l_m^f(t) \le x \le l \text{ (in congested regime)} \end{cases} \tag{16}$$

Then we can obtain the class specific number of vehicles in a link:

$$N_m(t) = \underbrace{\frac{1}{V}\int_0^{l_m^f(t)} u_m\left(t - \frac{x}{V}\right) dx}_{\text{Free-flow regime}} + \underbrace{\int_0^{l_m^c(t)} \left(K\alpha_m(t, l-x) - \frac{1}{W(t)} v_m\left(t - \frac{x}{W(t)}\right)\right) dx}_{\text{Congested regime}} \tag{17}$$

This equation indicates that the class specific number of vehicles in a link determined in each traffic regime by equation (16) has been implicitly incorporated and will be used latter to formulate the linear discrete link constraints for a dynamics system optimum problem.

*4.2. Definition of the discrete time variables*

In this paper, we shall use the time index $i$ or $k$ for discrete time, and $t$ or $h$ for continuous time. Let us define the following discrete time variables from the continuous TTM described above:

$$\begin{aligned} u_m(i) &= \int_i^{i+1^-} u_m(t)dt \\ v_m(i) &= \int_i^{i+1^-} v_m(t)dt \\ f_m(i) &= \int_i^{i+1^-} f_m(t)dt \qquad \forall i \in \mathbb{T} \\ D_m(i) &= \int_i^{i+1^-} D_m(t)dt \\ S_m(i) &= \int_i^{i+1^-} S_m(t)dt \end{aligned} \tag{18}$$

For the sake of simplicity without loss of generality let us assume that

$$\begin{matrix} \dfrac{l_a}{V} \in \mathbb{N} \\ \dfrac{l_a}{W} \in \mathbb{N} \end{matrix} \qquad \forall a \in \mathbb{E} \tag{19}$$

Basically, we will consider the class specific number of vehicles in a link $N_{am}(t)$ and the spatial average share factor $\hat{\alpha}_{am}(t)$ a major set of variables for the DNL problem. In continuous time, the class specific number of vehicles in a link at time instant $t \in \mathbb{T}_{\mathbb{R}}$ can also determined by the class specific in- and out-flow:

$$N_m(t) = \int_0^t [u_m(h) - v_m(h)]\,\mathrm{d}h. \tag{20}$$

In discrete time step, the class specific number of vehicles in a link at time step $i \in \mathbb{T}$ means:

$$N_m(i) = \lim_{t \to (i+1)^-} \int_0^t (u_m(h) - v_m(h))\,\mathrm{d}h = \sum_{k=0}^{i} [u_m(k) - v_m(k)] \tag{21}$$

Hence,

$$N(i) = \sum_{m \in \mathbb{M}} \sum_{k=0}^{i} [u_m(k) - v_m(k)] \tag{22}$$

Note, by definition, the link aggregate share factor is computed as: $\hat{\alpha}_m(t) = \dfrac{1}{l} \displaystyle\int_0^l \alpha_m(x,t)dx$. The corresponding discrete link aggregate share factor is $\hat{\alpha}_m(i)$. In the model formulation of the ensuing paper, we consider all the variables and dynamic parameters (i.e. aggregate capacity and wave speed) to be discrete in time.

*4.3. Formulation of side constraints*

These class specific number of vehicles in a link are bounded as follows.

**Proposition 1** (Lower bound of class specific number of vehicles in a link)**.**

$$N_m(i) \geq \sum_{k=i-\frac{l}{V}+1}^{i} u_m(k) \qquad \forall i \in \mathbb{T}. \tag{23}$$

*Proof.* See proof in Appendix A. □

Obviously, this condition also leads to the lower bound of the total number of vehicles in a link (Ngoduy et al., 2016):

$$N(i) = \sum_{m \in \mathbb{M}} N_m(i) \geq \sum_{m \in \mathbb{M}} \sum_{k=i-\frac{l}{V}+1}^{i} u_m(k) = \sum_{k=i-\frac{l}{V}+1}^{i} u(k). \tag{24}$$

**Proposition 2** (Upper bound of class specific number of vehicles in a link)**.**

$$N_m(i) \leq Kl\hat{\alpha}_m(i) - \sum_{k=i-\frac{l}{W(i)}+1}^{i} v_m(k) \qquad \forall i \in \mathbb{T}. \tag{25}$$

*Proof.* See proof in Appendix B □

This inequality also leads to:

$$N(i) = \sum_{m \in \mathbb{M}} N_m(i) \leq Kl \sum_{m \in \mathbb{M}} \hat{\alpha}_m(i) - \sum_{m \in \mathbb{M}} \sum_{k=i-\frac{l}{W(i)}+1}^{i} v_m(k) = Kl - \sum_{k=i-\frac{l}{W(i)}+1}^{i} v_m(k) \tag{26}$$

which is the linear upper bound of the total number of vehicles in a link (Ngoduy et al., 2016).

**Remark 1.** *In case the jam density is different for each vehicle class then equation (25) reads:*

$$N_m(i) \leq K_m l \hat{\alpha}_m(i) - \sum_{k=i-\frac{l}{W(i)}+1}^{i} v_m(k) \qquad \forall i \in \mathbb{T}. \tag{27}$$

**Theorem 1** (The existence and uniqueness of the class specific congested length $l_m^c(t)$)**.** *At time instant $t$, given $u_m(h), v_m(h)$ $(\forall h \leq t)$ and $\alpha_m(x,t)$ $(\forall x \in [0,l])$ satisfying conditions (23), (A.3), (A.4) and (25), there exists a value of $l_m^c(t)$ so that*

$$\int_0^{l-l_m^c(t)} \frac{u_m\left(t-\frac{x}{V}\right)}{V} dx + \int_0^{l_m^c(t)} \left[ K\alpha_m(l-x,t) - \frac{v_m\left(t-\frac{x}{W(t)}\right)}{W(t)} \right] dx = \sum_{h\leq t} [u_m(h) - v_m(h)].$$

*The solution of $l_m^c(t) = l^c$ is unique (at time $t$) if it satisfies the above condition and*

$$K\alpha_m(l-l^c,t) - \frac{v_m\left(t-\frac{l^c}{W(t)}\right)}{W(t)} - \frac{u_m\left(t-\frac{l-l^c}{V}\right)}{V} \geq 0.$$

*Proof.* Let $F(l^c)$ be a continuous function of $l^c$:

$$F(l^c) = \int_0^{l-l^c} \frac{u_m(t-\frac{x}{V})}{V} dx + \int_0^{l^c} \left[ K\alpha_m(l-x,t) - \frac{v_m(t-\frac{x}{W(t)})}{W(t)} \right] dx.$$

It is obvious that this function is a monotonically non-decreasing continuous function (of $l^c$) in $[0,l]$ because:

$$\frac{dF(l^c)}{dl^c} = K\alpha_m(l-l^c,t) - \frac{v_m(t-\frac{l^c}{W(t)})}{W(t)} - \frac{u_m(t-\frac{l-l^c}{V})}{V} \geq 0.$$

This means that we can obtain $l_m^c(i)$ given the class specific number of vehicles in a link $N_m(i)$. □

**Theorem 2** (Class specific queue dynamics)**.** *If time step is selected to be at least* $\max(l/W)$ *(Ban et al., 2012) so that $u_m(i)$ and $v_m(i)$ are steady in $i \in [0, \frac{l}{W} - 1]$ then the class specific queue dynamics is approximated as:*

$$l_m^c(i) = \frac{N_m(i) - \frac{u_m(i)}{V} l}{K\hat{\alpha}_m(i) - \frac{u_m(i)}{V} - \frac{v_m(i)}{W(i)}} \tag{28}$$

*Proof.* As $u_m(i)$ and $v_m(i)$ are steady in $i \in [0, \frac{l}{W(i)} - 1]$, we have:

$$u_m(t_0) = u_m(t_0+i) = u_m(i) \tag{29}$$
$$v_m(t_0) = v_m(t_0+i) = v_m(i) \tag{30}$$

then, based on Proposition 1 and Theorem 1, we can obtain:

$$N_m(i) = l_m^c(i) \left( K\hat{\alpha}_m(i) - \frac{u_m(i)}{V} - \frac{v_m(i)}{W(i)} \right) + \frac{u_m(i)}{V} l$$

which leads to:

$$l_m^c(i) = \frac{N_m(i) - \frac{u_m(i)}{V} l}{K\hat{\alpha}_m(i) - \frac{u_m(i)}{V} - \frac{v_m(i)}{W(i)}} \tag{31}$$

□

## 5. Dynamic system optimum problem with MTTM-based constraints

As DSO may have multiple solutions in which all the solutions share the same objective while the difference is where the queues are formed and dissipated in the network. This section formulates a problem to find DSO solutions which can evenly distribute the class specific queues (i.e. the length of the congested regime for class specific vehicles $l^c_{am}(t)$) over links in the network so that the spill-backs can be reduced.

### *5.1. Objective functions*

We formulate the DSO problem in this paper into two stages.

1. **Stage 1**: find a DSO solution via minimizing the total travel times or the total system cost. That is:

$$F = \max \sum_{m\in\mathbb{M}} \sum_{a\in\mathbb{E}_S} \sum_{i\in\mathbb{T}} (T+1-i)u_{am}(i) \tag{32}$$

   subject to the node and link constraints in sections 5.2. and 5.3. In the literature, Ngoduy et al. (2016), Shen and Zhang (2008) have proven that the objective function (32) is equivalent to the objective of minimizing the total travel time in case of the single user class traffic. The proof for multi-class traffic is rather straightforward and is given in Appendix E.
   Since Shen and Zhang (2014) have indicated that there exist many DSO solutions which share the same DSO objective (32) but have different queue lengths at nodes in the network. Stage 1 aims to find one of these solutions which gives such minimum total travel time while the purpose of Stage 2 is not minimisation of travel time, as that is already achieved in Stage 1, but a secondary objective that may be defined, separately from travel time minimisation. It is formulated as below.
2. **Stage 2**: Given the optimal total travel times found in Stage 1, find a traffic pattern which equally distributes the congestion in links.

$$\begin{aligned} &\text{Objective:} && G, \text{defined as the function of the queue length } l^c_a(i) \\ &\text{Constraints:} && \sum_{m\in\mathbb{M}} \sum_{a\in\mathbb{E}_S} \sum_{i\in\mathbb{T}} (T+1-i)u_{am}(i) = F \quad \text{(i.e. to achieve the DSO solutions)} \\ & && \text{and the node, link and non-holding-back constraints in Sections 5.3, 5.2 and 5.4.} \end{aligned} \tag{33}$$

   This step restricts the solution domain to system optimal solutions via the first constraint. As there exists a non-holding-back solution without changing the system optimal objective value (Shen and Zhang, 2014), it indirectly shows that **Stage 2** provides a feasible solution domain.

In this paper, we consider an objective functions $G$ below for the illustration purposes:

$$G = \min \sum_{m\in\mathbb{M}} \sum_{a\in\mathbb{A}} \sum_{i\in\mathbb{T}} \left( l^c - \frac{l^c_{am}(i)}{l_a} \right)^2 \tag{34}$$

where

$$l^c = \frac{1}{|\mathbb{M}||\mathbb{A}|T} \sum_{m\in\mathbb{M}} \sum_{a\in\mathbb{A}} \sum_{i\in\mathbb{T}} \frac{l^c_{am}(i)}{l_a}. \tag{35}$$

The Stage 2 formulated by the above $G$ function aims to find a DSO solution which can reduce the spillback issues due the propagation of the queues in the network. According to the estimation of congestion length shown in Theorem 2, the Statge 2 problem aims to minimize the variance of congestion level (equivalent to the ratio between the estimated congestion length and the link length) over all links in the network. We show in the numerical results in Section 7 and in the appendix Appendix F that Stage 2 can help reducing the spill-back traffic in the SO solution. Note that, the issue of spill-back happens when the congestion from downstream flow impacts on the upstream flow or $l^c_a = l_a$. This issue can be reduced

if travellers could divert to other less-congested routes. By doing this, we would reduce the congestion on heavy-traffic links and increase congestion on less congested links, meanwhile none of them would cause spill-back to the upstream, i.e. to guarantee $l_a^c < l_a$ for any link in the network. In our formulation, Stage 2 problem aims to divert the flow on heavy congested links to other free-flow or less-congested links. Therefore, it can help to reduce the issue of spillback in a DSO solution. Because the link lengths are different, the evaluation of congestion in each link is normalized with the link lengths in our objective function $G$.

**Remark 2.** *While the objective function $F$ in Stage 1 problem is linear, the objective function $G$ in Stage 2 problem is a quadratic (non-linear) function involving the dynamics of the physical queues. We will discuss the complexity of this problem in Section 6.*

The multi-class network-level constraints contain both link constraints and node constraints, which are described below.

*5.2. Link constraints*

From the Propositions and Theorems developed above, the class specific link constraints for the proposed model are summarized as below:

$\forall i \in \mathbb{T}$, $\forall m \in \mathbb{M}$, $\forall a \in \mathbb{E}$:

$$N_{am}(i) = \sum_{k=0}^{i} [u_{am}(k) - v_{am}(k)] \tag{36}$$

$$N_{am}(i) \geq \sum_{k=i+1-\frac{l_a}{V_a}}^{i} u_{am}(k) \tag{37}$$

$$N_{am}(i) \leq K_a l_a \hat{\alpha}_{am}(i) - \sum_{k=i+1-\frac{l_a}{W_a(i)}}^{i} v_{am}(k) \tag{38}$$

$$\hat{\alpha}_{am}(i) \geq 0 \tag{39}$$

$$\sum_{m \in \mathbb{M}} \hat{\alpha}_{am}(i) = 1 \tag{40}$$

Note that we have introduced an extra set of variables representing the aggregate link-based share factor $\hat{\alpha}_a^m(i)$.

*5.3. Node constraints and capacity constraints*

$\forall i \in \mathbb{T}, m \in \mathbb{M}$, $\forall a \in \mathbb{E}$:

$$\sum_{a \in \Gamma_n^-} v_{am}(i) = \sum_{a \in \Gamma_n^+} u_{am}(i) \tag{41}$$

$$u_{am}(i) = \sum_{b \in \Upsilon_a^-} f_{bam}(i) \tag{42}$$

$$v_{am}(i) = \sum_{b \in \Upsilon_a^+} f_{abm}(i) \tag{43}$$

$$S_a(i) = \min \left( Q_a(i); K_a l_a + \sum_m \sum_{k \le i - \frac{l_a}{W_a(i)}} v_{am}(k) - \sum_m \sum_{k \le i-1} u_{am}(k) \right) \tag{44}$$

$$D_a(i) = \min \left( Q_a(i); \sum_m \sum_{k \le i - \frac{l_a}{V_a}} u_{am}(k) - \sum_m \sum_{k \le i-1} v_{am}(k) \right) \tag{45}$$

$$\sum_m u_{am}(i) \le S_a(i) \tag{46}$$

$$\sum_m v_{am}(i) \le D_a(i) \tag{47}$$

$$f_{abm}(i) \ge 0 \tag{48}$$

Constraint (41) represents the conservation of the class specific vehicles at an immediate node while constraint (42) determines the total number of class specific vehicles entering a node from all upstream links of that node and constraint (43) defines the total number of class specific vehicles exiting a node from all downstream links of that node. In fact, equation (41) and pair of equations (42) and (43) are equivalent, and we only need one of them in the final model. Constraints (44) - (47) are the supply and demand constraints at the node, and the corresponding flow at the entry and the exit of the link. Note that the link capacity $Q_a(i)$ in these constraints is the aggregated capacity, taking into account the effect of each vehicle class on each other as illustrated in Figure 2. Constraint (48) guarantees the non-negative class specific flow on link.

*5.3.1. Source node constraints*

$$\begin{array}{c} u_{am}(i) = U_{am}(i) \\ \hat{\alpha}_{am}(i) = \dfrac{U_{am}(i)}{U_a(i)} \\ \{Q_a(i)\} \to \infty \end{array} \quad \forall a \in \mathbb{E}_R, i \in \mathbb{T}, m \in \mathbb{M} \tag{49}$$

where $U_{am}(i)$ and $U_a(i)$ denote, respectively, the class specific demand and the total demand to source link $a$ at time step $i$.

*5.3.2. Sink node constraints*

$$\begin{array}{c} v_{am}(i) = 0 \\ \{Q_a(i)\} \to \infty \end{array} \quad \forall a \in \mathbb{E}_R, i \in \mathbb{T}, m \in \mathbb{M} \tag{50}$$

*5.4. The non-holding-back constraints*

The non-holding back (NHB) solutions require the system to discharge flow as much as it can and have been widely discussed before in the literature (Shen et al., 2007). For our particular problem, the non-holding back condition is that if there is a positive flow $f_{abm}(t)$, then at least one of the above constraints (37, 38, 46, 47) has to meet the equality either at the downstream of link $a$ or at the upstream of link $b$. We formulate this constraint similar to the one in Ngoduy et al. (2016) as below:

$$c_a^1(i)c_b^2(i)c_a^3(i)c_b^4(i)\sum_{k\geq i+1} f_{abm}(k) = 0 \tag{51}$$

for all $i \in \mathbb{T}, m \in \mathbb{M}, a \in \mathbb{E}, b \in \Gamma_a^+$ where

$$c_a^1(i) = N_{am}(i) - \sum_{k=i+1-\frac{l_a}{V_a}}^{i} u_{am}(k)$$

$$c_b^2(i) = K_b l_b \hat{\alpha}_{bm}(i) - N_{bm}(i) - \sum_{k=i+1-\frac{l_b}{W_b(i)}}^{i} v_{bm}(k)$$

$$c_a^3(i) = Q_a(i) - \sum_m v_{am}(i)$$

$$c_b^4(i) = Q_b(i) - \sum_m u_{bm}(i).$$

## 6. The complexity and solution method

The summary of model in both Stages are presented below:

Stage 1: $\min F$ s.t. Eqs. (36)-(50).
Stage 2: $\min G$ s.t. Eqs. (33), (36)-(51).

The complexity of our optimization problems in terms of number of constraints and variables is $O(|\mathbb{A}|T)$. The challenge in solving these problems is caused by the non-linear objective function $G$ in Stage 2 and non-linear constraints in Eqs. (10), (13), (14) and (51). Note that in Eq. (14), because $\hat{\alpha}_1 + \hat{\alpha}_2 = 1$, we rewrite the formulation of $Q_a(i)$ as the function of $\alpha = \hat{\alpha}_1$ as below:

$$\hat{T}(\alpha) = \alpha^2(T_{AA} - T_{AH}) + \alpha(T_{AH} - T_H) + T_H$$

$$Q(\alpha) = \frac{KVW(\alpha)}{V + W(\alpha)}.$$

Then we have the following proposition:

**Proposition 3.** *The aggregate capacity $Q(\alpha)$ is always a convex function of the share factor $\alpha$.*

*Proof.* See Appendix C. □

**Remark 3.** *Due to the convex function $Q(\alpha)$, the Stage-1 model (i.e. finding general DSO solutions) is classified as a difference of the convex programming (DCP) problem. Lipp and Boyd (2016) showed that its local optimum is efficiently found by using the concave-convex procedure that leverages the ability of solving convex optimization problems. Alternatively, in this paper, we rely on the IPOPT solver (Biegler and Zavala, 2009) for both models. This solver could be applied to a general non-linear problem. Our numerical results show the reasonable computing time for solving Stage-1 model, i.e., less than 5 minutes in all results.*

***Non holding-back traffic for DSO solution.*** In Section 5.4, we present the constraints to avoid the issue of holding traffic at any nodes in the network. However, these constraints are non-convex which makes the Stage-2 problem more complicated than the Stage-1 problem. In this section, we propose a method to obtain a non-holding-back DSO solution without using these constraints. Let's consider the following problem $\mathcal{P}$:

$$[\mathcal{P}] \quad \max H \text{ s.t. Eqs. (33), (36)-(50)}$$

where

$$H = \sum_{a\in\mathbb{E}}\sum_{m\in\mathbb{M}}\sum_{i\in\mathbb{T}}(T+1-i)u_{am}(i).$$

Then we have the following proposition:

**Proposition 4.** *The solution of problem $\mathcal{P}$ is not holding traffic at any link in the network.*

*Proof.* See Appendix D. □

Note that, the objective function $H$ considers the aggregated total flow at any link in the network rather than only at the sink in the SO objective function $F$. By maximizing the incoming flow (and also outgoing flow via the Eq. (41)) at any link given the coming flow at destinations (via Eq. (33)), we can resolve the issue of non-holding-back traffic for the SO solution. By doing this (i.e. eliminating the non-holding back constraints), we can reduce the complexity of the Stage 2 problem and, similarly, we also solve this problem using the IPOPT solver.

## 7. Numerical results

The numerical results are studied in two different networks: a small-size Braess network and a larger-scale network (e.g. Sioux Falls network). In the Braess network, we demonstrate the system optimal solution via the optimal share factor and the distribution of traffic flow and physical queues in congested links. Whereas in the Sioux Falls network, we aim to show the computational complexity and its potential for solving larger scale networks.

### *7.1. Braess network*

The network topology in Figure 4 and its parameters are described in Table 2. The traffic demand departs from node $R$ (to the destination $S$) at a rate of 180 veh/min. With these network settings, we aim to create a bottleneck at the upstream of link (4,5) due to the merging traffic from links (2,4) and (3,4). As the congestion will propagate back to the source, it causes the reduced flow throughput and the decreased total travel time. In this example, we would like to study the impact of AVs on the system performance (i.e. the total travel times and the spill-back problems) in different situations:

- Case 1: Given the same total amount of demand, the AV penetration is varied.
- Case 2: Similar to case 1, but this case requires the equalization of average travel time (EATT) for each vehicle type.

In the numerical results below for case 1, we will show that the AVs are allocated more link capacity than HVs due to their higher capacity in congested links. In contrast, in case 2, the obtained solution has to guarantee the identical travel time for each vehicle type in order to avoid the bias control of traffic split per vehicle class in the network.

Let $U_m$ denote the total demand (over time) of a specific type of vehicles, $\text{TSTT}_m$ be the total system travel time for the vehicle class $m$. Hence, the additional EATT constraint for case 2 reads:

$$\frac{\text{TSTT}_1}{U_1} = \frac{\text{TSTT}_2}{U_2} \tag{52}$$

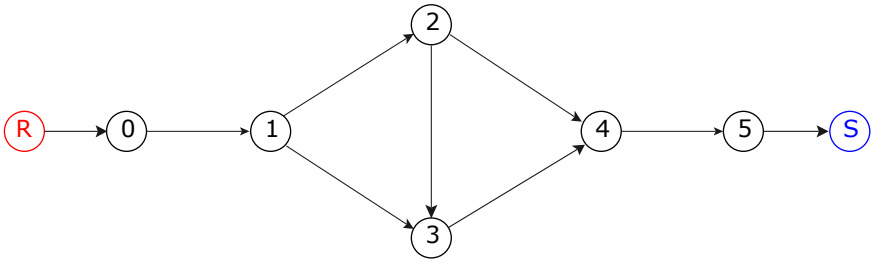


Figure 4: Braess network.

| Link | Length (m) | $K_{HV} = K_{AV}$ (veh/m) | $V$ (m/min) | $T_{AA}$ (sec) | $T_{AH}$ (sec) | $T_H$ (sec) |
|---|---|---|---|---|---|---|
| (1,3), (2,4), (2,3) | 400 | 0.12 | 1000 | 0.5 | 1.0 | 1.5 |
| (1,2), (3,4), (4,5) | 400 | 0.24 | 1000 | 0.5 | 1.0 | 1.5 |
| (0,1) | 400 | 0.36 | 1000 | 0.5 | 1.0 | 1.5 |

Table 2: Braess network parameters. Note that in our simulation set-up, links (1,3), (2,4) and (2,3) have 1 lane, links (1,2), (3,4) and (4,5) have 2 lanes, links (0,1) has 3 lanes

where $m = 2$ is for HVs, $m = 1$ is for AVs, and

$$\text{TSTT}_m = \sum_{a\in\mathbb{E}_R}\sum_{i\in\mathbb{T}} iU_{am}(i) - \sum_{a\in\mathbb{E}_S}\sum_{i\in\mathbb{T}}(T+1-i)u_{am}(i)$$
$$U_m = \sum_{a\in\mathbb{E}_R}\sum_{i\in\mathbb{T}} U_{am}(i).$$

This additional constraint is used in both Stages in case 2. The comparison between these two cases is further discussed below.

*Travel time versus the penetration of AVs.* The overall comparison between these cases is performed in terms of the effect of the penetration of AVs on the total travel time, as shown in Figure 5. This figure clearly shows that the total travel time is decreased with the increased AV penetration. Furthermore, the average travel time of AVs is always smaller than that of HVs. It certainly indicates that the AVs gain more priority over the HVs to pass through the network.

In contrast, the result for case 2 shows the solution where the equality of travel time for each class of vehicles is maintained in the system. It is interesting to observe that we could achieve this equality without increasing much the total travel time (less than 4%). In summary, the increase of AV penetration definitely improves the total travel times (at most 13% at 90% penetration of AV).

*Physical queue distribution versus the penetration of AVs.* In this part, we investigate the above scenarios in terms of the physical queue distribution, i.e., the smaller value of $G$ function, the better distribution of queue over the network. It is worth mentioning that, in the previous result of the total travel time, the Stage-1 objective value achieved in case 2 is slightly larger than in case 1 (about 0.5 %). Given these two different Stage-1 solutions, the result of Stage 2 in Figure 6 shows a similar impact of the AV penetration on the optimal distribution of physical queues. Furthermore, it also shows that more AV traffic would reduce the variance of queue length over the network.

Figures 7 and 8 show the spatio-temporal total density at links (1, 2) and (2, 4) obtained from two Stages in two scenarios. From these figures, we observe that the spill-back traffic between these links can be reduced in the Stage-2 solution. More specifically, spill-back does happen from link (2,4) to link (1,2) in Stage-1 solutions as the congestion in link (2,4) reaches upstream node (2). However, in Stage-2 solutions, we prevent the congestion in link (2,4) from reaching its upstream node (2) by shifting a part of it to other free-flow links, thus reduce the congestion in link (1,2).

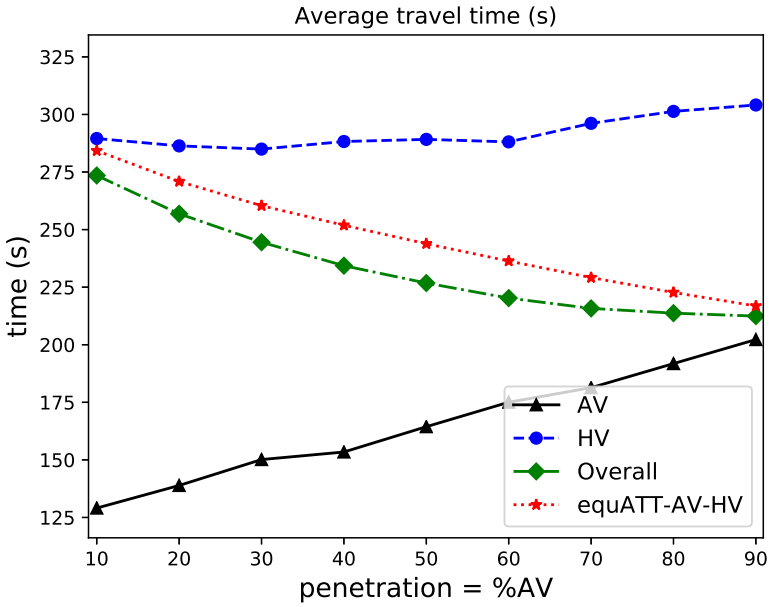


Figure 5: Travel time in Braess network with different penetrations of AV (with total demand of 360 vehicles).

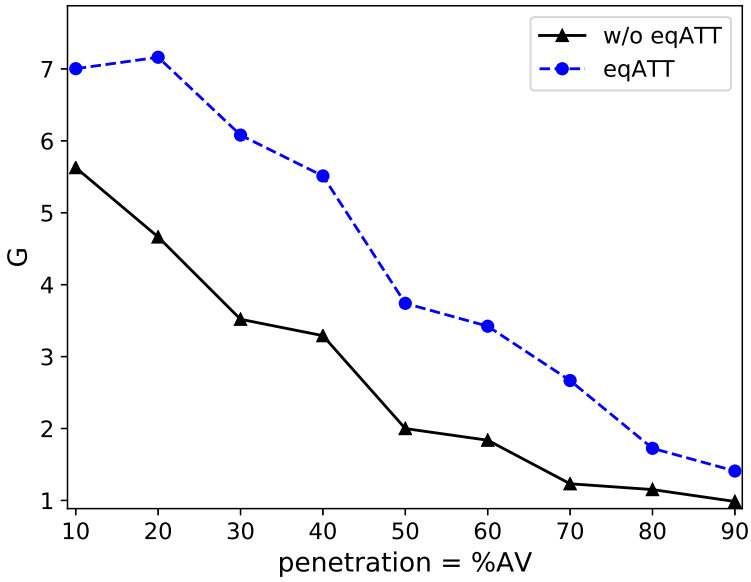


Figure 6: The performance of queue distribution in terms of $G$ in Braess network with different penetrations of AV

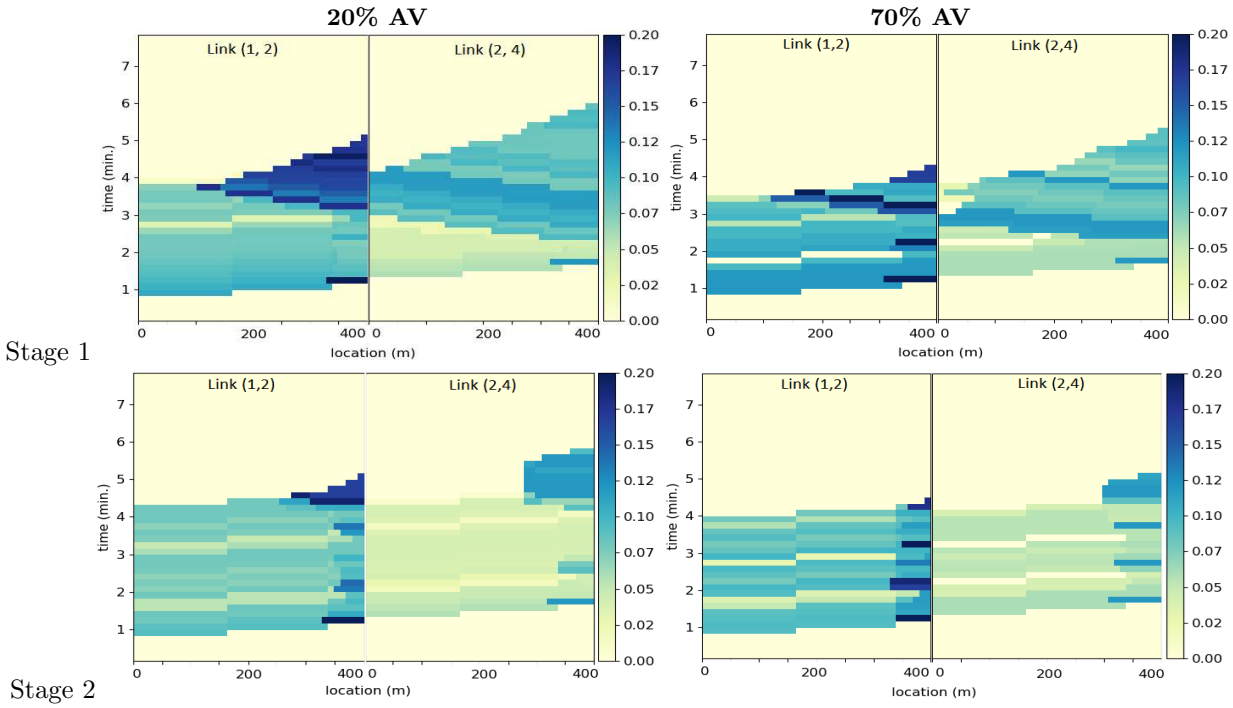


Figure 7: Time-space density at links (1,2) and (2,4) in the Braess network (case 1).

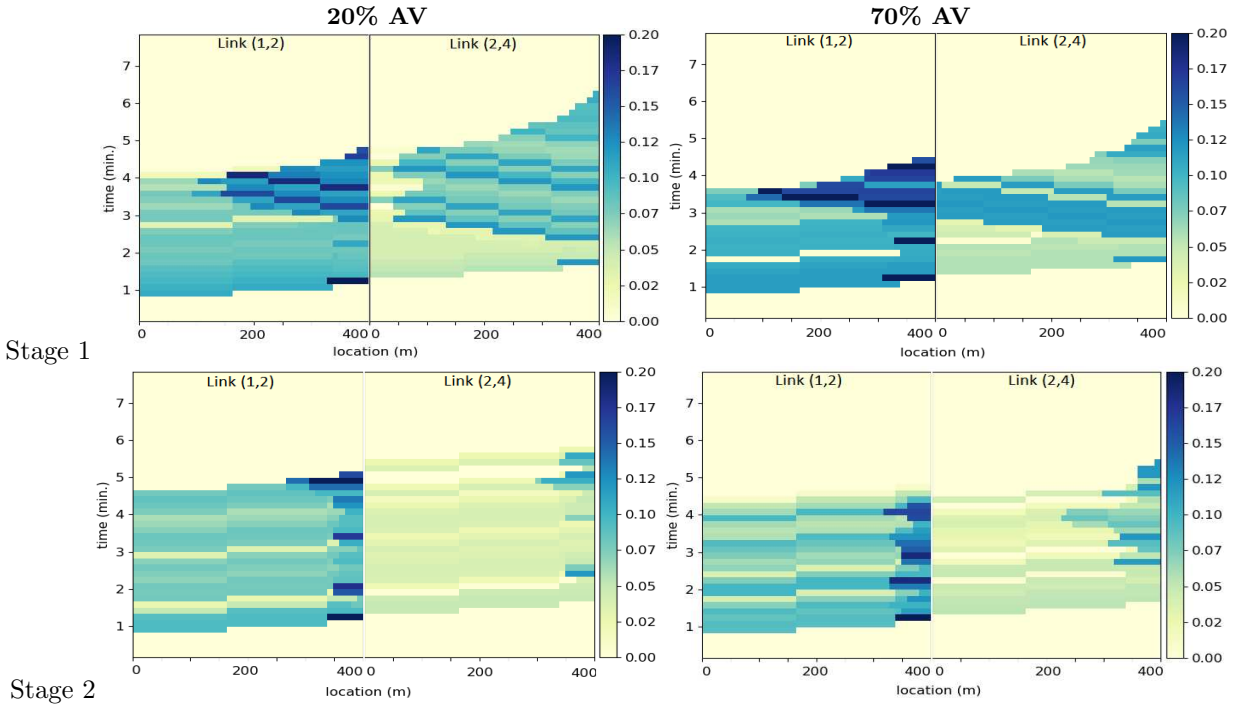


Figure 8: Time-space density at links (1,2) and (2,4) in the Braess network (case 2).

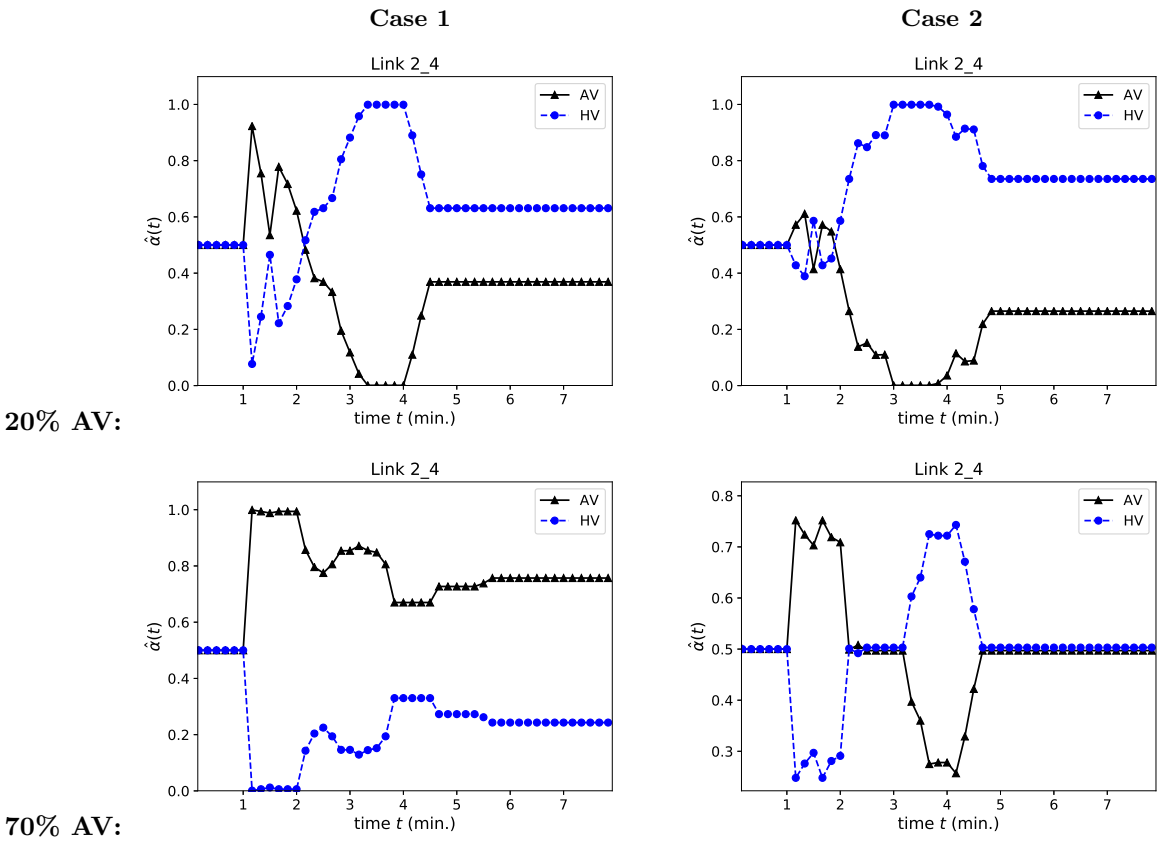


Figure 9: The solution of $\hat{\alpha}(t)$ at link (2,4) in the Braess network (case 1: left figure, case 2: right figure).

Furthermore, Figure 9 show the corresponding result of time varying share factor $\hat{\alpha}(t)$ at link (2, 4). It is clearly shown that $\hat{\alpha}^{HV}(t)$ is higher in case 2 than in case 1, therefore it enables the reduction of HVs' travel time observed in case 2.

### *7.2. Sioux Falls network*

In this section, we will evaluate the computational performance of our framework by using the Sioux Falls network in Figure 10, where the network settings (i.e., link length, number of lanes, and link jam density) in Bar-Gera (2002) are adopted. The parameters used in this case are similar to Table 2. The total demand (per minute) of both AVs and HVs is 333.34veh/min, the time domain is one hour with 3-min interval. Similar to the previous example, the increased AV penetration helps to improve the dissipation of congestion (e.g., at link (22, 20) in Figure 12) and reduce the total system travel time (Figure 11) where the AVs gain the most benefit over the HVs.

*Benefit of AVs in the congested network.* In the previous example of a small network, we observe the benefit of AVs in reducing significantly the total system travel time (TSTT). In fact, we are able to obtain the same result in the Sioux Falls network with a very high demand of traffic so that traffic congestion occurs, i.e., demand of 8000 vehicles in total, see Figure 11(a). For a smaller demand, that is the network is not congested, the effect of the increased AV penetration on the TSTT is not significant as seen in Figure 11(b). It is worth noticing that the AVs' travel time also increases due to the higher AV penetration (i.e. higher AV demand).

Figure 12 describes the spatio-temporal evolution of the total density in the congested link (22, 20) for different initial penetrations of AVs. It is clear that a higher penetration of AVs leads to less traffic

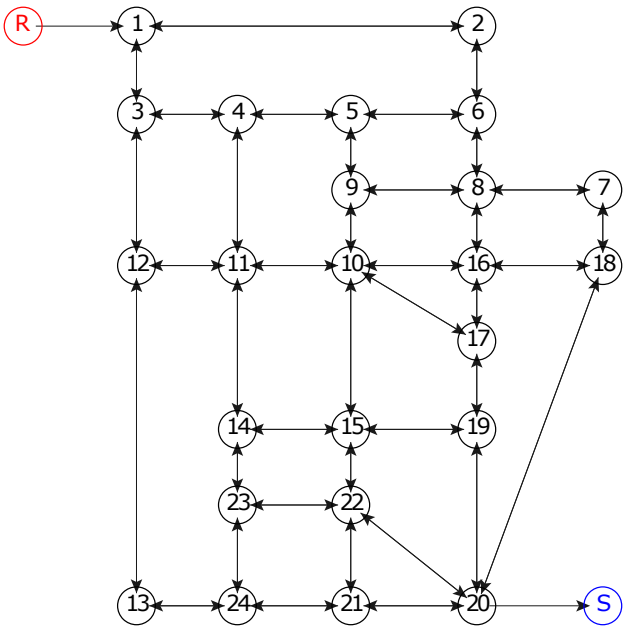


Figure 10: Sioux-Fall network.

(a) Demand = 8000 vehicles

(b) Demand = 2000 vehicles

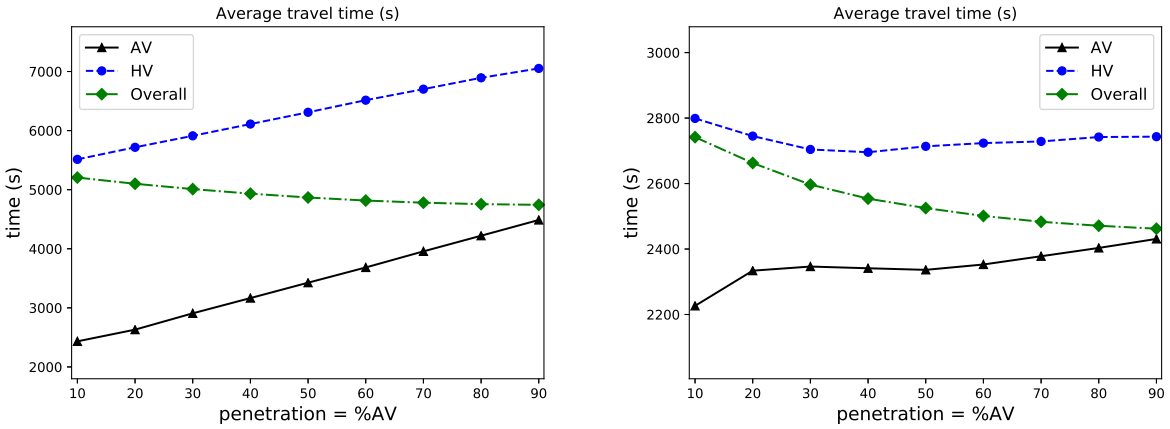


Figure 11: Travel time in Sioux Fall network with different penetrations of AV.

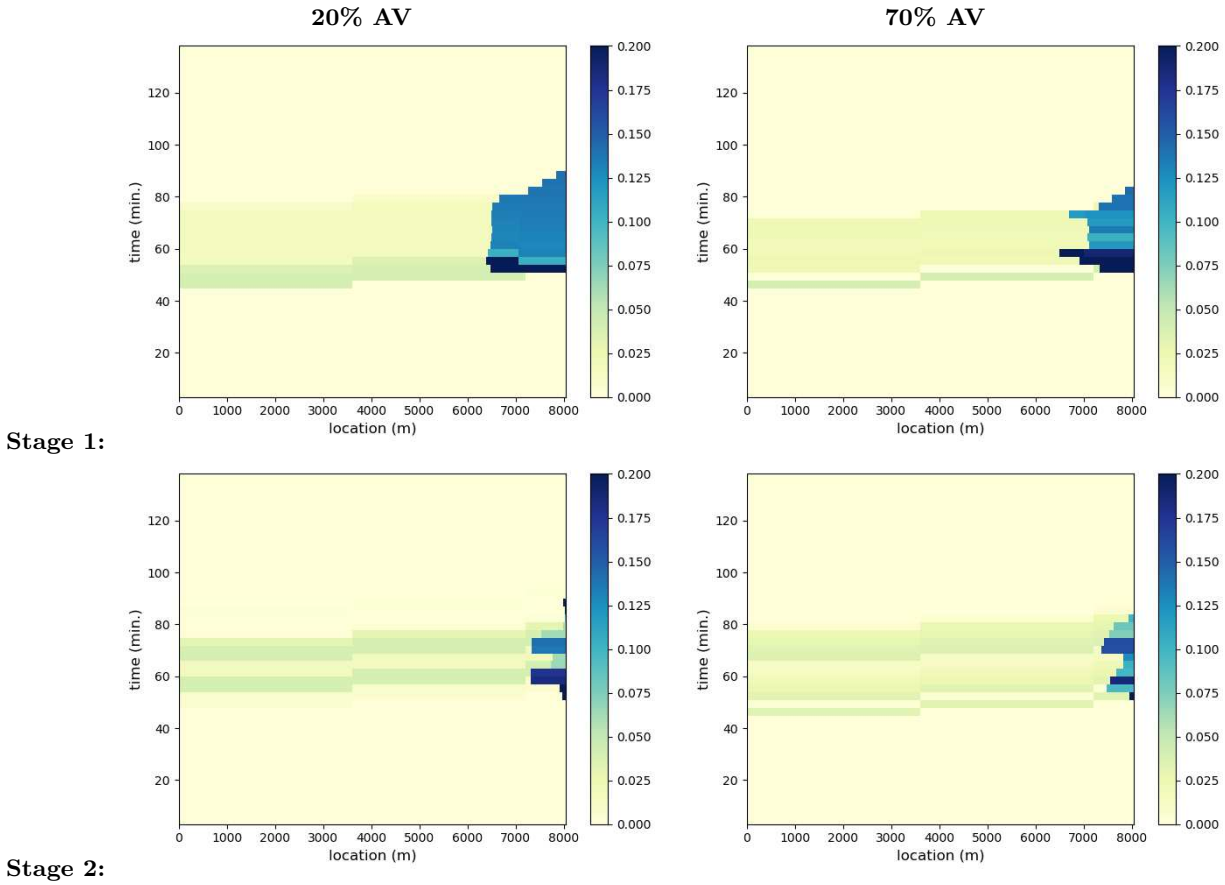


Figure 12: Time-space vehicle density at link (22, 20) in Sioux Falls network, demand = 4000 vehicles.

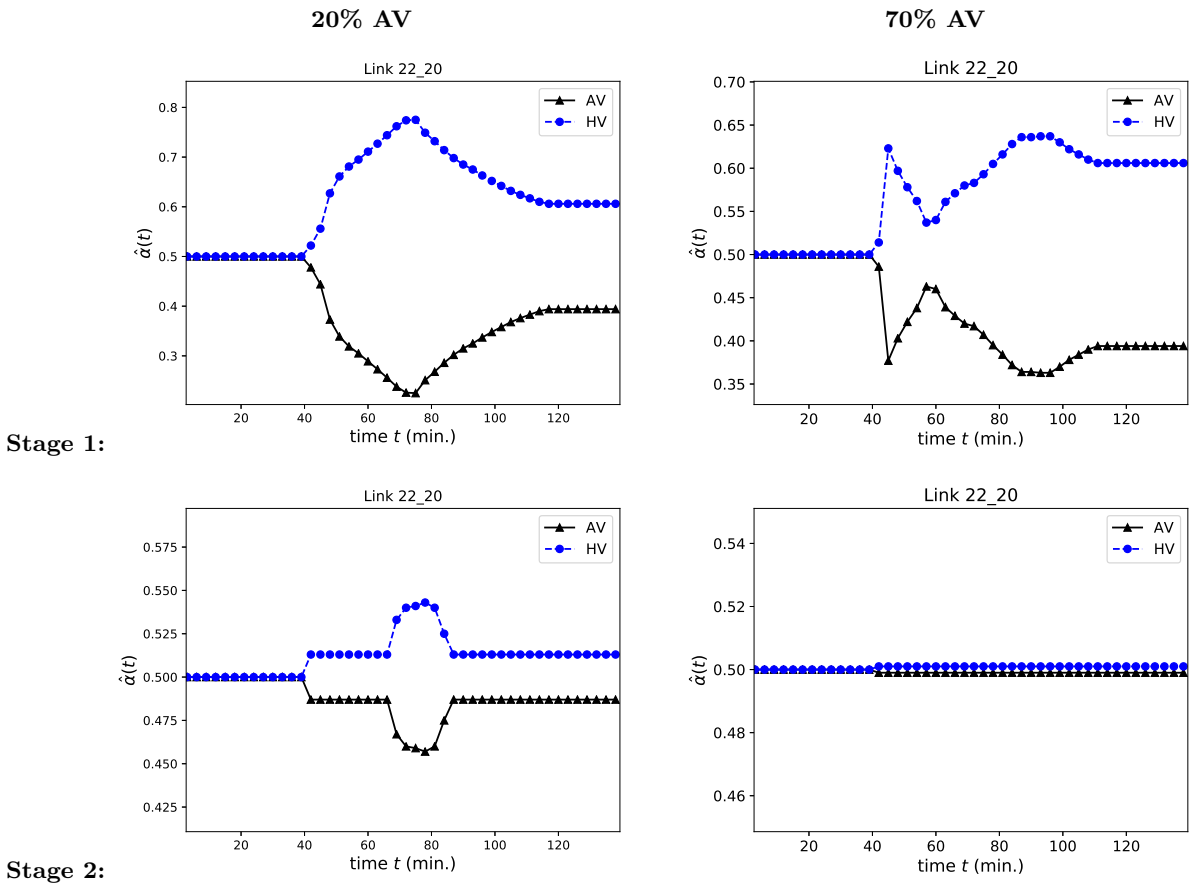


Figure 13: The value of $\hat{\alpha}(t)$ at link (22, 20) in Sioux Falls network, demand = 4000 vehicles.

| Setting | Stage 1 computing time (min.) | Stage 2 computing time (min.) |
|---|---|---|
| 10% AV | 5.0 | 50.8 |
| 30% AV | 4.0 | 40.1 |
| 50% AV | 2.4 | 61.7 |
| 70% AV | 1.9 | 40.1 |
| 90% AV | 1.1 | 46.7 |

Table 3: Computational demand using the Sioux Falls network with the total demand of 8000 vehicles.

| Demand (veh) | Time $T$ (min) | Stage 1 computing time (min.) |
|---|---|---|
| 2000 | 66 | 0.1 |
| 4000 | 96 | 0.4 |
| 6000 | 120 | 1.5 |
| 8000 | 150 | 3.1 |

Table 4: Computational demand of Stage 1 using the Sioux Falls network.

congestion. Moreover, the DSO solution with an optimal distribution of the physical queues (i.e. Stage 2 solution) tends to reduce the congestion level along the link.

Additionally, Figure 13 presents the time varying share factor $\hat{\alpha}^m(t)$ at the link (22, 20) that complements with Figure 12. It clearly shows that AVs use much of the link capacity in the high initial penetration of AVs while HVs share a portion of link capacity with AVs in the low initial penetration of AVs. This again indicates how the link capacity is dynamically allocated to different vehicle classes in order to optimize the total system travel costs. Furthermore, in terms of optimal queue distribution (i.e. at Stage 2), AVs tend to be allocated longer (in time) and larger value of $\hat{\alpha}^{AV}(t)$ in the comparison with the solution at Stage 1.

*Computational performance.* In terms of computational performance, Table 3 shows that Stage 1 performs much faster than Stage 2, i.e., about 1/10, in our proposed framework. Note that, the model in Stage 1 is less complex than Stage 2 due to the linear objective function $F$. For this reason, the result of minimizing the queue distribution takes a considerable amount of time. However, due to the fast computation of solving Stage 1, we are able to evaluate quickly the optimal system performance in the different scenarios (e.g., with and without EATT, with different class specific traffic demands) for planing purposes.

To further evaluate the first Stage, we measure its computing time by increasing the OD traffic demand with the result shown in Table 4. By increasing the demand, we also need to increase linearly the time domain to clear all the corresponding demand. However, the computational time (of Stage 1) increases dramatically.

In summary, we demonstrate the DSO solution for mixed human-driven and automated vehicles via our examples and show the benefit of increasing AV penetration to improve the system performance, in terms of both total system travel time and optimal physical queue distribution. Our numerical results also show how the link capacity is dynamically allocated to different vehicles classes in order to optimize the total system travel costs and the total physical queue distribution. We also present the impact of traffic demand on both average travel time for each vehicle type and on the computational demand of each Stage in our framework.

## 8. Concluding remarks

This paper explored the impact of AVs on traffic network performance using a multi-class system optimum dynamic traffic assignment (SO-DTA) approach. We derived the network level side constraints for the mixed AVs and HVs vehicles taking into account the variations in link capacities in response to the dynamics of the vehicle class proportions. More specially, our proposed model allows to obtain a DSO solution that optimally distributes the physical queues in the network considering the dynamics of the AV share factor. The possible impact of the AVs on the road capacity under different initial penetration rates was then numerically investigated. The simulation results indicated that the introduction of AVs changes the traffic

flow dynamics, reducing the spillback with the reduction in the total system travel time. Our results also demonstrated that the reduction in the total system travel time with increased initial penetrations of AVs is more profound with increased total demands, when heavy traffic congestion occurs.

The findings in this paper have important implications for traffic management and open up various new research directions, such as optimal dynamic road/lane allocations and priority signal control for specific vehicle classes. Although the proposed framework currently only applies for the single origin-destination network with two specific vehicle classes, it plays an important role in the state-of-the-art and paves a way for future work covering more generic scenario of multi-class networks and traffic demands (e.g. with more than two vehicle classes such as HVs, AVs and trucks). Our future work is to extend the proposed framework to the mixed DSO-DUE case, in which the AVs are controlled to follow the SO principle while the HVs pertain to the UE principle. In this mixed case, measures such as pricing may be used to influence the HVs towards a better solution for the system; the model considered in the present paper, of DSO for both HVs and AVs, will then serve as an important benchmark as the maximum that might be achieved from such pricing measures, as we move towards full marginal cost pricing.

## Appendix A. Proof of proposition 1

*Proof.* From equation (14) we obtain:

$$\frac{q_m(x,t)}{V} + \frac{q_m(x,t)}{W(t)} \leq K\alpha_m(x,t) \tag{A.1}$$

According to assumption 1:

$$q_m(x,t) = \begin{cases} u_m\left(t - \dfrac{x}{V}\right) & \text{in the free-flow regime} \\ v_m\left(t - \dfrac{l-x}{W(t)}\right) & \text{in the congested regime} \end{cases} \tag{A.2}$$

Then inequality (A.1) is converted into the following condition:

$$\frac{u_m(t - \frac{x}{V})}{V} + \frac{v_m(t - \frac{l-x}{W(t)})}{W(t)} \leq K\alpha_m(x,t) \tag{A.3}$$

$$\text{in which, by definition: } \sum_{m\in\mathbb{M}} \alpha_m(x,t) = 1 \tag{A.4}$$

This inequality implies that:

$$K\alpha_m(l-x,t) - \frac{1}{W(t)} v_m\left(t - \frac{x}{W(t)}\right) \geq \frac{1}{V} u_m\left(t - \frac{l-x}{V}\right).$$

which is substituted to equation (17) to obtain:

$$\begin{aligned} N_m(t) &\geq \frac{1}{V}\int_0^{l_m^f(t)} u_m\left(t - \frac{x}{V}\right)\,\mathrm{d}x + \frac{1}{V}\int_0^{l_m^c(t)} u_m\left(t - \frac{l-x}{V}\right)\,\mathrm{d}x \\ &= \frac{1}{V}\int_0^{l} u_m\left(t - \frac{x}{V}\right)\,\mathrm{d}x \end{aligned}$$

which is converted to the discrete time as:

$$N_m(i) \geq \sum_{k=i-\frac{l}{V}+1}^{i} u_m(k).$$

□

## Appendix B. Proof of proposition 2

*Proof.* Similarly, inequality (A.3) implies that

$$\frac{1}{V}u_m\left(t-\frac{x}{V}\right) \leq K\alpha_m(x,t) - \frac{1}{W(t)}v_m\left(t-\frac{l-x}{W(t)}\right).$$

which is substituted to equation (17) to obtain:

$$\begin{aligned} N_m(t) &\leq \int_0^{l_m^f(t)} \left[K\alpha_m(x,t) - \frac{1}{W(i)}v_m\left(t-\frac{l-x}{W(t)}\right)\right] \,\mathrm{d}x + \int_0^{l_m^c(t)} \left[K\alpha_m(l-x,t) - \frac{1}{W(t)}v_m\left(t-\frac{x}{W(t)}\right)\right] \,\mathrm{d}x \\ &= -\int_0^l \frac{1}{W(t)}v_m\left(t-\frac{x}{W(t)}\right) \,\mathrm{d}x + \int_0^l K\alpha_m(x,t) \,\mathrm{d}x \end{aligned}$$

which is converted to the discrete time as:

$$N_m(i) \leq Kl\hat{\alpha}_m(i) - \sum_{k=i-\frac{l}{W(i)}+1}^{i} v_m(k).$$

□

## Appendix C. Proof of Proposition 3

*Proof.* First, we have $\frac{d\hat{T}(\alpha)}{d\alpha} = 2\alpha(T_{AA}-T_{AH}) + (T_{AH}-T_H) < 0$ since $T_{AA} < T_{AH} < T_H$. Hence $\frac{dW(\alpha)}{d\alpha} = -\frac{1}{K\hat{T}^2}\frac{d\hat{T}(\alpha)}{d\alpha} > 0$. Moreover, $\frac{d^2\hat{T}(\alpha)}{d\alpha^2} = 2(T_{AA}-T_{AH}) < 0$. Let us compute the derivatives of $Q(\alpha)$ w.r.t. $\alpha$:

$$\frac{dQ(\alpha)}{d\alpha} = \frac{KV^2}{(V+W)^2}\frac{dW}{d\alpha} > 0 \tag{C.1}$$

and

$$\begin{aligned} \frac{d^2Q(\alpha)}{d\alpha^2} &= \frac{KV^2}{(V+W)^2}\left[\frac{d^2W(\alpha)}{d\alpha^2} - \frac{2}{(V+W)}\left(\frac{dW}{d\alpha}\right)^2\right] \\ &= \frac{V^2}{\hat{T}^2(V+W)^2}\left[-\frac{d^2\hat{T}}{d\alpha^2} + \frac{2V}{\hat{T}(V+W)}\left(\frac{d\hat{T}}{d\alpha}\right)^2\right] > 0 \end{aligned} \tag{C.2}$$

so $Q(\alpha)$ is a convex function of $\alpha$. □

## Appendix D. Proof of Proposition 4

Assume that the optimal solution of problem $\mathcal{P}$ holds traffic from link $a$ to link $b$ at time $i$ in the network. According to the condition Eq. (51), all the conditions are not met, i.e.,

$$\begin{aligned} N_{am}(i) - \sum_{k=i+1-\frac{l_a}{V_a}}^{i} u_{am}(k) > 0 \\ K_b l_b \hat{\alpha}_{bm}(i) - N_{bm}(i) - \sum_{k=i+1-\frac{l_b}{W_b(i)}}^{i} v_{bm}(k) > 0 \\ Q_a(i) - \sum_m v_{am}(i) > 0 \\ Q_b(i) - \sum_m u_{bm}(i) > 0 \\ \sum_{k \geq i+1} f_{abm}(k) > 0. \end{aligned}$$

All these conditions guarantee that we can take a small traffic $\epsilon$ in $f_{abm}(k) > 0$ $(k > i)$and put into $f_{abm}(i)$ without violating any constraints. Therefore, we achieve a new solution that is identical to the optimal one except for

$$\begin{aligned} u_{am}^{new}(i) = u_{am}^{old}(i) + \epsilon \\ u_{am}^{new}(k) = u_{am}^{old}(k) - \epsilon. \end{aligned}$$

As the result, we achieve a new value of $H$ which larger than the optimal one:

$$H^{new} = H^{optimal} + \epsilon(k - i) > H^{optimal}$$

where $\epsilon > 0, k > i$. Therefore, the solution of problem $\mathcal{P}$ does not hold traffic at any link in the network.

## Appendix E. Minimising total travel time with the objective function $F$

Let $f_{pm}(i, k)$ be the amount of class-$m$ traffic on path $p$ departing the source at time $i$ and arriving the sink at time $k$. Let $\mathbb{P}$ be the set of all path in the network. The source link is denoted as link $r$ and the sink link is denoted as link $s$. Remind that we denote $U_{rm}(i)$ as the demand of class-$m$ traffic from the source $r$ at time $i$. Therefore, we can infer that:

$$\sum_{p \in \mathbb{P}} \sum_{k > i} f_{pm}(i, k) = U_{rm}(i) \qquad \forall i \in \mathbb{T}.$$

The total system travel time (TSTT) for class-$m$ vehicle is computed by the following equation:

$$\begin{aligned} \text{TSTT}_m &= \sum_{p \in \mathbb{P}} \sum_{i \in \mathbb{T}} \sum_{k > i} (k - i) f_{pm}(i, k) = \sum_{p \in \mathbb{P}} \sum_{i \in \mathbb{T}} \sum_{k > i} k \ f_{pm}(i, k) - \sum_{p \in \mathbb{P}} \sum_{i \in \mathbb{T}} \sum_{k > i} i \ f_{pm}(i, k) \\ &= \sum_{p \in \mathbb{P}} \sum_{i \in \mathbb{T}} \sum_{k > i} k \ f_{pm}(i, k) - \sum_{i \in \mathbb{T}} i \ U_{rm}(i). \end{aligned}$$

Due to $\sum_{p\in\mathbb{P}}\sum_{i\in\mathbb{T}} f_{pm}(i,k) = u_{sm}(k)$ (where $s$ is the sink link in path $p$), we can continue inferring that:

$$\begin{aligned}\text{TSTT}_m &= \sum_{i\in\mathbb{T}} i\ u_{sm}(i) - \sum_{i\in\mathbb{T}} i\ U_{rm}(i)\\ &= -\sum_{i\in\mathbb{T}}(T+1-i)u_{sm}(i) + \sum_{i\in\mathbb{T}}(T+1)u_{sm}(i) - \sum_{i\in\mathbb{T}} i\ U_{rm}(i)\\ &= \underbrace{\sum_{i\in\mathbb{T}}(T+1-i)U_{rm}(i)}_{\text{Constant}} - \sum_{i\in\mathbb{T}}(T+1-i)u_{sm}(i).\end{aligned}$$

Note that, above transformation is based on the conservation of travel demand at sources and destinations that $\sum_{i\in\mathbb{T}} u_{sm}(i) = \sum_{i\in\mathbb{T}} U_{rm}(i)$ where $(r,s)$ is the pair of source and sink links. The above equation is true for any O-D pair and any class $m$. Therefore, minimizing the total travel time is equivalent to maximizing the objective function $F$ (i.e., aggregation of total flow at destinations).

## Appendix F. Justification of the two-stage framework in the DSO problem

Since the distribution of traffic in the network (i.e., nodes and links between sources and destinations) can form different traffic patterns, Stage 2 aims to select the one which balances the queue among those links, and consequently helps slow down the propagation of the spillback. In the following, we show an example that different patterns (or solutions) can be observed, one with more desirable properties compared to the other. Let's consider a simple network with 4 links as shown in Figure F.14 where the network setting is shown in Table F.5. It is clearly that there are two identical paths from $R$ to $S$ with the same free-flow travel time. The traffic demand (with a total of 1500 vehicles) from $R$ to $S$ runs at full capacity of link (0,1), however, the bottle-neck link $(2,3)$ cannot get all traffic from two path at the full capacity. Therefore, the congestion will happen in one or both paths and cause the traffic congestion to spill back from node 2 to node 1. There are two paths called $a$ and $b$ as shown in Figure F.14 (in this example, by mentioning path $a$ (or $b$) we mean the path from R to S passing through link $a$ (or $b$)).

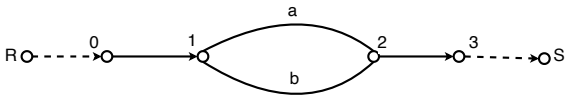


Figure F.14: The simple network.

Table F.5: The network setting.

| Link(s) | Length (m) | #lanes | $V$ (veh/min.) | $W$ (veh/min.) (AV) | $W$ (veh/min.) (CV) | $K$ (veh/m per lane) |
|---|---|---|---|---|---|---|
| 0-1 | 400 | 8 | 1000 | 600 | 400 | 0.12 |
| 1-2 (a, b) | 800 | 4 | 1000 | 600 | 400 | 0.12 |
| 2-3 | 400 | 5 | 1000 | 600 | 400 | 0.12 |

Figure F.15 shows the difference in the number of vehicles staying in the network in the comparison between Stage 1 and Stage 2 traffic patterns in different settings of AV penetration ($\alpha$). In this figure, the positive value means that more (Stage 2) vehicles are in the network than Stage 1 vehicles at any point in time. Note that, these two traffic patterns give the same DSO objective value (i.e. the same total travel time). However, because the spillback at node 1 happens quicker in Stage 1 (i.e. the tail of the congestion

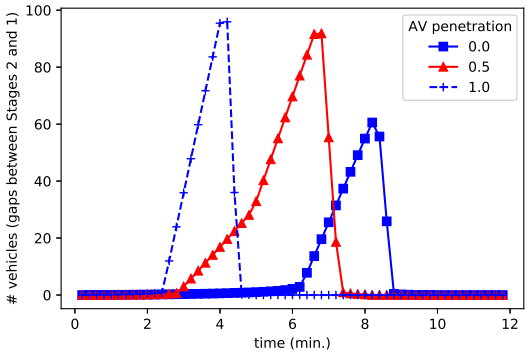


Figure F.15: The difference of vehicles in the comparison between the Stage 2 and Stage 1 solutions.

reaches node 1), the flow to the network reduces quicker to the bottle-neck flow (at link (2,3)) in this stage. Particularly, the two traffic patterns shown in Figure F.15 are described below:

- (Stage 1) All traffic flow on path $a$ passes through node 2 without delay where the remaining capacity on link (2,S) is for path $b$.
- (Stage 2) Traffic splits equally at node 1 and merges equally at node 2.

These two traffic patterns fully utilize the capacity at node 2 (therefore they are DSO solutions) while they are different in splitting and merging traffic at nodes 1 and 2.

In the following, we show that different congestion patterns can lead to different traffic states in terms of network occupancy and traffic flow. For any link $a$, let $u_{a,t}$ and $v_{a,t}$ denote the (class-aggregated) traffic flow at the upstream and downstream of link $a$. The dynamic number of vehicles in link $a$, defined as $N_{a,t}$, is equal to

$$N_{a,t} = \sum_{h \le t} (u_{a,h} - v_{a,h}) .$$

Both stages above share the same traffic patterns when the congestion is still in the middle of link (1,2) (either on path $a$ or $b$). We will investigate how long the congestion (or queue) reaches the node 1 and causes the spillback on link (0,1). Let $T_V = \dfrac{l_a}{V_a}$ be the free-flow travel time and $T_W = \dfrac{l_a}{W_a}$ be the travel time of backward shockwave from the downstream. Let $t_s$ be the time that the spillback happens at the upstream of link $a$ (meaning that the whole link is in congested regime from time $t_s$). Assume that the inflow to link $a$ is a constant $u$ and the outflow capacity of link $a$ is $v$. Due to the fact that the value of $v$ depends on the available flow capacity of the downstream link(s) of $a$, it can happen that $u \ge v$ leading to the congestion on this link. In this typical condition, we can observe three periods in link $a$ as follows:

- $t \le T_V$: $u_{a,t} = u$, $v_{a,t} = 0$, $N_{a,t} = ut$.
- $T_V + 1 \le t \le t_s$: $u_{a,t} = u$, $v_{a,t} = v$, $N_{a,t} = T_V u + (t - T_V)(u - v)$.
- $t \ge t_s + 1$: $u_{a,t} = v_{a,t} = v$, $N_{a,t} = T_V u + (t_s - T_V)(u - v)$.

This clearly shows that larger value of $t_s$ can lead to more flow and more vehicles staying in the network.

For any $T_V + 1 \leq t \leq t_s$, according to Eq. (36) regarding the shockwave propagation, we can derive that

$$K_a l_a \geq \sum_{h \leq t} u_{a,h} - \sum_{h \leq t - T_W} v_{a,h} = ut \Rightarrow t \leq \frac{K_a l_a}{u} \qquad t \leq T_V + T_W$$

$$K_a l_a \geq \sum_{h \leq t} u_{a,h} - \sum_{h \leq t - T_W} v_{a,h} = ut - (t - T_V - T_W)v \Rightarrow t \leq \frac{K_a l_a - v(T_V + T_W)}{u - v} \qquad t \geq T_V + T_W$$

With the above inference, it can be easily proved that

- If $u = \dfrac{K_a V_a W_a}{V_a + W_a} = \dfrac{K_a l_a}{T_V + T_W}$ then $t_s = T_V + T_W$ (spillback is independent on outflow $v$).
- If $v < u < \dfrac{K_a V_a W_a}{V_a + W_a}$ then $t_s = \dfrac{K_a l_a - v(T_V + T_W)}{u - v}$ (spillback is dependent on both inflow $u$ and outflow $v$)

If $u \leq v$, no queue can happen or grow in this link, therefore $t_s$ is undefined in this case. The above discussion shows quantitatively the dependence of spillback on the inflow $u$ and outflow $v$. It clearly indicates that different DSO traffic patterns (i.e., different values of $u$ and $v$) create different patterns of spillback reflected by the number of vehicles accommodated in the network as shown in Figure F.15.

To further describe how the two-stage framework works, in the following part, we show that Stage 2 traffic pattern actually minimizes the spreading of spillback. Let $u_a, v_a, u_b, v_b$ be the inflow and outflow of links $a$ and $b$ (from node 1 to node 2) respectively. Both links $a$ and $b$ have the same traffic jam density $K$, length $l$, free-flow speed $V$ and backward shockwave speed $W$. According to the analytical study of spillback in the example above, the spillback happens on these links at $t_s^a$ and $t_s^b$, respectively, as follows.

$$t_s^a = \frac{Kl - v_a(T_V + T_W)}{u_a - v_a} \tag{F.1}$$

$$t_s^b = \frac{Kl - v_b(T_V + T_W)}{u_b - v_b} \tag{F.2}$$

where $v_a < u_a < \dfrac{KVW}{V + W}$ and $v_a < u_a < \dfrac{KVW}{V + W}$. We also have the conservation of travel demand $D$ where $u_a + u_b = D$ and bottleneck capacity as $v_a + v_b = Q$. Let $\bar{t}_s = \min(t_s^a, t_s^b)$ be the time when the shockwave reaches the upstream node (i.e., node 1 in the example). Therefore, to reduce (or slow down) the spillback, we need to maximize the value of $\bar{t}_s$. Because this optimization problem is symmetric, the optimal solution happens when $u_a = u_b = \dfrac{D}{2}$ and $v_a = v_b = \dfrac{Q}{2}$, which also balances queues in both paths (Stage 2 traffic pattern).

## References


Bagloee, S., Sarvi, M., Patriksson, M., Rajabifard, A., 2017. A mixed userequilibrium and systemoptimal traffic flow for connected vehicles stated as a complementarity problem. Computer Aided Civil and Infrastructure Engineering 23, 562–580.

Ban, X., Pang, J. S., Liu, X., Ma, R., 2012. Modeling and solution of continuous-time instantaneous dynamic user equilibria: A differential complementarity systems approach. Transportation Research Part B 46, 389–408.

Bar-Gera, H., 2002. Transportation network test problems.
URL `http://www.bgu. ac.il/ bargera/tntp/`

Biegler, L. T., Zavala, V. M., 2009. Large-scale nonlinear programming using ipopt: An integrating framework for enterprise-wide dynamic optimization. Computers & Chemical Engineering 33 (3), 575–582.

Bretti, G., Natalini, R., Piccoli, B., 2006. Fast algorithms for the approximation of a traffic flow model on networks. Discrete and Continuous Dynamical Systems, Series B 6, 427–448.

Chen, Z., He, F., Zhang, L., Y., Y., 2016. Optimal deployment of autonmous vehicle lanes with endogenous market penetration. Transportation Research Part C 72, 143–156.

Dafermos, S., 1972. The traffic assignment problem for Multiclass User Transportation Networks. Transportation Science 6, 73–87.
Daganzo, C. F., 1995. The cell transmission model, part 2: network traffic. Transportation Research Part B 29 (2), 79–93.
Delis, A. I., Nikolos, I. K., Papageorgiou, M., 2015. Macroscopic traffic flow modeling with adaptive cruise control: Development and numerical solution. Computers and Mathematics with Applications 70, 1921–1947.
Gentile, G., Meschini, L., Papola, N., 2007. Spillback congestion in dynamic traffic assignment: A macroscopic flow model with time-varying bottlenecks. Transportation Research Part B 41, 1114–1138.
Ghiasi, A., Li, X., Ma, J., 2019. A mixed traffic speed harmonization model with connected autonomous vehicles. Transportation Research Part C 104, 210–233.
Han, K., Piccoli, B., Szeto, W., 2016. Continuous-time link-based kinematic wave model: formulation, solution existence, and well-posedness. Transportmetrica B 4, 187–222.
Han, K., Piccoli, B., Szeto, W. Y., 2015. Continuous-time link-based kinematic wave model: formulation, solution existence, and well-posedness. Transportmetrica B:Transport Dynamics, "DOI: 10.1080/21680566.2015.1064793".
Jia, D., Ngoduy, D., 2016a. Enhanced cooperative car-following traffic model with the combination of V2V and V2I communication. Transportation Research Part B 90, 172–191.
Jia, D., Ngoduy, D., 2016b. Platoon based Cooperative Driving Model with Consideration of Realistic Inter-vehicle Communication. Transportation Research Part C 68, 245–264.
Jia, D., Ngoduy, D., Vu, H., 2019. A multi-class microscopic model for heterogeneous platoon with vehicle-to-vehicle communication. Transportmetrica B 7, 448–472.
Jiang, Y., Szeto, W., Long, J., Han, K., 2016. Multi-class dynamic traffic assignment with physical queues: intersection-movement-based formulation and paradox. Transportmetrica A: Transport Science 12, 878–908.
Jin, W. L., 2015. Continuous formulations and analytical properties of the link transmission model. Transportation Research Part B 74, 88–103.
Kesting, A., Treiber, M., Helbing, D., 2010. Enhanced Intelligent Driver Model to Access the Impact of Driving Strategies on Traffic Capacity. Philosophical Transactions of the Royal Society A 368, 4585–4605.
Levin, M., Boyles, S., 2016a. A cell transmission model for dynamic lane reversal with autonomous vehicles. Transportation Research Part C 68, 126–143.
Levin, M., Boyles, S., 2016b. A multi-class cell transmission model for shared human and autonomous vehicle roads. Transportation Research Part C 62, 103–116.
Lighthill, M. H., Whitham, G. B., 1955. On kinematic waves 2: A theory of traffic flow on long, crowded roads. In: Proceedings of the Royal Society of London. A 229. pp. 317–345.
Lipp, T., Boyd, S., 2016. Variations and extension of the convex–concave procedure. Optimization and Engineering 17 (2), 263–287.
Liu, B., Jia, D., Lu, K., Ngoduy, D., Wang, J., Wu, L., 2017. A Joint Control-Communication Design for Reliable Vehicle Platooning in Hybrid Traffic. IEEE Transactions on Vehicular Technology 66, 9394–9409.
Lo, H. K., Szeto, W. Y., 2002. A cell-based variational inequality formulation of the dynamic user optimal assignment problem. Transportation Research Part B 36, 421–443.
Long, J., Chen, J., Szeto, W., Shi, Q., 2018. Link-based system optimum dynamic traffic assignment problems with environmental objectives. Transportation Research Part D 60, 56–75.
Ma, R., Ban, X., Pang, J. S., 2014. Continuous-time dynamic system optimal for single-destination traffic networks with queue spillbacks. Transportation Research Part B 68, 98–122.
Melson, C., Levin, M., Hammit, B., Boyles, S., 2018. Dynamic traffic assignment of cooperative adaptive cruise control. Transportation Research Part C 90, 114–133.
Mohajerpoor, R., Ramezani, M., 2019. Mixed flow of autonomous and human-driven vehicles: Analytical headway modeling and optimal lane management. Transportation Research Part C 109, 194–210.
Newell, G. F., 1993. A simplified theory of kinematic waves in highway traffic (part i, ii, iii). Transportation Research Part B: Methodological 27 (4), 281–313.
Ngoduy, D., 2012. Application of gas-kinetic theory to modelling mixed traffic of manual and adaptive cruise control vehicles. Transportmetrica A: Transport Science 8, 43–60.
Ngoduy, D., 2013a. Analytical studies on the instabilities of heterogeneous intelligent traffic flow. Communications in Nonlinear Science and Numerical Simulation 18, 2699–2706.
Ngoduy, D., 2013b. Instabilities of cooperative adaptive cruise control traffic flow: a macroscopic approach. Communications in Nonlinear Science and Numerical Simulation 18, 2838–2851.
Ngoduy, D., 2013c. Platoon-based macroscopic model for intelligent traffic flow. Transportmetrica B: Transport Dynamics 1, 153–169.
Ngoduy, D., Hoang, N., Vu, H., Watling, D., 2016. Optimal queue placement in dynamic system optimum solutions for single origin-destination traffic networks. Transportation Research Part B 92, 148–169.
Osorio, C., Flotterod, G., 2014. Capturing dependency among link boundaries in a stochastic dynamic network loading model. Transportation Science 49, 420–431.
Osorio, C., Flotterod, G., Bierlaire, M., 2011. Dynamic network loading: a stochastic differentiable model that derives link state distributions. Transportation Research Part B 45, 1410–1423.
Sharma, A., Zheng, Z., Bhaskar, A., Haque, M., 2019. Modelling car-following behaviour of connected vehicles with a focus on driver compliance. Transportation Research Part B 126, 256–279.
Shen, W., Nie, Y., Zhang, H. M., 2007. A dynamic network simplex method for designing emergency evacuation plans. Transportation Research Record 2022, 83–93.

Shen, W., Zhang, H. M., 2008. What do different traffic flow models mean for system-optimal dynamic traffic assignment in a many-to-one network? Transportation Research Record 2008, 157–166.
Shen, W., Zhang, H. M., 2014. System optimal dynamic traffic assignment: Properties and solution procedures in the case of a many-to-one network. Transportation Research Part B 65 (10), 1–17.
Shladover, S. amd Su, D., Lu, X., 2012. Impacts of cooperative adaptive cruise control on freeway traffic flow. Transportation Research Record 2324, 63–70.
Szeto, W., Jiang, Y., Sumalee, A., 2011. A cellbased model for multiclass doubly stochastic dynamic traffic assignment. Computer Aided Civil and Infrastructure Engineering 26, 595–611.
Treiber, M., Kesting, A., Helbing, D., 2006. Delays, inaccuracies and anticipation in microscopic traffic model. Physica A 360, 71–88.
van Vuren, T., Watling, D., 1991. Multiple user class assignment model for route guidance. Transportation Research Record 1306, 22–32.
Wang, J., Peeta, S., He, X., 2019. Multiclass traffic assignment model for mixed traffic flow of human-driven vehicles and connected and autonomous vehicles. Transportation Research Part B 126, 139–168.
Wang, M., Hoogendoorn, S., Daamen, W., Arem, B., Happee, R., 2015. Game theoretic approach for predictive lane-changing and car-following control. Transportation Research Part C 58, 73–92.
Xie, D., Fang, Z., Jia, B., He, Z., 2019. A data-driven lane-changing model based on deep learning. Transportation Research Part C 106, 41–60.
Ye, L., Yamamoto, T., 2018. Modelling connected autononous vehicles in heterogeneous traffic flow. Physica A 490, 269–277.
Ye, L., Yamamoto, T., 2019. Evaluating the impact of connected and autonomous vehicles on traffic safety. Physica A 526, 121009.
Yperman, I., Logghe, S., Immers, L. H., Tampere, C., 2005. Multicommodity link transmission model for dynamic network loading. In: The 85th Annual Meeting of the Transportation Research Board. Washington, D.C.
Zhang, X., Sun, J., Qi, X., Sun, J., 2019. Simultaneous modeling of car-following and lane-changing behaviors using deep learning. Transportation Research Part C 104, 287–304.
Zhao, W., Ngoduy, D., Shepherd, S., Liu, R., Papageorgiou, M., 2018. A platoon based cooperative eco-driving model for mixed automated and human-driven vehicles at a signalised intersection. Transportation Research Part C 95, 802–821.