Equilibrium in Mixed-Autonomy Ride-Hailing Networks: Endogenous Fleets, Behavioral Heterogeneity, and Customer Patience

Jiaxin Hou[a], Kexin Wang[b], Ruolin Li[b,*], Jong-shi Pang[a]

[a] *Daniel J. Epstein Department of Industrial and Systems Engineering, University of Southern California, Los Angeles, California 90089-0193, U.S.A.*

[b] *Sonny Astani Department of Civil and Environmental Engineering, University of Southern California, Los Angeles, California 90089-2531, U.S.A.*

Abstract

The rapid advancement of autonomous vehicle (AV) technology is reshaping ride-hailing markets, where transportation network companies (TNCs) increasingly operate mixed AV and human-driven vehicle (HV) fleets. This transition demands an enhanced planning framework that is flexible—accommodating varying AV penetration rates, expressive—capturing operational heterogeneity of mixed autonomy, and unifying. Termed MAGE (Mixed-Autonomy General Equilibrium with Customer Patience), the resulting framework is structured around four coupled modules and distinguishes three behavioral classes: AVs, centrally coordinated by TNCs for profit-driven dispatch and fleet composition; HVs, which serve TNC demand but exhibit decentralized behavior due to human spontaneity and individual preferences; and solo drivers (SVs), who minimize individual travel costs. The behavioral asymmetry between AVs and HVs is explicitly preserved across both pickup and service stages–the two operational phases unique to TNC fleets–where deviations from the Wardrop principle arise through distinct mechanisms, while SVs adhere to this principle. Customer patience endogenously links supply-side heterogeneity to demands: travelers trade waiting for ride-hailing service within patience against solo driving, thus capping wait times and influencing fleet deployment within a flow-dependent congestion network. The overall system is mathematically formulated as a Nonlinear Complementarity Problem, which we show is equivalent to a Variational Inequality. We establish equilibrium existence through a new proof that relaxes a restrictive fleet-size assumption in prior work. Numerical results reveal that AV routing lever amplifies competitive advantages among TNCs but exhibits diminishing marginal returns, eventually saturating, while traveler patience stabilizes the market by implicitly disciplining excessive AV routing control.

Keywords: Autonomous Vehicles, Mixed Autonomy, Mixed-Fleet Traffic Network Companies, Ride-Hailing, Variational Inequality, Nonlinear Complementarity

*Corresponding author

Email addresses: jiaxinho@usc.edu (Jiaxin Hou), kwang255@usc.edu (Kexin Wang), ruolinl@usc.edu (Ruolin Li), jongship@usc.edu (Jong-shi Pang)

1. Introduction

The ride-hailing industry is undergoing a structural transformation driven by the rapid advancement of autonomous vehicle (AV) technology. Recent projections suggest that by 2030, approximately 12% of new passenger cars will feature Level 3+ autonomous capabilities, with this figure potentially rising to 37% by 2035 and generating $300-$400 billion in market value Deichmann et al. (2023). As this technological transition accelerates, a diverse set of companies has emerged or is actively redefining their roles in the emerging AV ecosystem, each pursuing distinct pathways toward automation. Technology-driven pioneers, such as Waymo, operate fully autonomous ride-hailing services across several U.S. cities, providing over 250,000 paid trips per week Waymo (2025). Tesla, as a new entrant, launched robotaxi services with safety drivers in Austin in 2025 and intends to expand to additional cities while progressing toward full autonomous operations Tesla, Inc. (2025). Meanwhile, platform incumbents such as Uber and Lyft are approaching automation through strategic partnerships. Uber began integrating Waymo's AVs into its ride-hailing platform in Atlanta in 2025 Uber Technologies, Inc. (2025), while Lyft plans to introduce Waymo's fully AVs to its service and expand operations to Nashville in 2026 Lyft, Inc. (2025). A defining feature of this transition is that ride-hailing companies are not moving from human-driven to fully autonomous operations overnight. Instead, they are managing, and will continue to manage for the foreseeable future, mixed fleets of AVs and HVs simultaneously. This new operator type, which we term the **Mixed-Fleet Traffic Network Company (MiFleet TNC)**, represents a genuinely new modeling challenge that existing transportation equilibrium frameworks are not designed to address.

The challenge is structural, not merely technical. A MiFleet TNC operating in a mixed-autonomy environment faces decisions that existing unified planning frameworks cannot capture across different levels. *First*, it must *endogenously* (rather than exogenously) determine its fleet composition, i.e, how many AVs versus HVs to deploy, given their structurally different pricing structures, operational costs, capabilities, and regulatory constraints, in order to suit the company's own profit goals. This fleet composition decision determines the AV penetration rate in the network and directly shapes system-level outcomes for companies, travelers, and regulators alike. *Second*, AVs and HVs behave differently on the network in ways that matter fundamentally for equilibrium analysis. AVs are centrally coordinated agents whose routing is governed by company optimization for profit. Company-dispatched HVs, while also serving for company profit goals, are operated by human drivers whose routing reflects individual spontaneity and preferences that do not fully align with company directives. Meanwhile, solo drivers follow shortest-path Wardrop behavior, minimizing their own travel time when routing on the network. This three-way behavioral heterogeneity must be explicitly represented in a credible model of mixed-autonomy ride-hailing traffic networks. *Third*, customer choice is not a passive result but an active force shaping the system. Customer waiting time depends on congestion-dependent pickup travel times, and traveler patience. When waiting becomes excessive, travelers switch their choices for the rides, or abandon the platform for solo driving, endogenously moderating ride-hailing demand and feeding back into fleet deployment decisions and network congestion. This patience mechanism closes the loop between supply and demand in mixed-autonomy markets. *Fourth*, as AV adoption accelerates, central regulators face a fundamental question: how

much should policy intervene in an emerging competitive market consisting of multiple stakeholders with misaligned interests? Instruments such as AV penetration caps, AV routing lever limits, and fleet size regulations can shape system outcomes, but their effects depend critically on how companies and travelers respond, interactions that can only be analyzed within a unified equilibrium framework. Together, these four dimensions define a problem that is both practically urgent and analytically new: one that demands a framework flexible enough to accommodate varying penetration rates of AVs and expressive enough to capture the full behavioral and structural complexity of mixed-autonomy ride-hailing systems.

To address this problem, we develop MAGE (Mixed-Autonomy General Equilibrium with Customer Patience), a unified planning framework structured around four coupled modules. The *MiFleet TNC operational module* allows each company to endogenously optimize its AV–HV fleet composition and dispatch decisions to maximize profit, subject to fleet capacity and regulatory constraints including an upper bound on AV share. The *traveler choice module* represents aggregate mode choice among solo driving versus ride-hailing by vehicle type and company as a function of fare, travel time, and waiting time disutility. The *capped customer waiting module* derives waiting time endogenously through a truncated, congestion-dependent queue formulation, in which travelers' patience serves as a natural cap on waiting. The *traffic congestion module* closes the system by linking path flows from all vehicle types, to network-wide travel times through a flow-dependent congestion model in which solo drivers follow the Wardrop principle. The four modules interact through coupled constraints, shared constraints and feedback loops: company dispatch decisions affect congestion, which affects travel and waiting times, which affect traveler mode choice, which in turn determines the demand each company faces. This coupling makes the overall system a generalized Nash equilibrium problem among competing MiFleet TNCs and a continuum of travelers, all interacting through the traffic network.

The contributions of this paper are threefold. **(i) A new problem and an enhanced unifying modeling framework** capturing endogenous AV–HV fleet composition, stage-dependent behavioral heterogeneity of AVs and HVs, endogenous customer waiting, and flow-dependent congestion, accommodating different AV penetration rates. **(ii) Theoretical analysis** via an NCP–VI formulation with a new equilibrium existence proof that removes the restrictive fleet-size assumption of prior work. **(iii) Policy-relevant insights via extensive case studies**: the AV routing lever is a double-edged sword, a competitive edge for individual companies yet a source of systemic inefficiency when overused; moderate AV penetration strikes the best balance across company, traveler, and system objectives; and traveler patience endogenously stabilizes the market by disciplining excessive automation. These findings suggest that *full automation is not always the answer and must be properly regulated.*

2. Gaps in Existing Literature

Our study is motivated by the gap in the existing transportation literature of how the presence of AVs affect traffic equilibrium and user behavior. On one hand, existing AV studies have largely focused on control-level automation without capturing system-wide interactions. On the other hand, traffic equilibrium models have not accounted for the presence of AVs whose characteristics are quite different from those of HVs. While there are significant advances in both streams,

to be reviewed below, they remain largely disconnected. In addition to bringing these two streams together, we add to their unification the aspects of vehicle heterogeneity between AVs and HVs, competing TNCs managing mixed-autonomy fleets, and interactions with travelers and congestion.

2.1. Lack of Equilibrium Models in Existing AV Studies

A growing body of literature has studied the system-level impacts of automation. These studies typically focus on routing and control, showing that coordinated AV behavior can improve the traffic efficiency and, in some cases, move the system toward socially optimal outcomes Fagnant and Kockelman (2015); Stern et al. (2018); Zheng et al. (2020); Wu et al. (2021); Chen et al. (2020); Battifarano and Qian (2023). At the same time, other works highlight unintended consequences of automation, including induced demand, increased vehicle miles traveled, and spatial imbalance Auld et al. (2017); Childress et al. (2015); Castro et al. (2024); Chen et al. (2024). These contrasting findings underscore the complexity of automation's system-level impacts, making it essential to answer the question: how do mixed fleets interact with platform-level decisions, traveler choice, and traffic congestion? However, existing models typically isolate automation's effects from other transportation actors and has limited attention from the perspective of MiFleet TNC planning. This gap highlights the need for an equilibrium framework that explicitly captures mixed-autonomy fleet heterogeneity and endogenous TNC decision-making.

2.2. Lack of Mixed Autonomy in Transportation Equilibrium

Extensive research has analyzed ride-hailing platforms from the perspective of pricing, fleet management, and matching mechanisms Li et al. (2021); Ni et al. (2021); Zha et al. (2018); Lai and Li (2023); Ke et al. (2020). Many of these works adopt equilibrium-based formulations to capture interactions among platforms, travelers, and, in some cases, the transportation network, jointly determining pricing, demand, and congestion outcomes Ban et al. (2019); Chen and Di (2024); Xu et al. (2021); Gu et al. (2025). Existing frameworks often impose simplifying assumptions: key system components such as fleet composition, customer waiting time, travel time and demand response are treated as exogenous, fixed, or subject to hard operational constraints Xu et al. (2021); Chen and Di (2024); Braverman et al. (2019); Feng et al. (2022). Moreover, although some models effectively capture how different travel modes influence congestion and yield useful policy insights, they typically assume a homogeneous TNC fleet structure across operators Di and Ban (2019); Ban et al. (2019). A related separation of behavioral roles appears in dual-sourcing ride-hailing models, where freelance drivers behave as Wardrop players while idle contracted drivers follow platform-directed repositioning Dong et al. (2024). Nevertheless, these models still do not allow TNCs to endogenously determine fleet composition or explicitly distinguish AV and HV behaviors within the same platform. This gap highlights the need for a mixed-autonomy equilibrium framework that captures endogenous fleet composition and heterogeneous vehicle behavior.

3. Problem Statement

We study a transportation system in which multiple TNCs operate mixed fleets of AVs and HVs, alongside SVs traveling on the same traffic network. In this system, companies make

strategic decisions on fleet composition and routing, while travelers choose among competing ride-hailing services and alternative travel options based on travel time, waiting time, and cost. These decisions are coupled through congestion and pickup dynamics on the underlying network.

The problem is to characterize the equilibrium outcome arising from the interactions among (i) company-level operational decisions at the macro scale, including endogenous fleet composition and vehicle dispatch, (ii) traveler choice behavior with waiting considerations, and (iii) network congestion under heterogeneous routing behaviors. To capture these interdependencies, we develop a unified equilibrium framework in which company strategies, traveler choices, and traffic flows are jointly determined.

3.1. Modeling Foundations

MiFleet TNC operational module. Each MiFleet TNC seeks to maximize its profit through vehicle dispatch decisions. In particular, each company can flexibly determine its fleet composition of AVs and HVs. This endogenous fleet composition determines the AV penetration rate in the transportation network and allows the framework to analyze how varying levels of AV adoption affect system-level performance. MiFleet-TNCs also dispatch vehicles to match customer requests while satisfying flow balance constraints at each destination node.

Traveler choice module. On the demand side, travelers choose between requesting MiFleet-TNC service or driving alone based on perceived disutility incorporating travel time, fare, and waiting time. These traveler decisions determine the demand faced by each company and influence its operational decisions, including fleet deployment and dispatching.

Vehicle behavior and routing. Due to AVs' coordinated characteristics, they can be centrally controlled by TNC, while HVs lack centralized coordination. To accurately capture these differences, MAGE distinguishes routing behaviors across two phases: pickup and service. During the pickup phase, both AVs and HVs may deviate from the Wardrop principle, but their routing behaviors are governed by fundamentally distinct mechanisms. AVs are strictly controlled by the TNC to optimize fleet distribution and maximize profit, whereas HVs are granted bounded freedom to follow individual preferences and relocation choices. During the service phase, TNCs strategically control AV routing to optimize profitability. Consequently, there may exist a network-wide congestion alleviation since AV flows are diverted from the shortest congestion-dependent paths favored by Wardropian drivers. In addition, SVs are assumed to follow the Wardrop principle, since they aim to minimize their individual travel costs. These behavioral distinctions are incorporated into the traffic model in Equation (6), which links path flows of SVs, AVs, and HVs to congestion and determines equilibrium travel times. Table 1 summarizes the driving behaviors of different vehicle types.

Vehicle type	Pickup phase	Service phase
AVs	May deviate under company control	May deviate under company control
HVs	May deviate per human driver	Follow Wardrop principle
SVs	Follow Wardrop principle	

Table 1: Driving Behaviors across Vehicle Types and Trip Stages

The principal system interactions are depicted schematically in Figure 1.

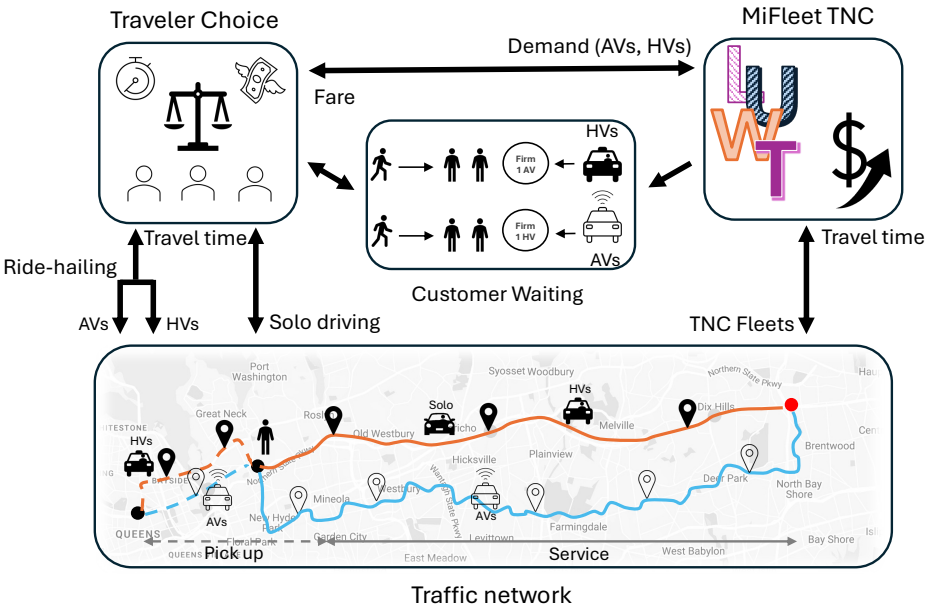


Figure 1: Mixed-Autonomy System Overview

3.2. Notations

With the major foundations established, we next define the mathematical notations used throughout this paper. These notations provide a formal representation of indices, parameters, and variables.

Sets associated with the system

$\mathcal{N}$	Set of nodes in the network
$\mathcal{A}$	Set of arcs in the network
$\mathcal{W}$	Set of Origin-Destination (OD) pairs, a subset of $\mathcal{N} \times \mathcal{N}$
$\mathcal{O}$	Set of origin nodes, $i \in \mathcal{O}$ if $\exists$ some $j \in \mathcal{N}$ such that $(i,j) \in \mathcal{W}$
$\mathcal{D}$	Set of destination nodes, $j \in \mathcal{D}$ if $\exists$ some $i \in \mathcal{N}$ such that $(i,j) \in \mathcal{W}$
$\mathcal{K}$	Set of Traffic Network Companies (TNCs)
$\mathcal{X}$	$\triangleq \{\text{AV}, \text{HV}\}$. Set of TNC vehicle types
$\mathcal{W}^{k,x}$	Set of OD pairs served by vehicle type $x \in \mathcal{X}$ of company $k \in \mathcal{K}$
$\mathcal{O}^{k,x}$	Set of origin nodes served by vehicle type $x \in \mathcal{X}$ of company $k \in \mathcal{K}$, $i \in \mathcal{O}^{k,x}$ if $\exists$ some $s \in \mathcal{N}$ such that $(i,s) \in \mathcal{W}^{k,x}$
$\mathcal{D}^{k,x}$	Set of destination nodes served by vehicle type $x \in \mathcal{X}$ of company $k \in \mathcal{K}$, $s \in \mathcal{D}^{k,x}$ if $\exists$ some $i \in \mathcal{N}$ such that $(i,s) \in \mathcal{W}^{k,x}$
$\mathcal{K}_{ij}^{x}$	Set of TNCs that provide vehicle type x to serve OD pair $(i,j) \in \mathcal{W}$
$\mathcal{P}$	Set of all paths in the network
$\mathcal{P}_{ij}$	Set of paths connecting node $i \in \mathcal{N}$ to node $j \in \mathcal{N}$
$\mathcal{P}_{si}$	Set of paths connecting node $s \in \mathcal{N}$ to node $i \in \mathcal{N}$

Model parameters (all positive):

$F_{ij}^{k,x}$ Fixed fare charged by vehicle type x of company k serving OD pair $(i,j) \in \mathcal{W}^{k,x}$

$\alpha_1^{k,x}$ Travel time-based fare rates for vehicle type x of company k

$\alpha_2^{k,x}$ Travel distance-based fare rates for vehicle type x of company k

$\beta_1^{k,x}$ Travel time conversion factor to monetary costs for vehicle type x of company k

$\beta_2^{k,x}$ Travel distance conversion factor to monetary costs for vehicle type x of company k

$\beta_3^{k,x}$ Waiting time conversion factor to monetary costs for vehicle type x of company k

$\mu_1^{k,x}$ Relaxation factor (≥ 1) of the Wardrop principle for AV operated by company k in serving an OD pair in $\mathcal{W}^{k,\mathrm{AV}}$; this non-Wardropian behavior may be due to regulations that prohibit AVs to take certain routes. For notational uniformity, we define $\mu_1^{k,\mathrm{HV}} = 1$

$\mu_2^{k,x}$ Relaxation factor (≥ 1) of the Wardrop principle for vehicle type x operated by company k in pre-service pick up at a node pair in $\mathcal{D}^{k,x}$; this reflects potential relocation and HV's permitted flexibility in their driving behavior in empty vehicles

t_{ij}^0 Free-flow travel time of the shortest path (in terms of free-flow traveling time) from node $i \in \mathcal{O}$ to node $j \in \mathcal{D}$

$t_{s,i}^0$ Free-flow travel time of the shortest path (in terms of free-flow traveling time) from node $s \in \mathcal{D}$ to node $i \in \mathcal{O}$

d_{ij}^0 Free-flow travel distance of the shortest path (in terms of free-flow traveling time) from node $i \in \mathcal{N}$ to node $j \in \mathcal{N}$

$\mu_{\mathrm{AV}}^{\mathrm{cap}}$ Maximum allowed fraction of AVs in TNC's fleet

N^k Total fleet size of company k

$\gamma_1^{k,x}$ Travel time conversion factor to monetary cost for travelers requesting vehicle x from company k

$\gamma_2^{k,x}$ Waiting time conversion factor to monetary cost for travelers requesting vehicle x form company k

α_1^{SV} Travel time conversion factor to monetary cost for driving vehicles

α_2^{SV} Travel distance conversion factor to monetary cost for solo driving vehicles

D_{ij} Total demand rate of OD pair $(i,j) \in \mathcal{W}$

Primary model variables

$z_{s,ij}^{k,x}$ Dispatch rate of vehicle type x of company k, originating at destination $s \in \mathcal{D}^{k,x}$, assigned to serve OD pair $(i,j) \in \mathcal{W}^{k,x}$

D_{ij}^{SV} Demand rate of OD pair (i,j) with solo driving

$D_{ij}^{k,x}$ Demand rate of OD pair (i,j) requesting vehicle type x from company k

h_p^{SV} Traffic flow of solo driving vehicles on path $p \in \mathcal{P}_{ij}$, for OD pair $(i,j) \in \mathcal{W}$

$h_p^{k,x}$ Traffic flow of vehicle x of company k on path $p \in \mathcal{P}_{ij}$, for OD pair $(i,j) \in \mathcal{W}^{k,x}$

$h_p^{k,x}$ Traffic flow of vehicle x of company k on path $p \in \mathcal{P}_{si}$, for $i \in \mathcal{O}^{k,x}$ and $s \in \mathcal{D}^{k,x}$

t_{ij}^{SV} Congestion dependent travel time of solo driving vehicles from i to j, where $(i,j) \in \mathcal{W}$

$t_{ij}^{k,x}$ Congestion dependent travel time of vehicle type x of company k from i to j where $(i,j) \in \mathcal{W}^{k,x}$

$t_{s,i}^{k,x}$ Congestion dependent travel time of vehicle type x of company k from s to i where $i \in \mathcal{O}^{k,x}$ and $s \in \mathcal{D}^{k,x}$

Induced model variables

$R_{s,ij}^{k,x}$ Per trip revenue for vehicle x of company k serving OD pair $(i,j) \in \mathcal{W}^{k,x}$ from $s \in \mathcal{D}^{k,x}$

$w_{ij}^{k,x}$ Waiting time for customers requesting vehicle x from company k

$\widehat{w}_{ij}^{k,x}$ Waiting time for TNC k's vehicle type x currently at node $s \in \mathcal{D}^{k,x}$ to serve OD pair $(i,j) \in \mathcal{W}^{k,x}$

$\phi_s^{k,x}$ Shadow price of the flow conservation constraint

$\lambda_{ij}^{k,x}$ Marginal price of OD demand (i,j) of vehicle type x, perceived by company k,

$\widehat{\lambda}_{ij}^{k,x}$ Marginal price of OD demand (i,j) of vehicle type x of company k, perceived by customer; assumed to be proportional to $\lambda_{ij}^{k,x}$

ν_{AV}^{k} Marginal price of company k's AV capacity

ν^{k} Marginal price of company k's fleet capacity

σ_{ij} Shadow price of the total demand satisfaction constraint

Auxiliary model variables

$\theta_{s,ij}^{k,x}$, $\zeta_{s,ij}^{k,x}$ Artificial variables employed to handle the ambiguity of the undefined fraction $0/0$ in customers' waiting costs.

In the model formulation, to be presented momentarily, we let $\boldsymbol{h}$ be the tuple consisting of all the path flows:

$$\left\{ h_p^{\mathrm{SV}} \,:\, p \in \mathcal{P}_{ij}\,,\, (i,j) \in \mathcal{W} \right\} \text{ for the solo vehicles,}$$
$$\left\{ h_p^{k,x} \,:\, (k,x) \in \mathcal{K} \times \mathcal{X},\, (i,j) \in \mathcal{W}^{k,x},\, p \in \mathcal{P}_{ij} \right\} \quad \text{and}$$
$$\left\{ h_p^{k,x} \,:\, (k,x) \in \mathcal{K} \times \mathcal{X},\, (s,i) \in \mathcal{D}^{k,x} \times \mathcal{O}^{k,x},\, p \in \mathcal{P}_{si} \right\} \text{for the TNC's vehicles.}$$

We also let $C_p(\boldsymbol{h})$ be the cost on path p, which we assume is continuous and satisfies the natural condition:

$$C_p(\boldsymbol{h}) \,\geq\, C_p(0) \,\geq\, 0, \quad \forall\, p \,\in\, \mathcal{P} \;\text{ and all }\; \boldsymbol{h} \geq 0.$$

The model and its analysis do not assume that the path costs are necessarily derived from an additive model of the link costs; in particular, the BPR link cost functions are not needed. Lastly, we define the free-flow path costs:

$$t_{ij}^0 \triangleq \min_{p \in \mathcal{P}_{ij}} C_p(0), \quad \forall (i,j) \in \mathcal{W} \text{ and } p \in \mathcal{P}_{ij};$$

$$t_{s,i}^0 \triangleq \min_{p \in \mathcal{P}_{si}} C_p(0), \quad \forall (s,i) \in \mathcal{D}^{k,x} \times \mathcal{O}^{k,x} \text{ for some } k \in \mathcal{K},\ x \in \mathcal{X}.$$

4. Mathematical Formulation

Four interrelated submodules constitute the model: the profit-making TNCs operating AVs and HVs, the traveler decision model, the truncated customer waiting model, and traffic conditions. Their interactions are depicted in Figure 2.

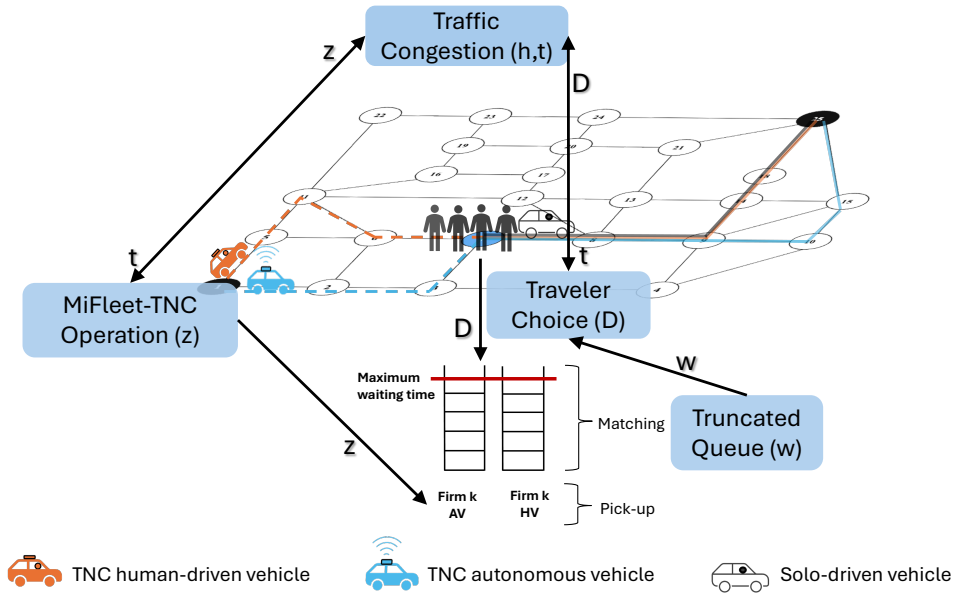


Figure 2: System Interactions

4.1. The MiFleet TNC operational module

For each TNC $k \in \mathcal{K}$, the total fleet size N^k is exogenously given. The TNC determines the dispatch rate of vehicles currently located at node $s \in \mathcal{D}^{k,x}$ and assigned to serve OD pair $(i,j) \in \mathcal{W}^{k,x}$, while endogenously optimizing the AV-HV composition to maximize profit, subject to a regulated upper bound on the AV share. The profit associated with assigning a type-x vehicle is given by $R_{s,ij}^{k,x}$, which is equal to

$$\begin{aligned} & F_{ij}^{k,x} + \underbrace{\alpha_2^{k,x} d_{ij}^0}_{\text{dist. based revenue}} + \underbrace{\alpha_1^{k,x}\left(t_{ij}^{k,x} - t_{ij}^0\right)}_{\text{time-based revenue}} - \underbrace{\beta_1^{k,x}\left(t_{ij}^{k,x} + t_{s,i}^{k,x}\right)}_{\text{time-based cost}} - \underbrace{\beta_2^{k,x}\left(d_{ij}^0 + d_{si}^0\right)}_{\text{dist. based cost}} \\ = & \underbrace{F_{ij}^{k,x} - \alpha_1^{k,x} t_{ij}^0 + \alpha_2^{k,x} d_{ij}^0 - \beta_2^{k,x}(d_{ij}^{k,x} + d_{si}^0)}_{\text{fixed part, denoted } \widetilde{R}_{s,ij}^{k,x}} + \underbrace{\alpha_1^{k,x} t_{ij}^{k,x} - \beta_1^{k,x}(t_{s,i}^{k,x} + t_{ij}^{k,x})}_{\text{variable part}}. \end{aligned}$$

As we adopt a macroscopic planning perspective and focus on overall TNC profitability, we model only the aggregate waiting time across the TNC vehicle types, characterized by the following flow-balancing equation:

$$\sum_{x\in\mathcal{X}}\sum_{(i,j)\in\mathcal{W}^{k,x}}\sum_{s\in\mathcal{D}^{k,x}} z_{sij}^{k,x}\widehat{w}_{ij}^{k,x} \;=\; N^k \;-\; \sum_{x\in\mathcal{X}}\left\{\sum_{(i,j)\in\mathcal{W}^{k,x}}\sum_{s\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x}t_{s,i}^{k,x} \;-\; \sum_{(i,j)\in\mathcal{W}^{k,x}} D_{ij}^{k,x}t_{ij}^{k,x}\right\}.$$

Substituting the right-hand side for the left-hand side into the objective function, TNC k's optimization problem is the following profit-maximization linear program:

$$\left\{\begin{array}{l}
\textbf{maximize}\limits_{z_{s,ij}^{k,\mathrm{AV}},\, z_{s,ij}^{k,\mathrm{HV}}} \\
\quad \text{company profit} \quad \displaystyle\sum_{x\in\mathcal{X}}\left\{\sum_{(i,j)\in\mathcal{W}^{k,x}}\sum_{s\in\mathcal{D}^{k,x}}\left(\underbrace{R_{s,ij}^{k,x}}_{\text{revenue}} - \underbrace{\beta_3^{k,x}\widehat{w}_{ij}^{k,x}}_{\text{monetary waiting cost}}\right) z_{s,ij}^{k,x}\right\} \\
\quad = \displaystyle\sum_{x\in\mathcal{X}}\sum_{(i,j)\in\mathcal{W}^{k,x}}\sum_{s\in\mathcal{D}^{k,x}}\left[\widetilde{R}_{s,ij}^{k,x} + \alpha_1^{k,x}t_{ij}^{k,x} - \beta_1^{k,x}(t_{s,i}^{k,x}+t_{ij}^{k,x})\right] z_{s,ij}^{k,x} + \\
\qquad \beta_3^{k,x}\displaystyle\sum_{x\in\mathcal{X}}\sum_{(i,j)\in\mathcal{W}^{k,x}}\sum_{s\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x}t_{s,i}^{k,x} + \text{constant exogeneous to module} \\
\textbf{subject to :} \\
\quad \text{flow feasibility} \quad \displaystyle\sum_{(i,j)\in\mathcal{W}^{k,x}} z_{s,ij}^{k,x} \le \sum_{i\in\mathcal{O}^{k,x}|(i,s)\in\mathcal{W}^{k,x}} D_{is}^{k,x}, \quad x\in\mathcal{X},\ s\in\mathcal{D}^{k,x} \\
\quad \text{fleet-demand constraint} \quad \displaystyle\sum_{s\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x} \ge D_{ij}^{k,x}, \quad x\in\mathcal{X},\ (i,j)\in\mathcal{W}^{k,x} \\
\text{AV capacity} \quad \displaystyle\sum_{(i,j)\in\mathcal{W}^{k,\mathrm{AV}}}\left\{\sum_{s\in\mathcal{D}^{k,\mathrm{AV}}} t_{s,i}^{k,\mathrm{AV}}z_{s,ij}^{k,\mathrm{AV}} + t_{ij}^{k,\mathrm{AV}}D_{ij}^{k,\mathrm{AV}}\right\} \le \mu_{\mathrm{AV}}^{\mathrm{cap}}N^k \\
\quad \text{fleet capacity} \quad \underbrace{\displaystyle\sum_{x\in\mathcal{X}}\sum_{(i,j)\in\mathcal{W}^{k,x}}\sum_{s\in\mathcal{D}^{k,x}} t_{s,i}^{k,x}z_{s,ij}^{k,x}}_{\substack{\text{vehicles}\\ \text{en route to service calls}}} + \underbrace{\displaystyle\sum_{x\in\mathcal{X}}\sum_{(i,j)\in\mathcal{W}^{k,x}} t_{ij}^{k,x}D_{ij}^{k,x}}_{\substack{\text{vehicles}\\ \text{serving travel demands}}} \le N^k \\
\text{nonnegativity} \quad z_{s,ij}^{k,x}\ge 0, \quad \text{for } x\in\mathcal{X},\ s\in\mathcal{D}^{k,x},\ (i,j)\in\mathcal{W}^{k,x}
\end{array}\right. \tag{1}$$

4.2. The traveler choice module

We model customers as an aggregate decision-making agent. Each traveler chooses a single travel mode per trip over all travel modes that include solo driving, HV or AV service from a TNC. If the disutility associated with any TNC service exceeds that of solo driving, customers shift to solo driving instead. Our model captures this patience mechanism explicitly through disutility comparison. For each OD pair $(i,j)\in\mathcal{W}$, let $V_{ij}^{k,x}$ be the traveler's disutility for

choosing vehicle type x of company k and let V_{ij}^0 be solo driver's disutility traveling from i to j. We have

$$V_{ij}^{k,x} \triangleq F_{ij}^{k,x} + \underbrace{\alpha_1^{k,x}(t_{ij}^{k,x} - t_{ij}^0)}_{\substack{\text{travel time} \\ \text{based disutility}}} + \underbrace{\alpha_2^{k,x} d_{ij}^0}_{\substack{\text{distance} \\ \text{disutility}}} + \underbrace{\gamma_1^{k,x} t_{ij}^{k,x}}_{\substack{\text{travel time} \\ \text{disutility}}} + \underbrace{\gamma_2^{k,x} w_{ij}^{k,x}}_{\substack{\text{waiting disutility} \\ \text{to be picked up}}}$$

$$V_{ij}^{\text{SV}} \triangleq \underbrace{\alpha_1^{SV} t_{ij}^{SV}}_{\substack{\text{travel time} \\ \text{based disutility}}} + \underbrace{\alpha_2^{\text{SV}} d_{ij}^0}_{\substack{\text{travel distance} \\ \text{based disutility}}} .$$

Note that unlike $V_{ij}^{k,x}$ whose sign is not predetermined, the solo driver's disutility V_{ij}^{SV} is always positive. Taking the waiting times $w_{ij}^{k,x}$ as exogenous variables (see the capped customer waiting module), the traveler choice optimization problem is the following disutility minimization linear program:

$$\left\{ \begin{array}{l} \textbf{minimize}_{\ D_{ij}^{k,x},\ D_{ij}^{\text{SV}}} \\ \quad \text{traveler disutility} \quad \displaystyle\sum_{(i,j)\in\mathcal{W}} \left(V_{ij}^{\text{SV}} D_{ij}^{\text{SV}} + \sum_{x\in\mathcal{X}} \sum_{k\in\mathcal{K}_{ij}^x} V_{ij}^{k,x} D_{ij}^{k,x} \right) \\ \textbf{subject to} \\ \quad \text{total demand satisfaction} \quad D_{ij}^{\text{SV}} + \displaystyle\sum_{x\in\mathcal{X}} \sum_{k\in\mathcal{K}_{ij}^x} D_{ij}^{k,x} \geq D_{ij}, \quad \forall\, (i,j) \in \mathcal{W} \\ \quad \text{fleet dictated demand constraint} \quad \displaystyle\sum_{s\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x} \geq D_{ij}^{k,x}, \quad \forall\, k\in\mathcal{K},\, x\in\mathcal{X},\, (i,j)\in\mathcal{W}^{k,x} \\ \quad \text{nonnegativity} \quad D_{ij}^{\text{SV}} \geq 0, \quad \forall\, (i,j)\in\mathcal{W} \\ \qquad\qquad D_{ij}^{k,x} \geq 0, \quad \forall\, k\in\mathcal{K},\, x\in\mathcal{X},\, (i,j)\in\mathcal{W}^{k,x} \end{array} \right\} \tag{2}$$

4.3. Capped customer waiting module

Patience. To formalize the customer waiting time, we begin by noting that travelers have limited patience and have the option of not waiting for TNC service. We model customer patience as a threshold, the maximum time a traveler is willing to wait, beyond which they switch to solo driving. Endogenous to the model, this threshold is naturally tied to the traveler choice model, as it depends on the relative disutility between solo driving and TNC services. The following Lemma 1 provides a key structural property that supports the characterization of such threshold.

Lemma 1. Suppose that $V_{ij}^{k,x} > V_{ij}^{\text{SV}}$ for some OD pair $(i,j) \in \mathcal{W}$ and company-vehicle type pair $(k,x) \in \mathcal{K}\times\mathcal{X}$, then there exists an optimal solution to the traveler choice problem (2) satisfying $D_{ij}^{k,x} = 0$.

Proof. It suffices to note that if $D_{ij}^{k,x} > 0$ in an optimal solution, then shifting $D_{ij}^{k,x}$ to D_{ij}^{SV} preserves feasibility while strictly improving the objective value. □

Lemma 1 implies that any TNC option whose disutility exceeds that of solo driving will not be chosen in equilibrium. Based on this observation, we define the maximum waiting time as

$$w_{ij}^{\max} \triangleq \max\left\{0,\ \frac{V_{ij}^{\mathrm{SV}} - \min\limits_{k\in\mathcal{K},x\in\mathcal{X}}\left\{F_{ij}^{k,x} + \alpha_1^{k,x}(t_{ij}^{k,x} - t_{ij}^0) + \alpha_2^{k,x} d_{ij}^0 + \gamma_1^k t_{ij}^{k,x}\right\}}{\min\limits_{k\in\mathcal{K},x\in\mathcal{X}} \gamma_2^{k,x}}\right\},$$

which is a measure of customer's patience.

Waiting time. Given that customers are willing to wait (up to the threshold $w_{ij}^{\max}$), we model the total waiting time as the sum of matching (dispatching) time and pickup travel time, both measured from the customer's perspective. There are multiple ways to model these two components.

In general, the matching time can be represented as an extended-value function of the tuple $\{z_{ij}^{k,x}, D_{ij}^{k,x}\}$ of company's vehicle allocation and OD demands. A common approach in the literature is to adopt queuing-based formulation. For instance, an M/M/1 queue can be used to represent the matching process. According to the this model, the customer waiting for company k's vehicle type x traveling between OD pair $(i,j) \in \mathcal{W}^{k,x}$ is considered as a queue, in particular an M/M/1 model, where vehicles are the servers and passengers are customers. Company k's vehecle x arrive exponentially with mean $1/\sum\limits_{s\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x}$ for $x \in \mathcal{X}$. The interarrival times of customers requesting x from company k follow an exponential distribution with mean $1/D_{ij}^{k,x}$. Mean matching time is modeled as the mean waiting time in the queuing system. Thus we have the queuing-based steady-state matching time:

$$\frac{1}{\sum\limits_{s\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x} - D_{ij}^{k,x}}, \quad \forall\, k\in\mathcal{K},\, x\in\mathcal{X},\, (i,j)\in\mathcal{W}^{k,x}, \tag{3}$$

where the denominator is nonnegative due to the fleet-demand constraint; it can equal to zero, leading to a infinite matching time that will be resolved by capping. Such queue-based formulations are widely used; for example, Gu et al. (2025) employs a queue-based matching mechanism, though it does not account for congestion effects on pickup delays. Alternatively, some studies approximate matching time using the marginal price associated with the fleet demand constraint $\sum\limits_{s\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x} \geq D_{ij}^{k,x}$. Since the multiplier is complementary to the slack of this constraint, an issue with this definition is that when equality holds, the matching time is not well defined.

In contrast to matching, the pickup travel time is naturally linked to the spatial distribution of available vehicles. In general, it can also be modeled as a continuous function of the tuple $\{z_{ij}^{k,x}, D_{ij}^{k,x}\}$. In this work, we adopt a probabilistic formulation, whereby pickup time is derived from the distribution of idle vehicles and their travel times to customers. The mean pickup time is computed as a weighted average of travel times from all candidate dispatch locations. Specifically, we have

$$\sum_{s\in\mathcal{D}^{k,x}} t_{s,i}^{k,x}\ \theta_{s,ij}^{k,x}, \quad \forall\, k\in\mathcal{K},\, x\in\mathcal{X},\, (i,j)\in\mathcal{W}^{k,x}.$$

Here, $\theta_{s,ij}^{k,x} \in [0,1]$ represents the share of vehicles coming from location s, with $\sum_{s \in \mathcal{D}^{k,x}} \theta_{s,ij}^{k,x} = 1$. When vehicles are available, these shares are naturally given by

$$\theta_{s,ij}^{k,x} = \frac{z_{s,ij}^{k,x}}{\sum_{s' \in \mathcal{D}^{k,x}} z_{s',ij}^{k,x}}.$$

The capped customer waiting time module. In the rest of the paper, we take the customer's waiting time to be

$$w_{ij}^{k,x} \triangleq \min \left\{ w_{ij}^{\max},\ \widetilde{w}_{ij}^{k,x} + \sum_{s \in \mathcal{D}^{k,x}} t_{s,i}^{k,x}\, \theta_{s,ij}^{k,x} \right\} \tag{4}$$

where $\widetilde{w}_{ij}^{k,x}$ is a nonnegative, possibly extended-valued function of the tuple $\{z_{ij}^{k,x}, D_{ij}^{k,x}\}$ that is continuous on its domain of finiteness, and, computationally,

$$\theta_{s,ij}^{k,x} = \arg \min_{\theta \in [0,1]} \left\{ -z_{s,ij}^{k,x}\theta + \frac{1}{2} \Big(\sum_{s' \in \mathcal{D}^{k,x}} z_{s',ij}^{k,x} \Big) \theta^2 \right\} \tag{5}$$

For the purpose of analysis, the explicit form of the waiting time function is not important, it is the continuity of the function in the arguments $\{ z_{ij}^{k,x}, D_{ij}^{k,x}, t_{s,i}^{k,x}, \theta_{s,ij}^{k,x} \}$ that is needed. Nevertheless, for the purpose of computation, an explicit form of the matching time $\widetilde{w}_{ij}^{k,x}$, such as (3), is needed.

4.4. The traffic congestion module

This is an extension of the basic Wardrop model; The most distinctive feature of this extension is that we relax the Wardrop shortest-path principle for AVs in both service and pickup trips, as they operate under centralized TNC coordination and may be strategically relocated. For HVs, we impose the Wardrop principle when they serve passenger demand, but relax it during pickup trips, reflecting the operational freedom granted to HVs, accounting for potential relocation decisions and personal routing preferences.

Travel delay

$$\begin{aligned}
&\text{SV} \quad 0 \le t_{ij}^{\mathrm{SV}} \quad \perp \quad \sum_{p \in \mathcal{P}_{ij}} h_p^{\mathrm{SV}} - D_{ij} + \sum_{x \in \mathcal{X}} \sum_{k \in \mathcal{K}_{ij}^x} D_{ij}^{k,x} \,(= h_p^{\mathrm{SV}} - D_{ij}^{\mathrm{SV}}) \ge 0,\ \forall (i,j) \in \mathcal{W} \\
&\text{service} \quad 0 \le t_{ij}^{k,x} \quad \perp \quad \sum_{p \in \mathcal{P}_{ij}} h_p^{k,x} - D_{ij}^{k,x} \ge 0, \qquad \forall x \in \mathcal{X},\ k \in \mathcal{K},\ (i,j) \in \mathcal{W}^{k,x} \\
&\text{pickup} \quad 0 \le t_{s,i}^{k,x} \quad \perp \quad \sum_{p \in \mathcal{P}_{si}} h_p^{k,x} - \sum_{j \in \mathcal{D}^{k,x}} z_{s,ij}^{k,x} \ge 0, \quad \forall x \in \mathcal{X},\ k \in \mathcal{K},\ (s,i) \in \mathcal{D}^{k,x} \times \mathcal{O}^{k,x}
\end{aligned} \tag{6}$$

Vehicle flow

$$\text{SV} \quad 0 \le h_p^{\mathrm{SV}} \quad \perp \quad C_p(\boldsymbol{h}) - t_{ij}^{\mathrm{SV}} \ge 0, \quad \forall (i,j) \in \mathcal{W},\ \forall p \in \mathcal{P}_{ij}$$

$$\begin{aligned}
&\text{service}\ \ 0 \le h_p^{k,x} \quad \perp \quad \mu_1^{k,x} C_p(\boldsymbol{h}) - t_{ij}^{k,x} \ge 0,\ \forall\, x \in \mathcal{X},\ k \in \mathcal{K},\ (i,j) \in \mathcal{W}^{k,x},\ p \in \mathcal{P}_{ij} \\
&\text{pickup}\ \ 0 \le h_p^{k,x} \quad \perp \quad \mu_2^{k,x} C_p(\boldsymbol{h}) - t_{s,i}^{k,x} \ge 0,\ \forall\, x \in \mathcal{X},\ k \in \mathcal{K}, (s,i) \in \mathcal{D}^{k,x} \times \mathcal{O}^{k,x}, p \in \mathcal{P}_{si}
\end{aligned}$$

Overall, the mixed-fleet traffic equilibrium problem with the coexistence of HV, AV, and SV is to solve for the primary model variables satisfying the conditions in the four modules introduced above. The problem can be studied as a noncooperative game where the primary agents (the TNCs and the travelers) are competing for the use of the traffic network subject to the Wardrop principle (with extensions to allow for their relaxations for the TNCs' fleets) and the customer waiting times for pick-up vehicles. The game is of the generalized type as the fleet-demand constraints appear in both the TNC and customer modules and the flow conservation constraint is of the coupled and shared type.

5. Existence of a generalized equilibrium

Our goal is to establish the existence of a generalized equilibrium 2 of the mixed-fleet transportation system by examining the details and interactions of its four submodules.

Definition 2 (Normalized MAGE)**.** A normalized MAGE equilibrium is a tuple $\{\mathbf{z}, \mathbf{D}, \boldsymbol{\theta}, \mathbf{h}, \mathbf{t}\}$ such that

1. For each $k \in \mathcal{K}$, the tuple $\mathbf{z} \triangleq \{z_{s,ij}^{k,x} : x \in \mathcal{X}, s \in \mathcal{D}^{k,x}, (i,j) \in \mathcal{W}^{k,x}\}$ solves Problem (1) for given $\{\mathbf{D}, \boldsymbol{\theta}, \mathbf{h}, \mathbf{t}\}$;
2. Tuple $\mathbf{D} \triangleq \{D_{ij}^{k,x}, D_{ij}^{\mathrm{SV}} : (i,j) \in \mathcal{W}, x \in \mathcal{X}, k \in \mathcal{K}_{ij}^x\}$ solves Problem (2) for given $\{\mathbf{z}, \boldsymbol{\theta}, \mathbf{h}, \mathbf{t}\}$, where each $w_{ij}^{k,x}$ is a continuous function of $\{ z_{ij}^{k,x}, D_{ij}^{k,x}, t_{s,i}^{k,x}, \theta_{s,ij}^{k,x} \}$ such as that given by (4);
3. Tuple $\boldsymbol{\theta} \triangleq \{\theta_{s,ij}^{k,x} : x \in \mathcal{X}, s \in \mathcal{D}^{k,x}, (i,j) \in \mathcal{W}^{k,x}\}$ solves (5) for given $\{\mathbf{z}, \mathbf{D}, \mathbf{h}, \mathbf{t}\}$;
4. Tuple $\mathbf{h} \triangleq \{h_p^{\mathrm{SV}}, t_{ij}^{\mathrm{SV}}, h_p^{k,x}, t_{ij}^{k,x}, t_{s,i}^{k,x} : (i,j) \in \mathcal{W}, x \in \mathcal{X}, k \in \mathcal{K}_{ij}^x, (s,i) \in \mathcal{D}^{k,x} \times \mathcal{O}^{k,x},$ and $p \in \mathcal{P}_{ij} \cup \mathcal{P}_{si}\}$ satisfies (6) for given $\{\mathbf{z}, \mathbf{D}\}$
5. The multipliers associated with the shared fleet-demand constraint $\sum_{s \in \mathcal{D}^{k,x}} z_{s,ij}^{k,x} \ge D_{ij}^{k,x}$ in Problems (1) and (2) are proportional.

This is Rosen's solution concept for generalized Nash games with shared constraints Rosen (1965).

Theorem 3. Let the travel demands D_{ij} be positive for all $(i,j) \in \mathcal{W}$. Suppose that the path cost functions $C_p(\boldsymbol{h})$ are continuous, and nonnegative and satisfy the weak positivity conditions (see Appendix A1, and see (Facchinei and Pang, 2003, Proposition 1.4.6) for background), and that the customer waiting times $w_{ij}^{k,x}$ are continuous functions of the tuples $\{ z_{ij}^{k,x}, D_{ij}^{k,x}, t_{s,i}^{k,x}, \theta_{s;ij}^{k,x} \}$. Then there exists a normalized MAGE equilibrium.

Proof sketch. For simplicity, the proof is for a *variational equilibrium* in which the multipliers of the share constraints are equal. This simplification does not affect the analysis for a normalized equilibrium and leads to a nonlinear complementarity problem (NCP) formulation of the model. The NCP is then shown to be equivalent a variational inequality (VI), to which we apply a fundamental degree theory based theorem—Proposition A2—to complete the proof. A key step in the proof is to establish the boundedness of the model's essential variables. See details in Appendix A3.

Why a new proof is needed. There are two related models whose proofs are established under restricted settings. First is the model in Ban et al. (2019), wherein the existence theorem is proved under a restriction on the TNC fleet sizes. Ideally, such a restriction should be dropped since travelers have the option of solo driving. The other related model is Gu et al. (2025), which assumes fixed travel times and ignores traffic congestion. Such assumption is far from ideal in traffic equilibrium analysis. Our proof of Theorem 3 makes a contribution to the analysis of the model.

6. Benchmark Numerical Results

Our model is implemented in GAMS and solved by the PATH solver, which is designed to solve MiCPs Dirkse and Ferris (1995). The PATH algorithm employs a Newton-based approach to efficiently locate solutions that satisfy complementarity and feasibility conditions GAMS Development Corporation (2025). To evaluate the proposed model, two networks are used: a small network and the Sioux-Falls network LeBlanc et al. (1975). The small network is used to validate the model by comparing its results with other equilibrium framework such as Ban et al. (2019). The Sioux-Falls network, representing a realistic urban-scale case, is used to conduct a comprehensive analysis of how TNCs adjust AV shares under maximum allowed AV fractions $\mu_{\text{AV}}^{\text{cap}}$ to achieve higher profit, and how difference between AV and HV routing behaviors impact the mixed-autonomy transportation systems.

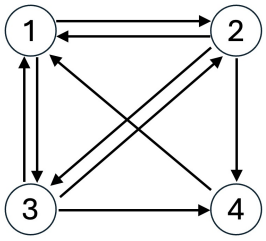


Figure 3: Illustration of the Small Network

The small network is a "4-node–9-link" network, as shown in Fig. 3. Node 1 serves as the origin, while nodes 2, 3, and 4 are destinations. The total travel demands are 50, 40, and 50, respectively. The free flow travel time, link length and link capacity of the small network are provided in Table 2. Company parameters related to fees and costs are provided in Table 3.

Table 2: Small Network Geometry Parameters

Links	**From node**	**To node**	**Free flow travel time (h)**	**Length (mile)**	**Capacity (veh/h)**
1	1	2	0.3	10	40
2	1	3	0.5	20	40
3	2	3	0.4	20	60
4	2	4	0.4	10	40
5	3	4	0.3	20	40
6	4	1	1.0	40	60
7	2	1	0.4	15	50
8	3	1	0.4	20	60
9	3	2	0.5	20	40

Table 3: Small Network Company Parameters

Parameters	**Notation(x=HV)**	**Company 1**	**Company 2**
Fixed fare ($)	$F^{k,x}$	3	2
Time-based fare rate ($/hr)	$\alpha_1^{k,x}$	20	15
Distance-based fare rate ($/mile)	$\alpha_2^{k,x}$	2	1.5
Time-based conversion factor ($/hr)	$\beta_1^{k,x}$	2	2
Distance-based conversion factor ($/mile)	$\beta_2^{k,x}$	0.55	0.9
Waiting time conversion factor ($/hr)	$\beta_3^{k,x}$	0.2	0.1
Value of travel time of customer ($/hr)	$\gamma_1^{k,x}$	7	18
Value of waiting time of customer ($/hr)	$\gamma_2^{k,x}$	3	2
Relaxation factor	$\mu_1^{k,x}$	1.0	1.0
Number of vehicles	N^k	400	400

To validate our model, we consider the HV-only case with $\mu_1^{k,\text{HV}} = 0$ to maintain consistency with the equilibrium model in Ban et al. (2019). Specifically, we compare customer demands, VMT, VHT, DHM, and average waiting time in Table 4. The first three columns (Solo, Company 1, Company 2) indicate the customer demand proportions, which are highly consistent between two models. The differences in VMT and VHT are both within 10%, indicating strong alignment in equilibrium outcomes. Importantly, our model achieves a substantially lower average waiting time of 0.714 minutes per order, representing nearly a 50% reduction (1.39 minutes). This improvement demonstrates that the capped customer waiting formulation introduced in our model effectively enhances system efficiency. These findings confirm the validity of our model and provide a reliable foundation for further analysis in mixed-autonomy systems.

Table 4: Comparison of Our Model with Other Model ($x = \text{HV}, \mu_1^{k,x} = 1.0$)

Model	Solo(%)	Company 1(%)	Company 2(%)	VMT	VHT	DHM	Avg w (min)
Ban et al.	25.5	69.9	4.6	4943.307	1.456	3005.978	**1.390**
Our Model	29.8	64.3	6.7	4834.412	1.457	2940.647	**0.714**

7. Case Study

For large-scale analysis, we use the Sioux-Falls network shown in Figure 4. The network contains 24 nodes and 76 directed links LeBlanc et al. (1975), and the dataset is openly available Stabler (2025). We select five nodes (1, 2, 4, 7, 9) as origins and five nodes (13, 19, 20, 23, 24) as destinations, resulting in a total of 25 OD pairs.

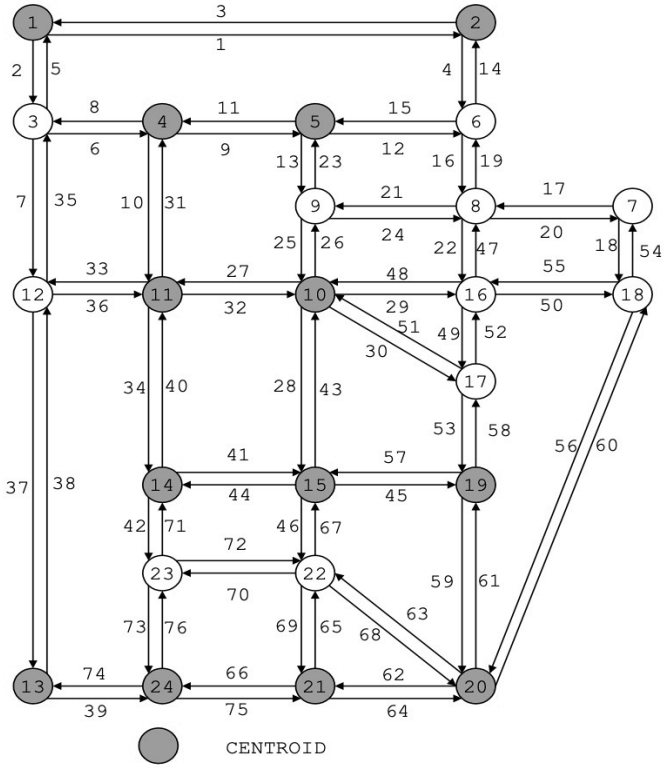


Figure 4: Illustration of the Sioux-Falls Network Stabler (2025)

To capture the heterogeneity in fare and cost strategies across ride-hailing companies, we define four representative company types in Figure 5, following Porter's Generic Strategies framework Miller and Friesen (1986).

• **Company 1 (Technology driven)**: adopts a differentiation strategy, emphasizing advanced IT capabilities and autonomous driving technologies. Denoted k1, this company offers medium–low fare levels to attract early adopters while maintaining low operational costs through automation and optimized fleet management.

• **Company 2 (Aggressive entrant)**: follows a cost-leadership strategy with an aggressive market-entry approach. Denoted k2, this company adopts low fares to rapidly gain market share, despite incurring high operating costs due to substantial capital expenditures and rapid fleet deployment.

• **Company 3 (Market leader)**: adopts a scale-based dominance strategy. Denoted k3, this company leverages its large user base and operational experience and sets high prices while benefiting from medium–low costs, owing to economies of scale and optimized routing.

• **Company 4 (Competitive co-player)**: acts as a focus-differentiation competitor, a smaller

but strategically adaptive company operating alongside the market leader. Denoted k4, this company sets medium–high fare levels and experiences medium–high operating costs, reflecting its intermediate market position.

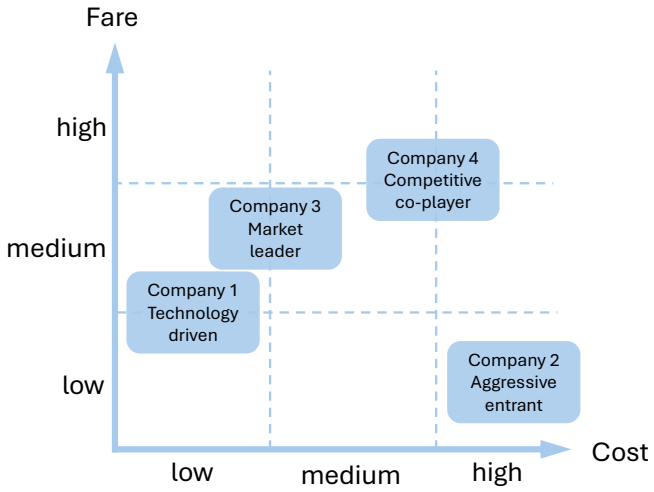


Figure 5: Company Strategy

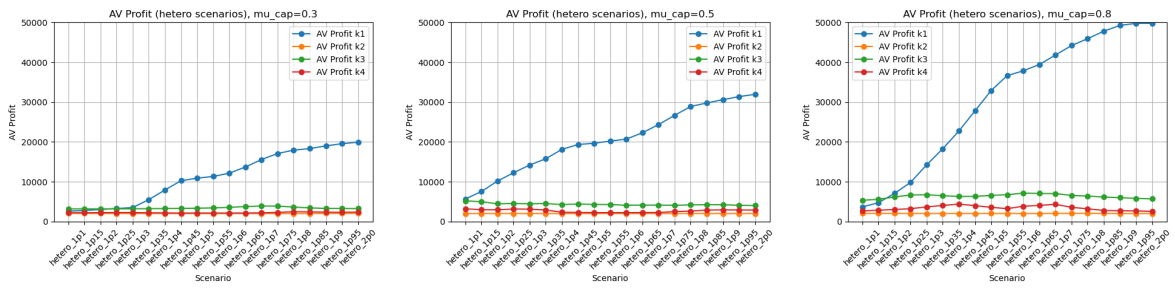
7.1. Company Analysis

We investigate how variations in fleet composition—the maximum allowed AV proportion $\mu_{\text{AV}}^{\text{cap}}$—and AV routing strategies—the AV relaxation factor $\mu_1^{k,\text{AV}}$ affect company profit and customer demands. Unless stated otherwise, $\mu_1^{k,\text{HV}} = 1.1$.

7.1.1. Impacts of AV Routing Strategy

Two AV routing scenarios are analyzed to show their impact among company competition. **Homogeneous Scenario:** all four companies adopt the same relaxation factor $\mu_1^{k,\text{AV}}$. **Heterogeneous Scenario:** company 1 adopts a more aggressive strategy with $\mu_1^{1,AV} > 1$, and the remaining companies ($k = 2, 3, 4$) maintain $\mu_1^{k,AV} = 1$.

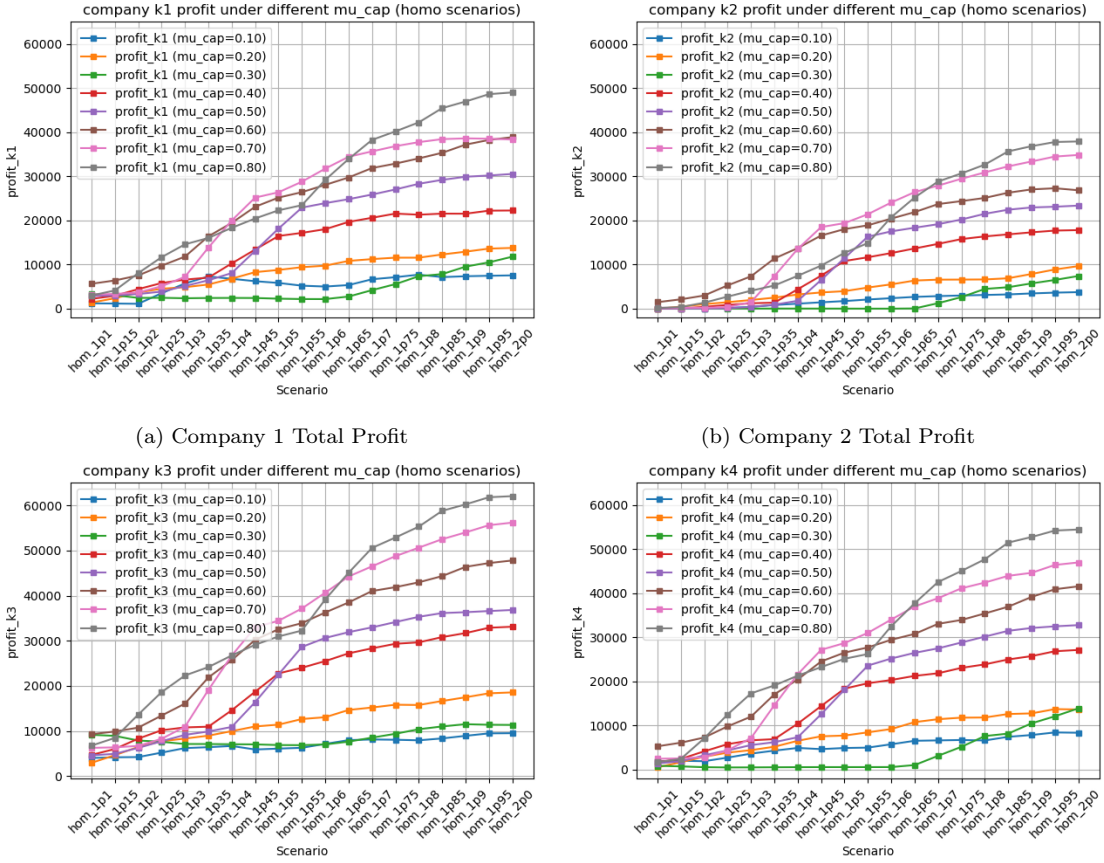
Figure 6 presents AV profit under heterogeneous scenarios, where x-axis represents $\mu_1^{1,AV}$. Results indicate that the technology-driven company (k1) with more relaxation on AV routing, consistently secures substantially greater profits compared to other competitors. For each company, the AV routing lever amplifies competitive advantages among companies, and increasing AV routing relaxation can therefore yield significant profits. However, this also highlights a potential risk that overly profit-pursuing strategies may undermine fair competition.



(a) AV Profit ($\mu_{\text{AV}}^{\text{cap}} = 0.3$) (b) AV Profit ($\mu_{\text{AV}}^{\text{cap}} = 0.5$) (c) AV Profit ($\mu_{\text{AV}}^{\text{cap}} = 0.8$)

Figure 6: AV Profit under Different $\mu_{\text{AV}}^{\text{cap}}$ (Heterogeneous Scenarios), 1p1 stands for 1.1

Figure 7 presents total company profits under homogeneous scenarios, where the x-axis represents $\mu_1^{k,\mathrm{AV}}$. Each subplot corresponds to one company's performance, while each curve within the subplot represents different values of $\mu_{\mathrm{AV}}^{\mathrm{cap}}$. From low to moderate $\mu_1^{k,\mathrm{AV}}$, companies earn significantly higher profits as $\mu_1^{k,\mathrm{AV}}$ increases. For moderate to high $\mu_1^{k,\mathrm{AV}}$, profits approach a saturation point, where further increases contribute little to company profits. This indicates that a suitable level of AV routing control is beneficial for companies, but excessively high routing control provides limited additional benefits and inefficiency.



(a) Company 1 Total Profit

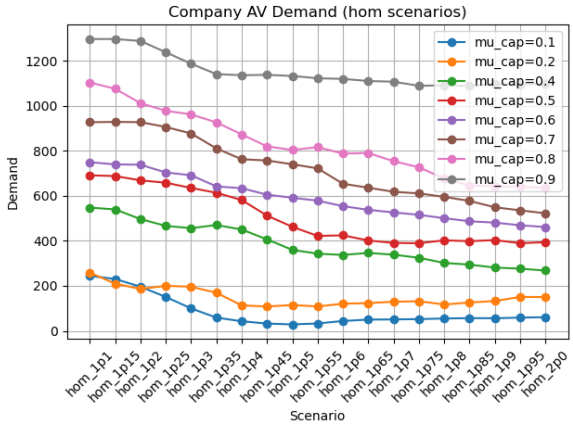
(b) Company 2 Total Profit

(c) Company 3 Total Profit

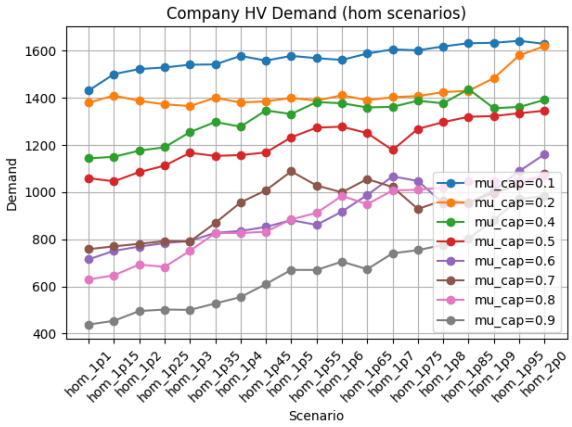
(d) Company 4 Total Profit

Figure 7: Total Profit under Different $\mu_{\mathrm{AV}}^{\mathrm{cap}}$ (Homogeneous Scenarios)

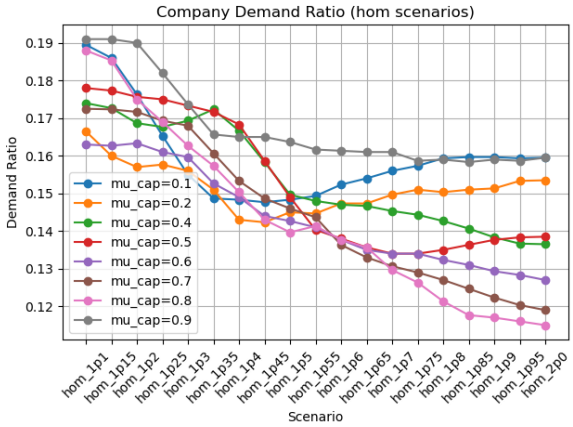
Figure 8 shows customer demand for AVs, HVs, and SVs. As $\mu^{k,\mathrm{AV}}$ increases, subfigures (a) and (b) exhibit opposite trends for AV and HV demand. This indicates that, as companies increase AV routing control, travelers tend to shift from AVs to HVs within ride-hailing services. Furthermore, as $\mu^{k,\mathrm{AV}}$ continues to increase, subfigures (c) and (d) show that the total demand for both AVs and HVs declines. This suggests that traveler patience endogenously stabilizes the market by disciplining excessive routing control, leading more travelers to switch to solo driving, as illustrated by profit saturation points in Figure 7.



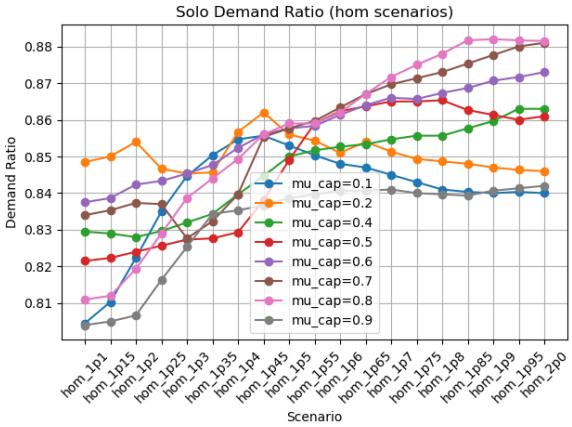
(a) AV Demand for All Company



(b) HV Demand for All Company



(c) Company Demand (HV and AV) Ratio



(d) Solo Driver Demand Ratio

Figure 8: Market Share under Different $\mu_1^{k,\mathrm{AV}}$ (Homogeneous Scenarios)

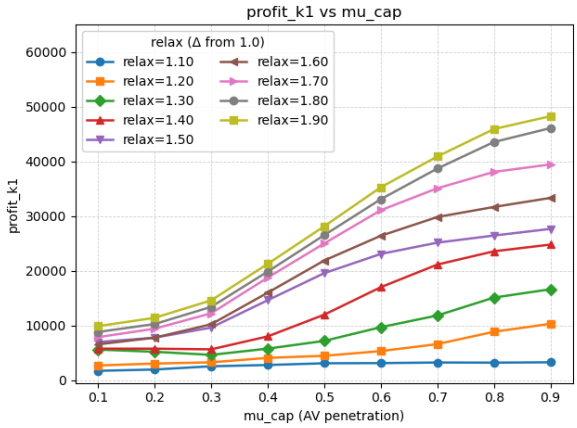
7.1.2. Impacts of AV Fleet Composition

Figure 9 illustrates the impact of the maximum allowed AV fraction $\mu_{\mathrm{AV}}^{\mathrm{cap}}$ on total company profit. When $\mu_{\mathrm{AV}}^{\mathrm{cap}}$ is very low (below 0.2) or very high (above 0.8), the profit growth remains modest. In contrast, intermediate values of $\mu_{\mathrm{AV}}^{\mathrm{cap}}$ exhibit a notable increase in profit as it rises. These findings suggest that a moderate AV penetration is most beneficial for companies, striking a balance between operational efficiency and market demand.

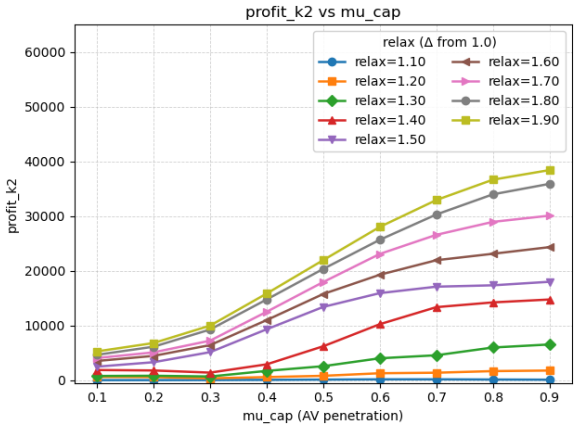
7.2. System Analysis

In this subsection, we examine how variations in fleet composition and AV routing strategies affect system-level metrics, including VMT, VHT, and average Wardrop travel time.

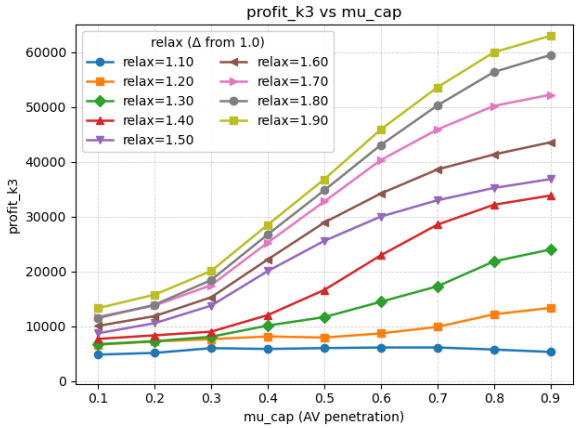
Figure 10 shows that VMT and VHT generally decrease as the AV relaxation factor $\mu_1^{k,\mathrm{AV}}$ increases, particularly when the maximum allowed AV fraction $\mu_{\mathrm{AV}}^{\mathrm{cap}}$ is maintained within a moderate to high range (approximately 0.4–0.9). However, under high $\mu_1^{k,\mathrm{AV}}$, such benefits are less pronounced. This suggests that a suitable level of AV penetration can improve overall network efficiency, whereas excessively AV routing control may lead to system inefficiency.



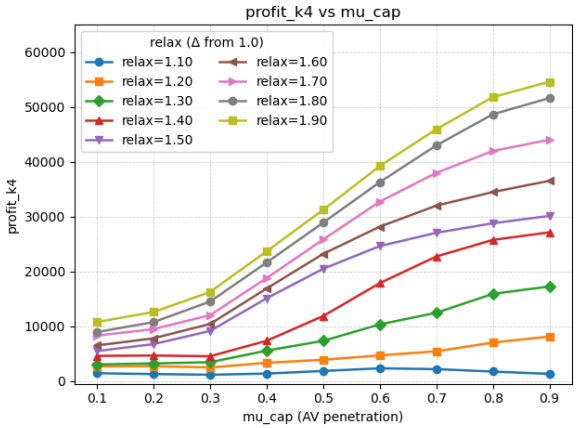
(a) Company 1 Total Profit



(b) Company 2 Total Profit

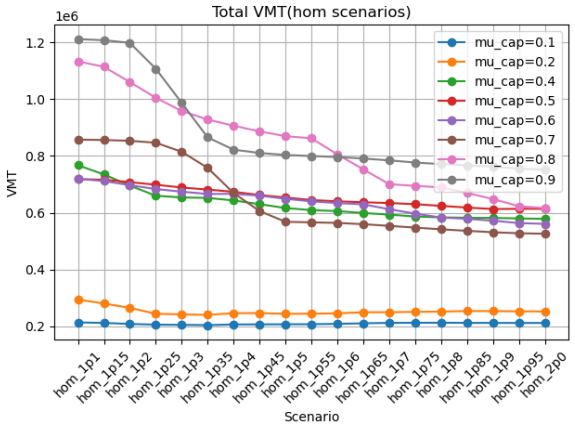


(c) Company 3 Total Profit

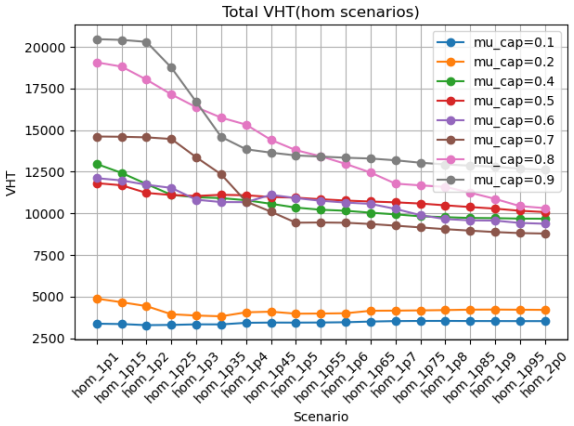


(d) Company 4 Total Profit

Figure 9: Total Profit under Different $\mu_{\text{AV}}^{\text{cap}}$



(a) total VMT

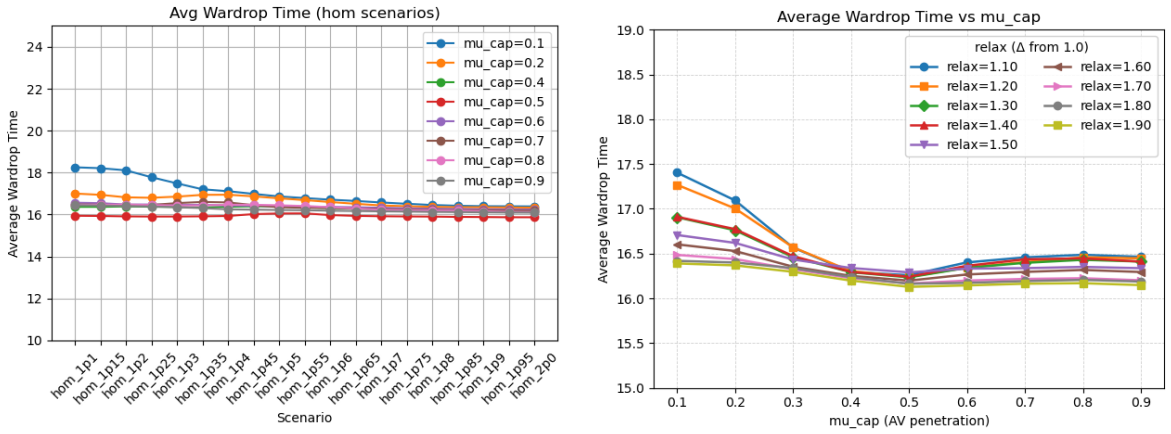


(b) total VHT

Figure 10: System-level Impact under Different $\mu_1^{k,\text{AV}}$

Figure 11 shows the average Wardrop time (defined as the average minimum travel time across all OD pairs and travelers) is not significantly affected by variations in $\mu_{\text{AV}}^{\text{cap}}$ and $\mu_1^{k,\text{AV}}$. However, subfigure (b) reveals a consistent pattern: the average Wardrop time achieves its minimum at a moderate $\mu_{\text{AV}}^{\text{cap}} \approx 0.5$ across different relaxation factors. This also suggests that a moderate level

of AV penetration can maximize the positive travel impact for HVs and SVs.



(a) Average Wardrop Time via $\mu_1^{k,\mathrm{AV}}$

(b) Average Wardrop Time via $\mu_{\mathrm{AV}}^{\mathrm{cap}}$

Figure 11: System-level Time Impact under Different $\mu_1^{k,\mathrm{AV}}$ and $\mu_{\mathrm{AV}}^{\mathrm{cap}}$

Overall, the numerical experiments show that the AV routing lever is a double-edged sword. While stronger routing control provides a competitive advantage for individual companies (Figure 6), it may also lead to firm-level and system-level inefficiencies when overused (Figures 7 and 10). Furthermore, moderate AV penetration achieves the best balance across company, traveler, and system objectives (Figures 9, 10, and 11), while traveler patience endogenously stabilizes the market by disciplining excessive automation (Figure 8). These findings suggest that full automation is not always optimal and should be appropriately regulated.

8. Conclusion

This paper proposes a unified equilibrium model for mixed-fleet ride-hailing systems, integrating the interactions among MiFleet TNCs, travelers, and traffic. The framework is flexible and can be extended to accommodate multiple heterogeneous fleet types, in particular, explicitly differentiating AV and HV behaviors across operational stages. It further introduces the idea of customer waiting functions that include as a special case of a truncated, congestion-dependent, queue-based formulation for customer waiting time that endogenously links customer patience to network congestion. We provide a new proof of equilibrium existence under the mere continuity of the model functions and the weak positivity of the path costs. These theoretical advances enable a more realistic and analytically tractable representation of customer waiting and traffic flow interactions in mixed-autonomy environments.

The proposed model effectively captures multilevel impacts across company, system, and traveler scales, uncovering key patterns in AV deployment. The numerical results indicate that a moderate or properly tuned AV penetration rate achieves the most balanced outcomes, enhancing company profitability, mitigating congestion, and improving traveler welfare. In contrast, excessive automation introduces inefficiencies, such as system congestion, underscoring the importance of appropriate regulatory oversight.

Beyond these immediate findings, the proposed framework provides a foundation for future research on equilibrium modeling in intelligent transportation systems and ride-hailing company

strategies, including the analysis of market performance and design of suitable pricing strategies under low and high AV penetration rates. By combining analytical rigor with practical interpretability, our study contributes to the broader vision of sustainable, coordinated, and human-centered automation, offering guidance to policymakers and platform operators navigating the evolving landscape of mixed-autonomy mobility.

Appendix

A1. Weak Positivity Conditions on Path Costs

Naturally satisfied when $C_p(h)$ are all positive for nonzero flow vectors h, the following conditions are standard and date back to the early studies of traffic equilibrium; (Facchinei and Pang, 2003, Proposition 1.4.6).

• For all OD pairs $(i,j) \in \mathcal{W}$:

$$\left[\sum_{p \in \mathcal{P}_{ij}} h_p^{\mathrm{SV}} \, C_p(\boldsymbol{h}) \, = \, 0; \, h_p^{\mathrm{SV}} \, \geq \, 0 \; \forall \, p \in \mathcal{P}_{ij} \right] \; \Rightarrow \; h_p^{\mathrm{SV}} \, = \, 0 \; \forall \, p \, \in \, \mathcal{P}_{ij}; \tag{A.1}$$

• for all $(k,x) \in \mathcal{K} \times \mathcal{X}$, and all $(i,j) \in \mathcal{W}^{k,x}$:

$$\left[\sum_{p \in \mathcal{P}_{ij}} h_p^{k,\mathrm{x}} C_p(\boldsymbol{h}) \, = \, 0; \, h_p^{k,\mathrm{x}} \, \geq \, 0, \; \forall \, p \, \in \, \mathcal{P}_{ij} \right] \; \Rightarrow \; h_p^{k,\mathrm{AV}} \, = \, 0 \; \forall \, p \, \in \, \mathcal{P}_{ij}; \tag{A.2}$$

• for all $(k,x) \in \mathcal{K} \times \mathcal{X}$ and all $(s,i) \in \mathcal{D}^{k,x} \times \mathcal{O}^{k,x}$:

$$\left[\sum_{p \in \mathcal{P}_{si}} h_p^{k,x} \, C_p(\boldsymbol{h}) \, = \, 0; \, h_p^{k,x} \, \geq \, 0, \; \forall \, p \, \in \, \mathcal{P}_{si} \right] \; \Rightarrow \; h_p^{k,x} \, = \, 0 \; \forall \, p \, \in \, \mathcal{P}_{si}. \tag{A.3}$$

A2. A Technical Proposition

Proposition 4. Let $F(\boldsymbol{x},\boldsymbol{y}) \triangleq \begin{pmatrix} \Phi(\boldsymbol{x},\boldsymbol{y}) \\ \Psi(\boldsymbol{x},\boldsymbol{y}) \end{pmatrix}$ be a continuous function from $\mathbb{R}^{n+m}$ into itself. Let $\boldsymbol{X}$ and $\boldsymbol{Y}$ be closed convex sets in $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, with $\boldsymbol{X}$ being additionally bounded. If there exists a vector $\boldsymbol{y}^{\mathrm{ref}} \in \boldsymbol{Y}$ such that the solutions of the VI defined by the function $F^{\tau}(\boldsymbol{x},\boldsymbol{y}) \triangleq \begin{pmatrix} \Phi(\boldsymbol{x},\boldsymbol{y}) \\ \Psi(\boldsymbol{x},\boldsymbol{y}) + \tau(\boldsymbol{y} - \boldsymbol{y}^{\mathrm{ref}}) \end{pmatrix}$ on the set $\boldsymbol{X} \times \boldsymbol{Y}$ over all scalars $\tau > 0$ are bounded, then the VI $(F, \boldsymbol{X} \times \boldsymbol{Y})$ has a solution.

Proof. We apply the homotopy invariance principle of the degree Facchinei and Pang (2003)[Definition 2.11, part (A3)] of a continuous mapping to the homotopy

$$H(\boldsymbol{x},\boldsymbol{y},t) \, \triangleq \, \begin{pmatrix} \boldsymbol{x} - \Pi_{\boldsymbol{X}}(\boldsymbol{x} - \Phi(\boldsymbol{x},\boldsymbol{y})) \\ \boldsymbol{y} - \Pi_{\boldsymbol{Y}}(t(\boldsymbol{y} - \Psi(\boldsymbol{x},\boldsymbol{y})) + (1-t)\boldsymbol{y}^{\mathrm{ref}}) \end{pmatrix}, \quad \text{for } t \in [\,0,1\,],$$

where Π_S is the Euclidean projector onto a closed convex set S. It suffices that the set: $\bigcup_{t \in [\,0,1\,)} H(\bullet,\bullet,t)^{-1}(0)$ is bounded. It is easy to see that a pair $(\boldsymbol{x},\boldsymbol{y}) \in H(\bullet,\bullet,t)^{-1}(0)$ for some $t \in [\,0,1\,)$ if and only if $(\boldsymbol{x},\boldsymbol{y})$ is a solution of the VI $(F^{\tau}, \boldsymbol{X} \times \boldsymbol{Y})$ for $\tau = \dfrac{1-t}{t}$. By assumption, such pair $(\boldsymbol{x},\boldsymbol{y})$ is bounded. □

A3. Proof of Theorem 3

Proof. Suppose the costs are nonnegative and satisfy three weak positivity conditions (A.1) – (A.3). The later conditions are trivially satisfies when the path cost functions are positive. Then, at equilibrium, the following constraints hold as equalities: the fleet–demand feasibility constraint in the TNC module (1), the total demand satisfaction constraint in the traveler module (2), the right-hand side of delay complementarity conditions in the traffic module (6). In the sequel, we assume these equalities hold. In particular,

$$\sum_{s\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x} = D_{ij}^{k,x}, \quad \forall, x\in\mathcal{X}, (i,j)\in\mathcal{W}^{k,x}.$$

$$D_{ij}^{\mathrm{SV}} + \sum_{x\in\mathcal{X}}\sum_{k\in\mathcal{K}_{ij}^{x}} D_{ij}^{k,x} = D_{ij}, \quad \forall\,(i,j)\in\mathcal{W}$$

$$h_p^{\mathrm{SV}} = D_{ij} - \sum_{x\in\mathcal{X}}\sum_{k\in\mathcal{K}_{ij}^{x}} D_{ij}^{k,x}\,(=\, h_p^{\mathrm{SV}} - D_{ij}^{\mathrm{SV}}\,), \qquad \forall\,(i,j)\in\mathcal{W}$$

$$h_p^{k,x} = D_{ij}^{k,x}, \qquad \forall\, x\in\mathcal{X},\, k\in\mathcal{K},\, (i,j)\in\mathcal{W}^{k,x}$$

$$h_p^{k,x} = \sum_{j\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x}, \qquad \forall\, x\in\mathcal{X},\, k\in\mathcal{K},\, (s,i)\in\mathcal{D}^{k,x}\times\mathcal{O}^{k,x}$$

Throughout the analysis, we note that the customer waiting time, including the special case of (3), is a continuous function of the model's primary variables $z_{s,ij}^{k,x}$, $D_{ij}^{k,x}$ and $t_{s,i}^{k,x}$. As it turns out, the denominator in the above expression is always (in particular, at equilibrium) equal to zero when all vehicle-trips are balanced; rendering $w_{ij}^{k,x} = w_{ij}^{\max}$ and obscuring the effect of matching.

First-step — NCP formulation. We introduce multipliers for all the constraints in the MAGE model and write down the optimality conditions for each of the optimization problems in the modules. Most importantly, we invoke the postulate of proportional multipliers (denoted $\lambda_{ij}^{k,x}$) for the shared fleet-demand constraint: $\sum_{s\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x} \geq D_{ij}^{k,x}$ to define the variational equilibrium. We then concatenate all the optimality conditions into a large-scale nonlinear complementarity problem, which we denote as $(\mathrm{NCP})_{\mathrm{main}}$:

TNC Problem (1) optimality conditions

$$0 \leq z_{s,ij}^{k,\mathrm{AV}} \perp -\widetilde{R}_{s,ij}^{k,\mathrm{AV}} - \alpha_1^{k,x} t_{ij}^{k,\mathrm{AV}} + \beta_1^{k,x}(t_{s,i}^{k,\mathrm{AV}} + t_{ij}^{k,\mathrm{AV}}) - \beta_3^{k,\mathrm{AV}} t_{s,i}^{k,\mathrm{AV}} + \phi_s^{k,\mathrm{AV}} - \lambda_{ij}^{k,\mathrm{AV}} + (t_{s,i}^{k,\mathrm{AV}} + t_{ij}^{k,\mathrm{AV}})\,(\nu^k + \nu_{\mathrm{AV}}^k) \geq 0, \quad \forall\, k\in\mathcal{K},\, s\in\mathcal{D}^{k,\mathrm{AV}},\, (i,j)\in\mathcal{W}^{k,\mathrm{AV}}$$

$$0 \leq z_{s,ij}^{k,\mathrm{HV}} \perp -\widetilde{R}_{s,ij}^{k,\mathrm{HV}} - \alpha_1^{k,x} t_{ij}^{k,\mathrm{HV}} + \beta_1^{k,x}(t_{s,i}^{k,\mathrm{HV}} + t_{ij}^{k,\mathrm{HV}}) - \beta_3^{k,\mathrm{HV}} t_{s,i}^{k,\mathrm{HV}} + \phi_s^{k,\mathrm{HV}} - \lambda_{ij}^{k,\mathrm{HV}} + (t_{s,i}^{k,\mathrm{HV}} + t_{ij}^{k,\mathrm{HV}})\,\nu^k \geq 0, \quad \forall\, k\in\mathcal{K},\, s\in\mathcal{D}^{k,\mathrm{HV}},\, (i,j)\in\mathcal{W}^{k,\mathrm{HV}}$$

$$0 \leq \phi_s^{k,x} \perp \sum_{i\in\mathcal{O}^{k,x}|(i,s)\in\mathcal{W}^{k,x}} D_{is}^{k,x} - \sum_{(i,j)\in\mathcal{W}^{k,x}} z_{s,ij}^{k,x} \geq 0,\ \forall\, k\in\mathcal{K},\, x\in\mathcal{X},\, s\in\mathcal{D}^{k,x}$$

$$0 \le \lambda_{ij}^{k,x} \perp \sum_{s\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x} - D_{ij}^{k,x} \ge 0, \quad \forall\, k \in \mathcal{K},\, x \in \mathcal{X},\, (i,j) \in \mathcal{W}^{k,x}$$

$$0 \le \nu^{k} \perp N^{k} - \sum_{x\in\mathcal{X}} \sum_{(i,j)\in\mathcal{W}^{k,x}} \sum_{s\in\mathcal{D}^{k,x}} (t_{s,i}^{k,x} + t_{ij}^{k,x}) z_{s,ij}^{k,x} \ge 0, \quad \forall\, k \in \mathcal{K}$$

$$0 \le \nu_{\mathrm{AV}}^{k} \perp \mu_{\mathrm{AV}}^{\mathrm{cap}} N^{k} - \sum_{(i,j)\in\mathcal{W}^{k,\mathrm{AV}}} \sum_{s\in\mathcal{D}^{k,\mathrm{AV}}} (t_{s,i}^{k,\mathrm{AV}} + t_{ij}^{k,\mathrm{AV}}) z_{s,ij}^{k,\mathrm{AV}} \ge 0, \quad \forall\, k \in \mathcal{K}$$

Traveler Problem (2) **optimality conditions**

$$0 \le D_{ij}^{\mathrm{SV}} \perp \alpha_1^{\mathrm{SV}} t_{ij}^{\mathrm{SV}} + \alpha_2^{\mathrm{SV}} d_{ij}^{0} - \sigma_{ij} \ge 0, \quad \forall\, (i,j) \in \mathcal{W}$$

$$0 \le D_{ij}^{k,x} \perp F_{ij}^{k,x} + \alpha_1^{k,x}(t_{ij}^{k,x} - t_{ij}^{0}) + \alpha_2^{k,x} d_{ij}^{0} + \gamma_1^{k,x} t_{ij}^{k,x} + \gamma_2^{k,x} w_{ij}^{k,x} - \sigma_{ij} + \lambda_{ij}^{k,x} \ge 0, \quad \forall\, (k,x) \in \mathcal{K}\times\mathcal{X},\, (i,j) \in \mathcal{W}^{k,x}$$

$$0 \le \sigma_{ij} \perp D_{ij}^{\mathrm{SV}} + \sum_{k\in\mathcal{K}} \sum_{x\in\mathcal{X}} D_{ij}^{k,x} - D_{ij} \ge 0, \quad \forall\, (i,j) \in \mathcal{W}$$

Customer waiting (4)**reformulation**

$$0 \le \theta_{s,ij}^{k,x} \perp \theta_{s,ij}^{k,x} \sum_{s'\in\mathcal{D}^{k,x}} z_{s',ij}^{k,x} - z_{s,ij}^{k,x} + \zeta_{s,ij}^{k,x} \ge 0, \quad \forall\, k \in \mathcal{K},\, x \in \mathcal{X},\, s \in \mathcal{D}^{k,x},\, (i,j) \in \mathcal{W}^{k,x}$$

$$0 \le \zeta_{s,ij}^{k,x} \perp 1 - \theta_{s,ij}^{k,x} \ge 0, \quad \forall\, k \in \mathcal{K},\, x \in \mathcal{X},\, s \in \mathcal{D}^{k,x},\, (i,j) \in \mathcal{W}^{k,x}$$

Traffic conditions (6)

$$0 \le t_{ij}^{\mathrm{SV}} \perp \sum_{p\in\mathcal{P}_{ij}} h_p^{\mathrm{SV}} - D_{ij}^{\mathrm{SV}} \ge 0, \quad \forall\, (i,j) \in \mathcal{W}$$

$$0 \le t_{ij}^{k,x} \perp \sum_{p\in\mathcal{P}_{ij}} h_p^{k,x} - D_{ij}^{k,x} \ge 0, \quad \forall\, x \in \mathcal{X},\, k \in \mathcal{K},\, (i,j) \in \mathcal{W}^{k,x}$$

$$0 \le t_{s,i}^{k,x} \perp \sum_{p\in\mathcal{P}_{si}} h_p^{k,x} - \sum_{j\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x} \ge 0, \quad \forall\, x \in \mathcal{X},\, k \in \mathcal{K},\, (s,i) \in \mathcal{D}^{k,x}\times\mathcal{O}^{k,x}$$

$$0 \le h_p^{\mathrm{SV}} \perp C_p(\boldsymbol{h}) - t_{ij}^{\mathrm{SV}} \ge 0, \quad \forall\, (i,j) \in \mathcal{W},\, p \in \mathcal{P}_{ij}$$

$$0 \le h_p^{k,x} \perp \mu_1^{k,x} C_p(\boldsymbol{h}) - t_{ij}^{k,x} \ge 0, \quad \forall\, x \in \mathcal{X},\, k \in \mathcal{K},\, (i,j) \in \mathcal{W}^{k,x},\, p \in \mathcal{P}_{ij}$$

$$0 \le h_p^{k,x} \perp \mu_2^{k,x} C_p(\boldsymbol{h}) - t_{s,i}^{k,x} \ge 0, \quad \forall\, x \in \mathcal{X},\, k \in \mathcal{K},\, (s,i) \in \mathcal{D}^{k,x}\times\mathcal{O}^{k,x},\, p \in \mathcal{P}_{si}$$

We define the polyhedron $\mathcal{Z}$ consisting of nonnegative tuples $\{z_{s,ij}^{k,x}, D_{ij}^{k,x}, D_{ij}^{\mathrm{SV}}\}$ for all (k,x) in $\mathcal{K}\times\mathcal{X}$, $s \in \mathcal{D}^{k,x}$, and $(i,j) \in \mathcal{W}$ satisfying the following flow conservation, fleet demand, and total demand constraints, respectively:

$$\sum_{(i,j)\in\mathcal{W}^{k,x}} z_{s,ij}^{k,x} \le \sum_{i\in\mathcal{O}^{k,x}|(i,s)\in\mathcal{W}^{k,x}} D_{is}^{k,x}, \quad \forall k \in \mathcal{K},\, x \in \mathcal{X},\, s \in \mathcal{D}^{k,x}$$

$$\sum_{s\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x} \ge D_{ij}^{k,x}, \quad \forall\, k \in \mathcal{K},\, x \in \mathcal{X},\, (i,j) \in \mathcal{W}^{k,x}$$

$$D_{ij}^{\mathrm{SV}} + \sum_{x\in\mathcal{X}} \sum_{k\in\mathcal{K}_{ij}^{x}} D_{ij}^{k,x} = D_{ij}, \quad \forall\, (i,j) \in \mathcal{W}.$$

Note that the above inequalities must hold as equalities for all tuples $\boldsymbol{z} \in \mathcal{Z}$ which are obviously bounded because of the last constraint.

Second step—bounds and a VI formulation: We establish upper bounds for the primary variables of the model that will be used in the VI formulation; these variables are:

$$\left\{ z_{s,ij}^{k,x},\, D_{ij}^{k,x},\, D_{ij}^{\mathrm{SV}},\, h_p^{k,x},\, h_p^{\mathrm{SV}},\, t_{ij}^{k,x},\, t_{s,i}^{k,x},\, t_{ij}^{\mathrm{SV}} \right\}, \tag{C.1}$$

with $\left\{ z_{s,ij}^{k,x}, D_{ij}^{k,x}, D_{ij}^{\mathrm{SV}} \right\}$ belonging to the polyhedron $\mathcal{Z}$, whose elements we denote $\boldsymbol{z}$, and with $\left\{ h_p^{k,x}, h_p^{\mathrm{SV}}, t_{ij}^{k,x}, t_{s,i}^{k,x}, t_{ij}^{\mathrm{SV}} \right\}$ satisfying (6). This is clear for the former family of variables. It is also clear for the path flow variables because

$$\sum_{p \in \mathcal{P}_{ij}} h_p^{\mathrm{SV}} \,=\, D_{ij}^{\mathrm{SV}}, \quad \sum_{p \in \mathcal{P}_{ij}} h_p^{k,x} \,=\, D_{ij}^{k,x}, \;\; \text{and} \;\; \sum_{p \in \mathcal{P}_{si}} h_p^{k,x} \,=\, \sum_{j \in \mathcal{D}^{k,x}} z_{s,ij}^{k,x}.$$

Indeed, by letting

$$\bar{h} > \max\left\{ \max_{(i,j)\in\mathcal{W}} D_{ij},\; \max_{(k,x)\in\mathcal{K}\times\mathcal{X}} \left(\max_{(i,j)\in\mathcal{W}^{k,x}} \max_{p\in\mathcal{P}_{ij}} h_p^{k,x},\; \max_{(s,i)\in\mathcal{D}^{k,x}\times\mathcal{O}^{k,x}} \max_{p\in\mathcal{P}_{si}} h_p^{k,x} \right),\; \max_{(i,j)\in\mathcal{W}} \max_{p\in\mathcal{P}_{ij}} h_p^{\mathrm{SV}} \right\},$$

it then follows that $\bar{h}$ is a (strict) upper bound of all tuples $\boldsymbol{z} \in \mathcal{Z}$ and all path tuples $\{h_p^{k,x}, h_p^{\mathrm{SV}}\}$. For the travel time variables, we have, as an illustration,

$$t_{i'j'}^{k',x'} \;\le\; \max_{(k,x)\in\mathcal{K}\times\mathcal{X}} \max_{(i,j)\in\mathcal{W}^{k,x}} \max_{p\in\mathcal{P}_{ij}} \max_{0\le \boldsymbol{h} \le \bar{h}\mathbf{1}} \mu^{k,x}\, C_p(\boldsymbol{h}).$$

Therefore, letting $\bar{t}$ be a (strict) upper bound of all the path costs (i.e., the right-hand maxima in the above expression), we obtain

$$\bar{t} \;>\; \max\left\{ \max_{(k,x)\in\mathcal{K}\times\mathcal{X}} \left(\max_{(i,j)\in\mathcal{W}^{k,x}} t_{ij}^{k,x},\; \max_{p\in\mathcal{P}_{si}} t_{s,i}^{k,x} \right),\; \max_{(i,j)\in\mathcal{W}} t_{ij}^{\mathrm{SV}} \right\}.$$

Based on the above bounds, we may now define the VI that will be shown to be equivalent to the NCP formulation of the traffic problem and which is the cornerstone for proving the existence of its solution. The VI variables are the tuples (C.1) along with the auxiliary variables $\boldsymbol{\theta} \triangleq \left\{ \{\theta_{ij}^{k,x}\}_{(i,j)\in\mathcal{W}^{k,x}} \right\}_{k\in\mathcal{K}}^{x\in\mathcal{X}}$, and $\boldsymbol{\nu} \triangleq \left\{ \nu_{\mathrm{AV}}^k,\, \nu^k \right\}_{k\in\mathcal{K}}$, which altogether belong to the feasible set: $\boldsymbol{V} \triangleq \mathcal{Z} \times \mathcal{H} \times [\,0,1\,]^K \times \mathbb{R}_+^{2|\mathcal{K}|}$, where $\mathcal{H}$ consists of all nonnegative tuples $\boldsymbol{h} \triangleq \left\{ h_p^{k,x}, h_p^{\mathrm{SV}}, t_{ij}^{k,x}, t_{s,i}^{k,x}, t_{ij}^{\mathrm{SV}} \right\}$ with upper bounds of $\bar{h}$ and $\bar{t}$, respectively, and K is the total number of the θ-variables. Thus, all the multipliers of the linear constraints are hidden in the VI formulation, but the multipliers ν^k and ν_{AV}^k for the (nonlinear) total fleet capacity and AV capacity constrains are kept explicitly. [This is a major departure from the analysis in Ban et al. (2019) where the nonlinear constraints are penalized.] We define the block partitioned function $\boldsymbol{F}(\boldsymbol{z}, \boldsymbol{h}, \boldsymbol{\theta}, \boldsymbol{\nu})$, whose blocks are arranged in the order consistent with its arguments,

- $\left(\begin{array}{l} \left(\begin{array}{l} -\widetilde{R}_{s,ij}^{k,\mathrm{AV}} - \alpha_1^{k,\mathrm{AV}} t_{ij}^{k,\mathrm{AV}} + \beta_1^{k,\mathrm{AV}} (t_{s,i}^{k,\mathrm{AV}} + t_{ij}^{k,\mathrm{AV}}) \\ -\beta_3^{k,\mathrm{AV}} t_{s,i}^{k,\mathcal{AV}} + (t_{s,i}^{k,\mathrm{AV}} + t_{ij}^{k,\mathrm{AV}})(\nu^k + \nu_{\mathrm{AV}}^k) \end{array} \right)_{k\in\mathcal{K};\, s\in\mathcal{D}^{k,\,AV},}^{(i,j)\in\mathcal{W}^{(k,\mathrm{AV})}} \\ \left(\begin{array}{l} -\widetilde{R}_{s,ij}^{k,\mathrm{HV}} - \alpha_1^{k,\mathrm{HV}} t_{ij}^{k,\mathrm{HV}} + \beta_1^{k,\mathrm{HV}} (t_{s,i}^{k,\mathrm{HV}} + t_{ij}^{k,\mathrm{HV}}) \\ -\beta_3^{k,\mathrm{HV}} t_{s,i}^{k,\mathcal{HV}} + (t_{s,i}^{k,\mathrm{HV}} + t_{ij}^{k,\mathrm{HV}})\nu^k \end{array} \right)_{k\in\mathcal{K};\, s\in\mathcal{D}^{k,\,HV},}^{(i,j)\in\mathcal{W}^{k,\mathrm{HV}}} \end{array} \right)$ the $\left(z_{ij}^{k,x}\right)$ block of $\boldsymbol{F}(\boldsymbol{z},\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{\nu})$
- $\left(\begin{array}{l} \left(\begin{array}{l} F_{ij}^{k,x} + \alpha_1^{k,x}(t_{ij}^{k,x} - t_{ij}^0) + \alpha_2^{k,x} d_{ij}^0 + \\ \gamma_1^{k,x} t_{ij}^{k,x} + \gamma_2^{k,x} w_{ij}^{k,x} \end{array} \right)_{(i,j)\in\mathcal{W}^{k,x}}^{(k,x)\in\mathcal{K}\times\mathcal{X}} \\ \left(\alpha_1^{\mathrm{SV}} t_{ij}^{\mathrm{SV}} + \alpha_2^{\mathrm{SV}} d_{ij}^0\right)_{(i,j)\in\mathcal{W}} \end{array} \right)$ the $\left(D_{ij}^{k,x}\right)$ block of $\boldsymbol{F}(\boldsymbol{z},\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{\nu})$
- $\left(\begin{array}{l} \left(\mu_1^{k,x} C_p(\boldsymbol{h}) - t_{ij}^{k,x}\right)_{(i,j)\in\mathcal{W},\, p\in\mathcal{P}_{ij}}^{(k,x)\in\mathcal{K}\times\mathcal{X}} \\ \left(\mu_2^{k,x} C_p(\boldsymbol{h}) - t_{s,i}^{k,x}\right)_{(s,i)\in\mathcal{D}\times\mathcal{O},\, p\in\mathcal{P}_{si}}^{(k,x)\in\mathcal{K}\times\mathcal{X}} \\ \left(C_p(\boldsymbol{h}) - t_{ij}^{\mathrm{SV}}\right)_{(i,j)\in\mathcal{W},\, p\in\mathcal{P}_{ij}} \end{array} \right)$ the $\left(h_p^{k,x}\right)$ block of $\boldsymbol{F}(\boldsymbol{z},\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{\nu})$
- $\left(\begin{array}{l} \left(\sum_{p\in\mathcal{P}_{ij}} h_p^{k,x} - D_{ij}^{k,x} \right)_{(i,j)\in\mathcal{W}^{k,x}}^{(k,x)\in\mathcal{K}\times\mathcal{X}} \\ \left(\sum_{p\in\mathcal{P}_{si}} h_p^{k,x} - \sum_{j\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x} \right)_{(s,i)\in\mathcal{D}^{k,x}\times\mathcal{O}^{k,x}}^{(k,x)\in\mathcal{K}\times\mathcal{X}} \\ \left(\sum_{p\in\mathcal{P}_{ij}} h_p^{\mathrm{SV}} - D_{ij}^{\mathrm{SV}} \right)_{(i,j)\in\mathcal{W}} \end{array} \right)$ the $\left(t_{ij}^{k,x}\right)$ block of $\boldsymbol{F}(\boldsymbol{z},\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{\nu})$
- $\left(\left(\theta_{s,ij}^{k,x} \sum_{s'\in\mathcal{D}^{k,x}} z_{s',ij}^{k,x} - z_{s,ij}^{k,x} \right)_{s\in\mathcal{D}^{k,x}}^{(i,j)\in\mathcal{W}^{k,x}} \right)_{k\in\mathcal{K}}^{x\in\mathcal{X}}$ the $\left(\theta_{s,ij}^{k,x}\right)$ block of $\boldsymbol{F}(\boldsymbol{z},\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{\nu})$
- $\left(\begin{array}{l} \left(N^k - \sum_{x\in\mathcal{X}} \sum_{(i,j)\in\mathcal{W}^{k,x}} \sum_{s\in\mathcal{D}^{k,x}} (t_{s,i}^{k,x} + t_{ij}^{k,x}) z_{s,ij}^{k,x} \right)_{k\in\mathcal{K}} \\ \left(\mu_{\mathrm{AV}}^{\mathrm{cap}} N^k - \sum_{(i,j)\in\mathcal{W}^{k,\mathrm{AV}}} \sum_{s\in\mathcal{D}^{k,\mathrm{AV}}} (t_{s,i}^{k,\mathrm{AV}} + t_{ij}^{k,\mathrm{AV}}) z_{s,ij}^{k,\mathrm{AV}} \right)_{k\in\mathcal{K}} \end{array} \right)$ the $\left(\nu^k,\, \nu_{\mathrm{AV}}^k\right)$ block of $\boldsymbol{F}(\boldsymbol{z},\boldsymbol{h},\boldsymbol{\theta},\boldsymbol{\nu})$

We have the following main result, which has two parts: the first part is the VI defined by the pair $\boldsymbol{F}$ and $\boldsymbol{V}$ is equivalent to the $(\mathrm{NCP})_{\mathrm{main}}$ formulation of the mixed-fleet transportation system; and the second part is the existence of a solution to the VI (Theorem 3), and thus a normalized equilibrium of the mixed-fleet transportation system. It is important to point

out that this result requires minimal assumptions; in particular, there is no restriction on the TNCs' available fleet sizes N^k except for their positivity. This is a significant improvement of the existence result compared to that of the previous model in Ban et al. (2019) which has a restriction on such fleets; see Lemma 3 therein.

We establish the first part by showing that the KKT conditions of the VI whose feasible set $\boldsymbol{V}$ is polyhedral is equivalent to the said NCP. The only difference between the former conditions and the latter is in the complementarity conditions of the t and h-variables. Specifically, the complementarity conditions of these travel time and path flow variables need to be modified to include the multipliers—denoted by u and v below—of the bound constraints, resulting in the following relevant modified constraints:

$$\left.\begin{array}{l} 0 \le t_{ij}^{\mathrm{SV}} \perp \displaystyle\sum_{p\in\mathcal{P}_{ij}} h_p^{\mathrm{SV}} - D_{ij}^{\mathrm{SV}} + u_{ij}^{\mathrm{SV}} \ge 0 \\ 0 \le u_{ij}^{\mathrm{SV}} \perp \bar{t} - t_{ij}^{\mathrm{SV}} \ge 0 \end{array}\right\} \quad \forall\, x \in \mathcal{X},\, k \in \mathcal{K},\, (i,j) \in \mathcal{W}^{k,x}$$

$$\left.\begin{array}{l} 0 \le t_{ij}^{k,x} \perp \displaystyle\sum_{p\in\mathcal{P}_{ij}} h_p^{k,x} - D_{ij}^{k,x} + u_{ij}^{k,x} \ge 0 \\ 0 \le u_{ij}^{k,x} \perp \bar{t} - t_{ij}^{k,x} \ge 0 \end{array}\right\} \quad \forall\, x \in \mathcal{X},\, k \in \mathcal{K},\, (i,j) \in \mathcal{W}^{k,x}$$

$$\left.\begin{array}{l} 0 \le t_{s,i}^{k,x} \perp \displaystyle\sum_{p\in\mathcal{P}_{si}} h_p^{k,x} - \sum_{j\in\mathcal{D}^{k,x}} z_{s,ij}^{k,x} + u_{s,i}^{k,x} \ge 0 \\ 0 \le u_{s,i}^{k,x} \perp \bar{t} - t_{s,i}^{k,x} \ge 0 \end{array}\right\} \forall\, x \in \mathcal{X},\, k \in \mathcal{K},\, (s,i) \in \mathcal{D}^{k,x} \times \mathcal{O}^{k,x}$$

$$\left.\begin{array}{l} 0 \le h_p^{\mathrm{SV}} \perp C_p(\boldsymbol{h}) - t_{ij}^{\mathrm{SV}} + v_p^{\mathrm{SV}} \ge 0 \\ 0 \le v_p^{\mathrm{SV}} \perp \bar{h} - h_p^{\mathrm{SV}} \ge 0 \end{array}\right\} \quad \forall\, (i,j) \in \mathcal{W},\, p \in \mathcal{P}_{ij}$$

$$\left.\begin{array}{l} 0 \le h_p^{k,x} \perp \mu_1^{k,x} C_p(\boldsymbol{h}) - t_{ij}^{k,x} + v_p^{k,x} \ge 0 \\ 0 \le v_p^{k,x} \perp \bar{h} - h_p^{k,x} \ge 0 \end{array}\right\} \quad \forall\, (i,j) \in \mathcal{W},\, p \in \mathcal{P}_{ij},\, x \in \mathcal{X},\, k \in \mathcal{K}$$

$$\left.\begin{array}{l} 0 \le h_p^{k,x} \perp \mu_2^{k,x} C_p(\boldsymbol{h}) - t_{s,i}^{k,x} + v_p^{k,x} \ge 0 \\ 0 \le v_p^{k,x} \perp \bar{h} - h_p^{k,x} \ge 0 \end{array}\right\} \quad \forall\, (s,i) \in \mathcal{D} \times \mathcal{O},\, p \in \mathcal{P}_{si},\, x \in \mathcal{X},\, k \in \mathcal{K}$$

We claim that all the multipliers u's and v's of the bound constraints are equal to zero. In fact, suppose that $u_{ij}^{k,x} > 0$ for some $(k,x) \in \mathcal{K} \times \mathcal{X}$ and $(i,j) \in \mathcal{W}^{k,x}$. We then have $t_{ij}^{k,x} = \bar{t} > 0$. Thus, by complementarity, it follows that

$$\bar{h} \,>\, D_{ij}^{k,x} \,=\, \sum_{p\in\mathcal{P}_{ij}} h_p^{k,x} + u_{ij}^{k,x} \,>\, h_p^{k,x}, \quad \forall\, p \,\in\, \mathcal{P}_{ij}.$$

Hence, by complementarity, $v_p^{k,x} = 0$ for all $p \in \mathcal{P}_{ij}$. This yields

$$t_{ij}^{k,x} \,=\, \mu_1^{k,x}\, C_p(\boldsymbol{h}) \,<\, \bar{t}$$

which contradicts the choice of $t_{ij}^{k,x}$. Similarly, all the other bound multipliers are zero.

Theorem 3 is an immediate consequence of Proposition A2 under the identifications: $\boldsymbol{x}$ being the tuple $(\boldsymbol{z}, \boldsymbol{h}, \boldsymbol{\theta})$, $\boldsymbol{y} = \boldsymbol{\nu}$, $\boldsymbol{X} = \mathcal{Z} \times \mathcal{H} \times [\,0,\,1\,]^K$, $\boldsymbol{Y}$ being $\mathbb{R}_+^{2|\mathcal{K}|}$ and $\boldsymbol{y}^{\mathrm{ref}}$ being the origin, provided that we can show that the tuples $\{\boldsymbol{\nu}^\tau\}$ are bounded for $\tau > 0$, where each $\boldsymbol{\nu}^\tau$ satisfies

$$
\begin{aligned}
0 \;\le\; \nu_{\mathrm{AV}}^{\tau;k} \;\perp\; & \tau\, \nu_{\mathrm{AV}}^{\tau;k} + \mu_{\mathrm{AV}}^{\mathrm{cap}} N^k - \\
& \sum_{(i,j) \in \mathcal{W}^{k,\mathrm{AV}}} \sum_{s \in \mathcal{D}^{k,\mathrm{AV}}} (t_{s,i}^{\tau;k,\mathrm{AV}} + t_{ij}^{\tau;k,\mathrm{AV}}) z_{s,ij}^{\tau;k,\mathrm{AV}} \;\ge\; 0, \quad \forall\, k \in \mathcal{K} \\
0 \;\le\; \nu^{\tau;k} \;\perp\; & \tau\, \nu^{\tau;k} + N^k - \\
& \sum_{x \in \mathcal{X}} \sum_{(i,j) \in \mathcal{W}^{k,x}} \sum_{s \in \mathcal{D}^{k,x}} (t_{s,i}^{\tau;k,x} + t_{ij}^{\tau;k,x}) z_{ij}^{\tau;k,x} \;\ge\; 0, \quad \forall\, k \in \mathcal{K}
\end{aligned}
\tag{C.2}
$$

for some (bounded) tuple $(\boldsymbol{z}^\tau, \boldsymbol{h}^\tau, \boldsymbol{\theta}^\tau) \in \mathcal{Z} \times \mathcal{H} \times [\,0,1\,]^K$, which along with suitable multipliers, satisfy the $\mathrm{NCP}_{\mathrm{main}}$. For convenience of reference, we re-write in full this NCP with all inequalities along with their respective (nonnegative) multipliers:

$$
\begin{aligned}
0 \;\le\; z_{s,ij}^{\tau;k,\mathrm{AV}} \;\perp\; & -\widetilde{R}_{s,ij}^{k,\mathrm{AV}} - \alpha_1^{k,x} t_{ij}^{\tau;k,\mathrm{AV}} + \beta_1^{k,x} (t_{s,i}^{\tau;k,\mathrm{AV}} + t_{ij}^{\tau;k,\mathrm{AV}}) - \beta_3^{k,\mathrm{AV}} t_{s,i}^{\tau;k,\mathrm{AV}} + \\
& \phi_s^{\tau;k,\mathrm{AV}} - \lambda_{ij}^{\tau;k,\mathrm{AV}} + (t_{s,i}^{\tau;k,\mathrm{AV}} + t_{ij}^{\tau;k,\mathrm{AV}})\, (\nu^{\tau;k} + \nu_{\mathrm{AV}}^{\tau;k}) \;\ge\; 0, \\
& \qquad \forall\, k \in \mathcal{K},\, s \in \mathcal{D}^{k,\mathrm{AV}},\, (i,j) \in \mathcal{W}^{k,\mathrm{AV}} \\
0 \;\le\; z_{s,ij}^{\tau;k,\mathrm{HV}} \;\perp\; & -\widetilde{R}_{s,ij}^{k,\mathrm{HV}} - \alpha_1^{k,x} t_{ij}^{\tau;k,\mathrm{HV}} + \beta_1^{k,x} (t_{s,i}^{\tau;k,\mathrm{HV}} + t_{ij}^{\tau;k,\mathrm{HV}}) - \beta_3^{k,\mathrm{HV}} t_{s,i}^{\tau;k,\mathrm{HV}} + \\
& \phi_s^{\tau;k,\mathrm{HV}} - \lambda_{ij}^{\tau;k,\mathrm{HV}} + (t_{s,i}^{\tau;k,\mathrm{HV}} + t_{ij}^{\tau;k,\mathrm{HV}})\, \nu^{\tau;k} \;\ge\; 0, \\
& \qquad \forall\, k \in \mathcal{K},\, s \in \mathcal{D}^{k,\mathrm{HV}},\, (i,j) \in \mathcal{W}^{k,\mathrm{HV}}
\end{aligned}
$$

$$
0 \;\le\; \phi_s^{\tau;k,x} \;\perp\; \sum_{i \in \mathcal{O}^{k,x}} D_{is}^{\tau;k,x} - \sum_{(i,j) \in \mathcal{W}^{k,x}} z_{s,ij}^{\tau;k,x} \;\ge\; 0,\; \forall\, k \in \mathcal{K},\, x \in \mathcal{X},\, s \in \mathcal{D}^{k,x}
$$

$$
0 \;\le\; \lambda_{ij}^{\tau;k,x} \;\perp\; \sum_{s \in \mathcal{D}^{k,x}} z_{s,ij}^{\tau;k,x} - D_{ij}^{\tau;k,x} \;\ge\; 0, \quad \forall\, k \in \mathcal{K},\, x \in \mathcal{X},\, (i,j) \in \mathcal{W}^{k,x}
$$

$$
\begin{aligned}
0 \;\le\; D_{ij}^{\tau;k,x} \;\perp\; & F_{ij}^{k,x} + \alpha_1^{k,x} (t_{ij}^{\tau;k,x} - t_{ij}^0) + \alpha_2^{k,x} d_{ij}^0 + \gamma_1^{k,x} t_{ij}^{\tau;k,x} + \gamma_2^{k,x} w_{ij}^{\tau;k,x} \\
& - \sigma_{ij}^\tau + \lambda_{ij}^{\tau;k,x} \;\ge\; 0, \qquad \forall\, (k,x) \in \mathcal{K} \times \mathcal{X},\, (i,j) \in \mathcal{W}^{k,x}
\end{aligned}
$$

$$
0 \;\le\; D_{ij}^{\tau;\mathrm{SV}} \;\perp\; \alpha_1^{\mathrm{SV}} t_{ij}^{\tau;\mathrm{SV}} + \alpha_2^{\mathrm{SV}} d_{ij}^0 - \sigma_{ij}^\tau \;\ge\; 0, \quad \forall\, (i,j) \in \mathcal{W}
$$

$$
0 \le \sigma_{ij}^\tau \;\perp\; D_{ij}^{\tau;\mathrm{SV}} + \sum_{k \in \mathcal{K}} \sum_{x \in \mathcal{X}} D_{ij}^{\tau;k,x} - D_{ij} \;\ge\; 0, \quad \forall\, (i,j) \in \mathcal{W}
$$

$$
0 \;\le\; t_{ij}^{\tau;\mathrm{SV}} \;\perp\; \sum_{p \in \mathcal{P}_{ij}} h_p^{\tau;\mathrm{SV}} - D_{ij}^{\tau;\mathrm{SV}} \;\ge\; 0, \qquad \forall\, (i,j) \in \mathcal{W}
$$

$$
0 \;\le\; t_{ij}^{\tau;k,x} \;\perp\; \sum_{p \in \mathcal{P}_{ij}} h_p^{\tau;k,x} - D_{ij}^{\tau;k,x} \;\ge\; 0, \qquad \forall\, x \in \mathcal{X},\, k \in \mathcal{K},\, (i,j) \in \mathcal{W}^{k,x}
$$

$$
0 \;\le\; t_{s,i}^{\tau;k,x} \;\perp\; \sum_{p \in \mathcal{P}_{si}} h_p^{\tau;k,x} - \sum_{j \in \mathcal{D}^{k,x}} z_{s,ij}^{\tau;k,x} \;\ge\; 0, \quad \forall\, x \in \mathcal{X},\, k \in \mathcal{K},\, (s,i) \in \mathcal{D}^{k,x} \times \mathcal{O}^{k,x}
$$

$$0 \le h_p^{\tau;\mathrm{SV}} \perp C_p(\boldsymbol{h}^\tau) - t_{ij}^{\tau;\mathrm{SV}} \ge 0, \qquad \forall\,(i,j) \in \mathcal{W},\, p \in \mathcal{P}_{ij}$$

$$0 \le h_p^{\tau;k,x} \perp \mu_1^{k,x} C_p(\boldsymbol{h}^\tau) - t_{ij}^{\tau;k,x} \ge 0, \quad \forall\,(i,j) \in \mathcal{W},\, p \in \mathcal{P}_{ij},\, x \in \mathcal{X},\, k \in \mathcal{K}$$

$$0 \le h_p^{\tau;k,x} \perp \mu_2^{k,x} C_p(\boldsymbol{h}^\tau) - t_{s,i}^{\tau;k,x} \ge 0, \quad \forall\,(s,i) \in \mathcal{D} \times \mathcal{O},\, p \in \mathcal{P}_{si},\, x \in \mathcal{X},\, k \in \mathcal{K}$$

We sum up the complementarity conditions for a sequence of positive scalars $\{\tau_n\}$ up to the σ_{ij}-complementarities and mark all the terms that can be canceled when these equations are added up (in what follows, we use the superscript "n" as a short-hand for τ_n in the variables):

$$\sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{D}^{k,\mathrm{AV}}} \sum_{(i,j)\in\mathcal{W}^{k,\mathrm{AV}}} \left\{ \begin{array}{l} -\widetilde{R}_{s,ij}^{k,\mathrm{AV}} z_{s,ij}^{n;k,\mathrm{AV}} + (\beta_1^{k,x} - \alpha_1^{k,\mathrm{AV}}) t_{ij}^{n;k,\mathrm{AV}} z_{s,ij}^{n;k,\mathrm{AV}} + \\ (\beta_1^{k,\mathrm{AV}} - \beta_3^{k,\mathrm{AV}}) t_{s,i}^{n;k,\mathrm{AV}} z_{s,ij}^{n;k,\mathrm{AV}} + \cancel{\phi_s^{n;k,\mathrm{AV}} z_{s,ij}^{n;k,\mathrm{AV}}} \\ \cancel{-\lambda_{ij}^{n;k,\mathrm{AV}} z_{s,ij}^{n;k,\mathrm{AV}}} + (t_{s,i}^{n;k,\mathrm{AV}} + t_{ij}^{n;k,\mathrm{AV}})(\nu_{\mathrm{AV}}^{n;k} + \nu^{n;k}) z_{s,ij}^{n;k,\mathrm{AV}} \end{array} \right\} +$$

$$\sum_{k\in\mathcal{K}} \sum_{s\in\mathcal{D}^{k,\mathrm{HV}}} \sum_{(i,j)\in\mathcal{W}^{k,\mathrm{HV}}} \left\{ \begin{array}{l} -\widetilde{R}_{s,ij}^{k,\mathrm{HV}} z_{s,ij}^{n;k,\mathrm{HV}} + (\beta_1^{k,\mathrm{HV}} - \alpha_1^{k,\mathrm{HV}}) t_{ij}^{n;k,\mathrm{HV}} z_{s,ij}^{n;k,\mathrm{HV}} + \\ (\beta_1^{k,\mathrm{HV}} - \beta_3^{k,\mathrm{HV}}) t_{s,i}^{n;k,\mathrm{HV}} z_{s,ij}^{n;k,\mathrm{HV}} + \cancel{\phi_s^{n;k,\mathrm{HV}} z_{s,ij}^{n;k,\mathrm{HV}}} \\ \cancel{-\lambda_{ij}^{n;k,\mathrm{HV}} z_{s,ij}^{n;k,\mathrm{HV}}} + (t_{s,i}^{n;k,\mathrm{HV}} + t_{ij}^{n;k,\mathrm{HV}}) \nu^{n;k} z_{s,ij}^{n;k,\mathrm{HV}} \end{array} \right\} +$$

$$\sum_{k\in\mathcal{K}} \sum_{x\in\mathcal{X}} \sum_{s\in\mathcal{D}^{k,x}} \left\{ \underbrace{\sum_{i\in\mathcal{O}^{k,x}} D_{is}^{n;k,x} \phi_s^{n;k,x}}_{\text{nonnegative}} - \cancel{\sum_{(i,j)\in\mathcal{W}^{k,x}} z_{s,ij}^{n;k,x} \phi_s^{n;k,x}} \right\} +$$

$$\sum_{k\in\mathcal{K}} \sum_{x\in\mathcal{X}} \sum_{(i,j)\in\mathcal{W}^{k,x}} \left\{ \cancel{\sum_{s\in\mathcal{D}^{k,x}} z_{s,ij}^{n;k,x} \lambda_{ij}^{n;k,x}} - \cancel{D_{ij}^{n;k,x} \lambda_{ij}^{n;k,x}} \right\} +$$

$$\sum_{(i,j)\in\mathcal{W}} \sum_{x\in\mathcal{X}} \sum_{k\in\mathcal{K}_{ij}^x} \left\{ \begin{array}{l} F_{ij}^{k,x} D_{ij}^{n;k,x} + \alpha_1^{k,x} (t_{ij}^{n;k,x} - t_{ij}^0) D_{ij}^{n;k,x} + \\ \alpha_2^{k,x} d_{ij}^0 D_{ij}^{n;k,x} + \gamma_1^{k,x} t_{ij}^{n;k,x} D_{ij}^{n;k,x} + \gamma_2^{k,x} w_{ij}^{n;k,x} D_{ij}^{n;k,x} - \\ \cancel{\sigma_{ij}^n D_{ij}^{n;k,x}} + \cancel{\lambda_{ij}^{n;k,x} D_{ij}^{n;k,x}} \end{array} \right\} +$$

$$\sum_{(i,j)\in\mathcal{W}} \left\{ \alpha_1^{\mathrm{SV}} t_{ij}^{n;\mathrm{SV}} D_{ij}^{n;\mathrm{SV}} + \alpha_2^{\mathrm{SV}} d_{ij}^0 D_{ij}^{n;\mathrm{SV}} - \cancel{\sigma_{ij}^n D_{ij}^{n;\mathrm{SV}}} \right\} +$$

$$\sum_{(i,j)\in\mathcal{W}} \left\{ \cancel{D_{ij}^{n;\mathrm{SV}} \sigma_{ij}^n} + \cancel{\sum_{k\in\mathcal{K}} \sum_{x\in\mathcal{X}} D_{ij}^{n;k,x} \sigma_{ij}^n} - D_{ij} \sigma_{ij}^n \right\} = 0$$

After all the cancellations, the only possibly negative unbounded terms on the left-hand side is $-D_{ij}\sigma_{ij}^n$. Since $\{t_{ij}^{n;\mathrm{SV}}\}$ is bounded, it follows from $\alpha_1^{\mathrm{SV}} t_{ij}^{n;\mathrm{SV}} + \alpha_2^{\mathrm{SV}} d_{ij}^0 \ge \sigma_{ij}^n$ that $\{\sigma_{ij}^n\}$ is bounded.

Consequently, since the left-hand side sums up to zero, it follows that the (nonnegative) terms

$$(t_{s,i}^{n;k,\mathrm{AV}} + t_{ij}^{n;k,\mathrm{AV}})\nu_{\mathrm{AV}}^{n;k} z_{s,ij}^{n;k,\mathrm{AV}} \quad \text{and} \quad (t_{s,i}^{n;k,\mathrm{AV}} + t_{ij}^{n;k,\mathrm{AV}})\nu^{n;k} z_{s,ij}^{n;k}$$

are bounded. Suppose there is a sequence of positive scalars $\{\tau_n\}$ such that $\{\nu^{n;k}\} \to \infty$ as $n \to \infty$ for some $k \in \mathcal{K}$. It follows from the above summation that $\left\{ \sum_{x \in X} \sum_{(i,j) \in \mathcal{W}^{k,x}} \sum_{s \in \mathcal{D}^{k;x}} z_{s,ij}^{n;k,x} (t_{s,i}^{n;k,x} + t_{ij}^{n;k,x}) \right\} \to 0$ as $n \to \infty$. Since $\nu^{n;k} > 0$ for all n sufficiently large, we have

$$\tau_n \, \nu^{n;k} + N^k - \underbrace{\sum_{x \in X} \sum_{(i,j) \in \mathcal{W}^{k,x}} \sum_{s \in \mathcal{D}^{k;x}} z_{s,ij}^{n;k,x} (t_{s,i}^{n;k,x} + t_{ij}^{n;k,x})}_{\text{converges to zero}} = 0$$

which is a contradiction because $N^k > 0$. In a similar way, we can also obtain a contradiction if $\{\nu_{\mathrm{AV}}^{n;k}\}$ is unbounded for some k. □

References

Auld, J., Sokolov, V., Stephens, T.S., 2017. Analysis of the effects of connected–automated vehicle technologies on travel demand. Transportation Research Record 2625, 1–8.

Ban, X.J., Dessouky, M., Pang, J.S., Fan, R., 2019. A general equilibrium model for transportation systems with e-hailing services and flow congestion. Transportation Research Part B: Methodological 129, 273–304.

Battifarano, M., Qian, S., 2023. The impact of optimized fleets in transportation networks. Transportation Science 57, 1047–1068. URL: https://api.semanticscholar.org/CorpusID:255722404.

Braverman, A., Dai, J.G., Liu, X., Ying, L., 2019. Empty-car routing in ridesharing systems. Operations Research 67, 1437–1452.

Castro, F., Gao, J., Martin, S., 2024. Autonomous vehicles in ride-hailing and the threat of spatial inequalities. Available at SSRN 4332493 .

Chen, X., Di, X., 2024. A network equilibrium model for integrated shared mobility services with ride-pooling. Transportation Research Part C: Emerging Technologies 167, 104837.

Chen, Z., Lin, X., Yin, Y., Li, M., 2020. Path controlling of automated vehicles for system optimum on transportation networks with heterogeneous traffic stream. Transportation Research Part C: Emerging Technologies 110, 312–329.

Chen, Z., Stuart, A.L., Guo, Y., Zhang, Y., Li, X., 2024. Distributional equity impacts of automated vehicles: A disaggregated approach. Transportation Research Part C: Emerging Technologies 167, 104828.

Childress, S., Nichols, B., Charlton, B., Coe, S., 2015. Using an activity-based model to explore the potential impacts of automated vehicles. Transportation Research Record 2493, 99–106.

Deichmann, J., Ebel, E., Heineke, K., Heuss, R., Kellner, M., Steiner, F., 2023. Autonomous driving's future: Convenient and connected. https://www.mckinsey.com/industries/automotive-and-assembly/our-insights/autonomous-drivings-future-convenient-and-connected.

Di, X., Ban, X.J., 2019. A unified equilibrium framework of new shared mobility systems. Transportation Research Part B: Methodological 129, 50–78.

Dirkse, S.P., Ferris, M.C., 1995. The path solver: a nommonotone stabilization scheme for mixed complementarity problems. Optimization Methods and Software 5, 123–156.

Dong, T., Luo, Q., Xu, Z., Yin, Y., Wang, J., 2024. Strategic driver repositioning in ride-hailing networks with dual sourcing. Transportation Research Part C: Emerging Technologies 158, 104450.

Facchinei, F., Pang, J.S., 2003. Finite-dimensional variational inequalities and complementarity problems. Springer.

Fagnant, D.J., Kockelman, K., 2015. Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations. Transportation Research Part A: Policy and Practice 77, 167–181.

Feng, S., Ke, J., Xiao, F., Yang, H., 2022. Approximating a ride-sourcing system with block matching. Transportation Research Part C: Emerging Technologies 145, 103920.

GAMS Development Corporation, 2025. Gams: General algebraic modeling system. https://www.gams.com/.

Gu, W., Ba, Q., Dessouky, M.M., Pang, J.S., 2025. Generalized traffic equilibrium with ride-hailing and customer waiting. Available at SSRN 5461575 .

Ke, J., Yang, H., Li, X., Wang, H., Ye, J., 2020. Pricing and equilibrium in on-demand ride-pooling markets. Transportation Research Part B: Methodological 139, 411–431.

Lai, Z., Li, S., 2023. Spatiotemporal pricing and fleet management of autonomous mobility-on-demand networks: A decomposition and dynamic programming approach with bounded optimality gap. IEEE Transactions on Intelligent Transportation Systems 25, 7057–7069.

LeBlanc, L.J., Morlok, E.K., Pierskalla, W.P., 1975. An efficient approach to solving the road network equilibrium traffic assignment problem. Transportation Research 9, 309–318.

Li, S., Yang, H., Poolla, K., Varaiya, P., 2021. Spatial pricing in ride-sourcing markets under a congestion charge. Transportation Research Part B: Methodological 152, 18–45.

Lyft, Inc., 2025. Lyft and waymo launch autonomous ride-hailing service in nashville. https://investor.lyft.com/news-and-events/news/news-details/2025/Lyft-and-Waymo-Launch-Partnership-to-Expand-Autonomous-Mobility-to-Nashville/default.aspx.

Miller, D., Friesen, P.H., 1986. Porter's (1980) generic strategies and performance: an empirical examination with american data: part i: testing porter. Organization Studies 7, 37–55.

Ni, L., Chen, C., Wang, X.C., Chen, X.M., 2021. Modeling network equilibrium of competitive ride-sourcing market with heterogeneous transportation network companies. Transportation Research Part C: Emerging Technologies 130, 103277.

Rosen, J.B., 1965. Existence and uniqueness of equilibrium points for concave n-person games. Econometrica: Journal of the Econometric Society , 520–534.

Stabler, B., 2025. Transportation networks repository: Sioux falls network. https://github.com/bstabler/TransportationNetworks/tree/master/SiouxFalls. Accessed October 2025.

Stern, R.E., Cui, S., Delle Monache, M.L., Bhadani, R., Bunting, M., Churchill, M., Hamilton, N., Haulcy, R., Pohlmann, H., Wu, F., et al., 2018. Dissipation of stop-and-go waves via control of autonomous vehicles: Field experiments. Transportation Research Part C: Emerging Technologies 89, 205–221.

Tesla, Inc., 2025. Tesla q2 2025 update. https://www.tesla.com/sites/default/files/downloads/TSLA-Q2-2025-Update.pdf.

Uber Technologies, Inc., 2025. Atlanta: The future is here with waymo and uber. https://investor.uber.com/news-events/news/press-release-details/2025/Atlanta-The-Future-is-Here-with-Waymo-and-Uber-2025-y1tuDCCSgu/default.aspx. Accessed October 16, 2025.

Waymo, 2025. Scaling our fleet through u.s. manufacturing. https://waymo.com/blog/2025/05/scaling-our-fleet-through-us-manufacturing. Accessed October 16, 2025.

Wu, C., Kreidieh, A.R., Parvate, K., Vinitsky, E., Bayen, A.M., 2021. Flow: A modular learning framework for mixed autonomy traffic. IEEE Transactions on Robotics 38, 1270–1286.

Xu, Z., Chen, Z., Yin, Y., Ye, J., 2021. Equilibrium analysis of urban traffic networks with ride-sourcing services. Transportation Science 55, 1260–1279.

Zha, L., Yin, Y., Xu, Z., 2018. Geometric matching and spatial pricing in ride-sourcing markets. Transportation Research Part C: Emerging Technologies 92, 58–75.

Zheng, Y., Wang, J., Li, K., 2020. Smoothing traffic flow via control of autonomous vehicles. IEEE Internet of Things Journal 7, 3882–3896.