HOKKAIDO UNIVERSITY

Title	Operation Model for Shared Autonomous Vehicles Which Affect Congestion
Author(s)	Matsumoto, Sorachi; Tani, Ryuichi; Uchida, Kenetsu
Citation	International journal of intelligent transportation systems research, 23(2), 1064-1079 https://doi.org/10.1007/s13177-025-00498-5
Issue Date	2025
Doc URL	https://hdl.handle.net/2115/97762
Rights	This version of the article has been accepted for publication, after peer review (when applicable) and is subject to Springer Nature' s AM terms of use, but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: http://dx.doi.org/10.1007/s13177-025-00498-5
Type	journal article
File Information	IJIT_matsumoto_finally submitte.pdf

Hokkaido University Collection of Scholarly and Academic Papers : HUSCAP

Operation Model for Shared Autonomous Vehicles Which Affect Congestion

Sorachi Matsumoto[1], Ryuichi Tani[1] & Kenetsu Uchida[1]
Graduate School of Engineering, Hokkaido University[1]
(Kita 13-jo Nishi 8-chome, Kita-ku, Sapporo, Hokkaido, Japan,
Email: matsumoto.sorachi.q2@elms.hokudai.ac.jp, r-tani@eng.hokudai.ac.jp, uchida@eng.hokudai.ac.jp)

Abstract

Shared autonomous vehicles are anticipated to be adopted widely in the future and have prompted the development of various models for planning their operations. Existing models fall into three primary categories: micro-simulations, routing problems such as the Dial-a-Ride problem (DARP) with fixed travel times, and routing problems involving dynamic traffic assignment. However, none of these provide detailed input/output corresponding to each vehicle's operation and theoretically accounting for congestion effects in the network. This paper introduces a novel approach that combines the DARP with a static user equilibrium traffic assignment model. Consequently, the DARP in this study is formulated as a mixed-integer nonlinear problem with equilibrium constraints. Finding an exact optimal solution to a nonlinear and nonconvex problem, such as the DARP addressed in this study, is generally regarded as a challenging task. This study proposes an algorithm for the DARP that finds a solution by repeatedly solving problems formulated using linearly approximated equilibrium link flows. Finally, numerical calculations are conducted to assess the model's validity.

Keywords Shared autonomous vehicle, Dial-a-Ride problem, Static user equilibrium traffic assignment, Sensitivity analysis, Branch-and-Cut, Fixed-point problem

1 Introduction

The sharing economy is spreading following the development of information technologies. Car-sharing and ride-sharing are prime examples in the field of transportation. They enable us to use vehicles more efficiently and reduce the cost per capita [1]. Furthermore, fully-autonomous (level-5) vehicles are highly compatible with these sharing services. For example, if vehicles are human-driven, it is hard to allow car-sharing users to drop off a vehicle at their destinations because vehicles are parked on the roadside, exacerbating congestion. Level-5 vehicles can solve this problem by moving unattended to a designated depot or to locations where other users are waiting. Consequently, services with the following characteristics are expected to spread.

・A local government has multiple level-5 autonomous vehicles.
・Users call a vehicle as necessary.
・Ride-sharing is implemented to improve the efficiency of operations.

In this paper, we define a vehicle used in such services as the shared autonomous vehicle (SAV). This service has advantages of both sharing services and level-5 vehicles. Moreover, the operations can be made as flexible as those of conventional taxis. However, we should consider some problems such as an increase in in-vehicle time due to the detour of ride-sharing vehicles and decrease in punctuality due to insufficiency of congestion prediction, as claimed by Krueger et al [2]. In addition, de Almeida Correia and van Arem [3] claimed aggravation of congestion around SAV depots. Hence, we have to predict congestion by considering the pattern of SAV operation and decide the routes and schedules of SAVs.

Micro-simulation methods provide a way to address this. Fagnant and Kockelman [4] used an agent-based model to estimate users' waiting times, all vehicles' distance traveled, and cold-start savings which are affected by both different scenarios and vehicle relocation strategies. Levin et al. [5] concluded that dynamic ride-sharing can reduce the congestion due to relocation of SAVs by using a model that combines a cell transmission model with an agent-based model. Other than these, many empirical agent-based simulations are carried out (e.g. [6-9]). Micro-simulation methods are suitable for analyzing a single scenario in detail but have some disadvantages. Firstly, they are often too complex due to many parameters that we can hardly calibrate simultaneously. Secondly, their algorithms are often ad hoc and not systematic. Many kinds of models and algorithms have been proposed, but their relationships are unclear. Therefore, a model which is useful for a specific scenario cannot be applied to others.

In contrast, mathematical models with simpler assumptions and fewer parameters usually overcome these problems. Especially, the Dial-a-Ride Problem (DARP) is often applied to the planning of SAV operations. Constraints of DARP make each SAV's route feasible as explained later. The optimal routes and schedules are the outputs obtained by solving the DARP. The DARP under constant travel times was first proposed in the 1970s according to Jaw et al. [10]. Many heuristic solution algorithms have been developed because of the NP-hardness (e.g. [11-14]). On the other hand, Cordeau [15] and Ropke et al. [16] obtained the exact solution by the Branch-and-Cut algorithm. The strong point of DARP is that it provides sufficient outputs, i.e., routes and schedules of SAVs which can be used as input to micro-simulation models as mentioned earlier. The weak point of DARP, in contrast, is that it does not take the congestion caused by SAVs themselves into account. Thus, the more SAVs are operated, the more inaccurate travel times are estimated.

In consequence, some models that combine a dynamic traffic assignment theory with the DARP are proposed. de Almeida Correia and van Arem [3] formulated the DARP by using privately owned vehicles as a dynamic user optimum problem. Levin [17] formulated the DARP as a linear programming by applying a link transmission model. They performed continuous approximation by assuming there is no carpooling. Tani et al. [18] constructed a model which also considers carpooling. On the other hand, Liang et al. [19] formulated the DARP as an integer nonlinear programming by applying a dynamic traffic assignment and solved it approximately by the Lagrangian relaxation method. Also, Fan et al. [20] formulated a similar problem to that of Liang et al. [19] and approximately solved it by performing linearization of variables. These models can express the dynamic nature of traffic flows and provide time-varying travel times. Hence, they are the most helpful to predict the congestion caused by the operations of many SAVs. Meanwhile, they have some problems to point out compared to the original DARP. Primarily, their inputs and outputs lack some important information. For instance, each user's arrival time constraint cannot be addressed. Also, since routes of SAVs are aggregated, the route and schedule of each SAV are not obtained. Consequently, there is a need to develop another method that combines a road network model with the conventional DARP.

Correia et al. [3] note that "The problem we want to define cannot be considered static over time. It is dynamic by nature". As they claim, users' choice of departure or arrival time should be designated freely, and travel times change moment by moment in real situations. At least, it is unequivocal that we have to take the information about requested departure or arrival times into consideration when we decide the routes and schedules of SAVs. In contrast, the static traffic assignment can obtain travel times with a certain degree of accuracy. As an illustration, Zhang et al. [21] calculated the travel times of SAVs by a static traffic assignment model. In their research, an OD-SAV is defined as an SAV shared by users whose origin and destination locations are the same. They formulated a problem for describing the route choice of OD-SAVs which is different from the one addressing ordinary SAVs.

Accordingly, we propose a DARP with user equilibrium (UE) constraints as an extension of the conventional DARP with constant travel times. The rest of the paper is organized as follows. The formulation of DARP with the static traffic assignment is presented in Section 2. Section 3 presents the solution algorithm for the original problem in Section 2. In Section 4, we report the result of a numerical example. Section 5 concludes the paper.

2 Model description

2.1 Notations

Networks	
$G^{\uparrow}$	Inter-personal network
$G^{\downarrow}$	Road network
Sets	
$\boldsymbol{U}^{+}$	$=\{1,2,\dots,n_u\}$, set of boarding nodes for users in the inter-personal network
$\boldsymbol{D}^{+}$	$=\{n_u+1,n_u+2,\dots,n\}$, set of departing depot nodes in the inter-personal network
$\boldsymbol{U}^{-}$	$=\{i+n \mid \forall i\in\boldsymbol{U}^{+}\}$, set of alighting nodes for users in the inter-personal network
$\boldsymbol{D}^{-}$	$=\{i+n \mid \forall i\in\boldsymbol{D}^{+}\}$, set of returning depot nodes in the inter-personal network
$\boldsymbol{U}$	$=\boldsymbol{U}^{+}\sqcup\boldsymbol{U}^{-}$, set of all users' nodes in the inter-personal network
$\boldsymbol{D}$	$=\boldsymbol{D}^{+}\sqcup\boldsymbol{D}^{-}$, set of all depot nodes in the inter-personal network
$\boldsymbol{N}^{\uparrow}$	$=\boldsymbol{U}\sqcup\boldsymbol{D}$, set of all nodes in the inter-personal network
$\boldsymbol{A}^{\uparrow}$	$=\{(i,j) \mid \forall i,j\in\boldsymbol{N}^{\uparrow}\}$, set of links in the inter-personal network
$\boldsymbol{S}$	Set of nodes on an infeasible route in the inter-personal network
$\mathfrak{S}$	Family of sets $\boldsymbol{S}$
$\boldsymbol{N}^{\downarrow}$	$=\{1,2,\dots,N\}$, set of nodes in the road network with N nodes
$\boldsymbol{A}^{\downarrow}$	$=\{1,2,\dots,A\}$, set of links in the road network with A links
$\boldsymbol{\Omega}$	$=\{rs \mid \forall r,s\in\boldsymbol{N}^{\downarrow}\}$, set of origin-destination (OD) pairs in the road network
$\boldsymbol{K}_{rs}$	Set of paths serving between nodes of OD pair $rs\in\boldsymbol{\Omega}$
Variables	
Z	Value of the objective function
x_{ij}	Equals 1 if an SAV moves on link $(i,j)(\in\boldsymbol{A}^{\uparrow})$, and 0 otherwise
R_i	Number of users in an SAV just after departing from node $i(\in\boldsymbol{N}^{\uparrow})$
B_i	SAV's visiting time to node $i(\in\boldsymbol{N}^{\uparrow})$
L_i	In-vehicle time of user $i(\in\boldsymbol{N}^{\uparrow})$
t_a	Travel time of link $a(\in\boldsymbol{A}^{\downarrow})$ in the road network
t_{rs}	Travel time between nodes of OD pair $rs\in\boldsymbol{\Omega}$ in the road network
f_k^{rs}	Flow on the kth path between nodes of OD pair $rs\in\boldsymbol{\Omega}$
ξ_a	Flow on link $a(\in\boldsymbol{A}^{\downarrow})$
ω_{rs}	Travel demand of SAVs between nodes of OD pair $rs\in\boldsymbol{\Omega}$

Parameters	
n_u	Number of users
n_d	Number of depots
n	$= n_u + n_d$, by which the boarding node i of a user can be connected with the corresponding alighting node $i + n$
r_i	Number of users boarding at node $i(\in \boldsymbol{N}^{\uparrow})$
$\hat{R}$	Capacity of an SAV
d_i	Service duration at node $i(\in \boldsymbol{N}^{\uparrow})$
e_i	Earliest arrival time at node $i(\in \boldsymbol{N}^{\uparrow})$
l_i	Latest arrival time at node $i(\in \boldsymbol{N}^{\uparrow})$
$\hat{L}$	Maximum in-vehicle time of a user
N	Number of nodes in the road network
A	Number of links in the road network
$\bar{\omega}_{rs}$	Travel demand of nonusers between nodes of OD pair $rs \in \boldsymbol{\Omega}$
y_{ij}^{rs}	Constant which equals 1 if a move on link $(i,j)(\in \boldsymbol{A}^{\uparrow})$ in the inter-personal network corresponds to a trip between nodes of OD pair $rs \in \boldsymbol{\Omega}$ on the road network, and 0 otherwise
δ_{ak}^{rs}	Constant which equals 1 if link $a(\in \boldsymbol{A}^{\downarrow})$ is on the kth path of OD pair $rs \in \boldsymbol{\Omega}$, and 0 otherwise
$t_a^{0\downarrow}$	Free flow travel time of link $a(\in \boldsymbol{A}^{\downarrow})$
α, β	Parameters of the BPR function
C_a	Capacity of link $a(\in \boldsymbol{A}^{\downarrow})$
Mapping	
ϕ	Address Book Mapping

2.2 Assumptions

The assumptions in this study are as follows:

- All SAVs are homogeneous.
- Each person in the problem is either an SAV user or a private car user, hereinafter referred to simply as a user or a nonuser, respectively.
- All users can specify their boarding and/or alighting time windows. Their in-vehicle times are equal to or shorter than a given maximum in-vehicle time $\hat{L}$.
- All SAVs depart from a depot and return to a depot which may be different. However, the number of vehicles departing and returning at each depot must be the same, ensuring that no relocations of SAVs are needed for the next operation.
- Each nonuser owns a private car and travels between his/her OD pair by the car.
- All vehicles can be parked at any node in the road network. Parking of vehicles at a node does not affect congestion in the road network.
- Travel demand for each person is given in advance. We do not consider travel times varying moment by moment. Consequently, traffic flows and travel times in the road network are determined by solving the hourly UE traffic assignment problem under fixed demands. Therefore, the operation time range should be at most an hour to obtain meaningful results.
- The BPR function [22] is used as a link cost function.
- The SAV operator optimizes its operation to minimize the total travel time in the road network.

2.3 Inter-personal network and road network

To solve the conventional DARP, a network with nodes representing both users' boarding and alighting locations, and SAVs' departure and return depots is typically used. This network is called the inter-personal network. By using such a network, we can determine respective routes and schedules of SAVs. In contrast, for traffic assignment calculations, we also use a road network with nodes representing intersections and links representing road sections. Thus, we introduce "Address Book Mapping" to connect these two networks, enabling them to be handled simultaneously in a single problem. This mapping converts a solution to DARP into an OD demand table for a traffic assignment model. Then, travel times between user locations are also obtained from the results of traffic assignment. In other words, we can calculate the value of the objective function of DARP while considering congestion in the road network theoretically.

Our model has an almost identical problem structure as the conventional DARP, which is solved solely on the inter-personal network $G^{\uparrow}(\boldsymbol{N}^{\uparrow}, \boldsymbol{A}^{\uparrow})$. $G^{\uparrow}$ has $2 \cdot n$ nodes where $n = n_u + n_d$. n_u and n_d are the number of users and that of depots, respectively. Nodes in $\boldsymbol{N}^{\uparrow}$ are classified broadly into nodes in $\boldsymbol{U}$ and those in $\boldsymbol{D}$. Nodes in $\boldsymbol{U}$ are further classified into nodes in $\boldsymbol{U}^+$ and those in $\boldsymbol{U}^-$. In a similar way, nodes in $\boldsymbol{D}$ are classified into nodes in $\boldsymbol{D}^+$ and those in $\boldsymbol{D}^-$ as shown in Figure 1. $\boldsymbol{U}^+ = \{1, 2, \dots, n_u\}$ and $\boldsymbol{U}^- = \{n+1, n+2, \dots, n+n_u\}$ represent the set of users' boarding nodes and that of alighting nodes, respectively. $\boldsymbol{U} = \boldsymbol{U}^+ \sqcup \boldsymbol{U}^-$ is the set of all users' nodes. Each user i is associated with both the boarding node i and the alighting node $n + i$. $\boldsymbol{D}^+ = \{n_u + 1, n_u + 2, \dots, n\}$ and $\boldsymbol{D}^- = \{n + n_u + 1, n + n_u + 2, \dots, 2n\}$ denote the set of departing nodes and that of returning depot nodes, respectively. $\boldsymbol{D} = \boldsymbol{D}^+ \sqcup \boldsymbol{D}^-$ is the set of all depot nodes. Each depot i is associated with both the departing node $n_u + i$ and the returning node $n + n_u + i$. Moreover, each pair of nodes in $\boldsymbol{N}^{\uparrow}$ is connected by a link in $\boldsymbol{A}^{\uparrow}$. The binary decision variable x_{ij} is defined for all $(i,j) \in \boldsymbol{A}^{\uparrow}$. It equals 1 if an SAV visits node j on the inter-personal network immediately after departing from node i, and 0 otherwise.

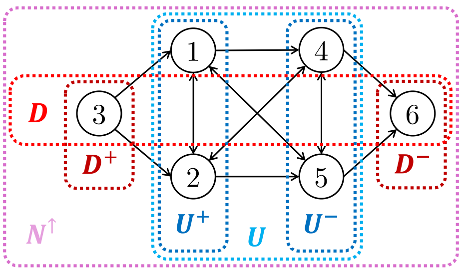


Figure 1. Inter-personal network, $n_u = 2, n_d = 1$.

The conventional DARP treats the travel time of a link in $\boldsymbol{A}^{\uparrow}$ as a constant. In contrast, we treat it as a variable to take congestion into account. To obtain link travel times, we solve the static UE assignment problem on a road network $G^{\downarrow}(\boldsymbol{N}^{\downarrow}, \boldsymbol{A}^{\downarrow})$. Therefore, we must use two different networks, $G^{\uparrow}$ and $G^{\downarrow}$, to solve our model.

Next, we define the mapping $\phi: \boldsymbol{N}^{\uparrow} \to \boldsymbol{N}^{\downarrow}$ to clarify the relationship between nodes of these two networks. Given a node in the inter-personal network as an input, this mapping returns the corresponding node number in the road network. We call this mapping “Address Book Mapping” because its function is similar to that of an address book, which provides the postal address when a person’s name is given.

Figure 2 illustrates an example of Address Book Mapping. For instance, $\phi(4) = 5$ indicates node 4 in the inter-personal network corresponds to node 5 in the road network. Additionally, a movement along link (2,4) in the inter-personal network is projected as traffic from node 3 to 5 in the road network. Then, $y_{2,4}^{3,5}$ equals one in such a case. All y_{ij}^{rs} are constants because we assume users’ travel demands are known in advance. By projecting all movements from the inter-personal network onto the road network, we can obtain an OD table containing information about SAV operation. Here, we assume that each SAV selfishly chooses the shortest path between its OD pair; in other words, Wardrop’s first principle [23] holds. Using the traffic assignment model, we obtain link travel times in the road network, and the shortest path travel time between nodes of OD pairs. In the next section, we specifically explain how to obtain an OD table from x_{ij} and y_{ij}^{rs}.

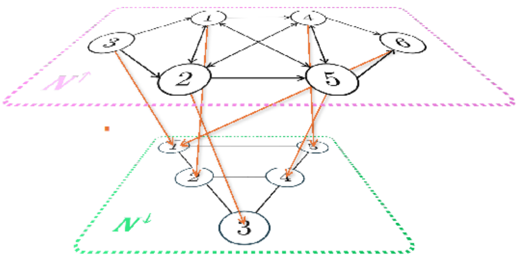


Figure 2. Example of Address Book Mapping

2.4 Formulation

Our model, the DARP with the UE constraints, can be formulated as the following mixed-integer nonlinear problem, referred to as [OP: original problem]:

[OP]

$$\min Z = \sum_{rs \in \boldsymbol{\Omega}} t_{rs} \cdot (\omega_{rs} + \bar{\omega}_{rs}) \qquad (1)$$

subject to

$$\sum_{i \in \boldsymbol{N}^{\uparrow}} x_{ij} = 1 \qquad \forall j \in \boldsymbol{U} \ (2)$$

$$\sum_{j \in \boldsymbol{N}^{\uparrow}} x_{ij} = 1 \qquad \forall i \in \boldsymbol{U} \ (3)$$

$$\sum_{j \in \boldsymbol{N}^{\uparrow}} x_{ij} = \sum_{j \in \boldsymbol{N}^{\uparrow}} x_{j,i+n} \qquad \forall i \in \boldsymbol{D}^{+} \ (4)$$

$$\sum_{i \in \boldsymbol{N}^{\uparrow}} x_{ij} = 0 \qquad \forall j \in \boldsymbol{D}^{+} \ (5)$$

$$\sum_{j \in \boldsymbol{N}^{\uparrow}} x_{ij} = 0 \qquad \forall i \in \boldsymbol{D}^{-} \ (6)$$

$$\sum_{i,j \in \boldsymbol{S}} x_{ij} \le |\boldsymbol{S}| - (n_d + 1) \qquad \forall \boldsymbol{S} \in \mathfrak{S} \ (7)$$

$$R_j \ge (R_i + r_j) \cdot x_{ij} \qquad \forall (i,j) \in \boldsymbol{A}^{\uparrow} \ (8)$$

$$\max\{0, r_i\} \le R_i \le \min\{\hat{R}, \hat{R} + r_i\} \qquad \forall i \in \boldsymbol{N}^{\uparrow} \ (9)$$

$$B_j \ge \left(B_i + d_i + t_{\phi(i)\phi(j)}\right) \cdot x_{ij} \qquad \forall (i,j) \in \boldsymbol{A}^{\uparrow} \ (10)$$

$$e_i \le B_i \le l_i \qquad \forall i \in \boldsymbol{N}^{\uparrow} \ (11)$$

$$t_{\phi(i)\phi(n+i)} \le L_i \le \hat{L} \qquad \forall i \in \boldsymbol{U}^{+} \ (12)$$

$$\omega_{rs} = \sum_{(i,j) \in \boldsymbol{A}^{\uparrow}} x_{ij} \cdot y_{ij}^{rs} \qquad \forall rs \in \boldsymbol{\Omega} \ (13)$$

$$f_k^{rs} \cdot \left(\sum_{a \in \boldsymbol{A}^{\downarrow}} \delta_{ak}^{rs} \cdot t_a - t_{rs} \right) = 0 \ \forall k \in \boldsymbol{K}_{rs}, \forall rs \in \boldsymbol{\Omega} \ (14)$$

$$\sum_{a \in \boldsymbol{A}^{\downarrow}} \delta_{ak}^{rs} \cdot t_a - t_{rs} \ge 0 \qquad \forall k \in \boldsymbol{K}_{rs}, \forall rs \in \boldsymbol{\Omega} \ (15)$$

$$\sum_{k \in \boldsymbol{K}_{rs}} f_k^{rs} = \omega_{rs} + \bar{\omega}_{rs} \qquad \forall rs \in \boldsymbol{\Omega} \ (16)$$

$$\xi_a = \sum_{rs \in \boldsymbol{\Omega}} \sum_{k \in \boldsymbol{K}_{rs}} \delta_{ak}^{rs} \cdot f_k^{rs} \qquad \forall a \in \boldsymbol{A}^{\downarrow} \ (17)$$

$$f_k^{rs} \ge 0, \xi_a \ge 0 \qquad \forall k \in \boldsymbol{K}_{rs}, \forall rs \in \boldsymbol{\Omega}, \forall a \in \boldsymbol{A}^{\downarrow} \ (18)$$

$$t_a(\xi_a) = t_a^0 \cdot \left\{ 1 + \alpha \cdot \left(\frac{\xi_a}{C_a} \right)^{\beta} \right\} \qquad \forall a \in \boldsymbol{A}^{\downarrow} \ (19)$$

Equation (1) represents the objective function of this study, which aims to minimize the total travel time in the road network.

The conditions given in (2)-(12) represent the constraints of the conventional DARP. Among them, (2)-(7) describe the constraints on the visiting order of SAVs. (2) and (3) indicate that each node in the inter-personal network representing a user is visited exactly once by an SAV. Equation (4) states that the number of vehicles departing from and returning to each depot must be the same, as mentioned in Section 2.2. Equation (4) can be replaced by other constraints depending on the assumptions set by the operator. (5) and (6) ensure that there are no vehicles flowing into departing depots or flowing out of returning depots. (7) is referred to as the “precedence constraint” in Ruland and Rodin [24], Ropke et al. [16], and other studies. This ensures that each user’s boarding and alighting nodes are situated on a route of the same vehicle, and that boarding node is visited before alighting one. Here, a set of nodes on an infeasible route, $\boldsymbol{S}$, is defined as $\boldsymbol{S} \subset \boldsymbol{N}^{\uparrow}$ such that $\boldsymbol{D}^{+} \subset \boldsymbol{S}, \boldsymbol{D}^{-} \subset \boldsymbol{N}^{\uparrow} \setminus \boldsymbol{S}$. There is at least one node $i \in \boldsymbol{U}^{+}$ which holds $i \notin \boldsymbol{S}$ and $i + n \in \boldsymbol{S}$. In other words, $\boldsymbol{S}$ is a set of nodes comprising a route which includes only the alighting node of a user, or which includes both boarding and alighting nodes of a user but the latter is visited earlier than the former. Additionally, $|\boldsymbol{S}|$ represents the number of nodes in $\boldsymbol{S}$.

Secondly, constraints on SAV capacities are shown by (8) and (9). (8) indicates that, if an SAV moves on link (i, j), the number of users in the SAV just after departing node j is the summation of that just after departing node i and the change in the number of users at node j. (9) shows that the number of users in an SAV is non-negative and equal to or smaller than the vehicle capacity. Here, r_i represents the change in the number of users at node i, as defined below:

$$r_i = \begin{cases} 1 & \text{if } i \in \boldsymbol{U}^+ \\ -1 & \text{if } i \in \boldsymbol{U}^- \\ 0 & \text{otherwise} \end{cases} \qquad \forall i \in \boldsymbol{N}^\uparrow \quad (20)$$

Following Ropke et al. [16], constraint (8) is replaced by the linearized function shown below.

$$R_j \geq R_i + r_j - W_{ij} \cdot \left(1 - x_{ij}\right) \qquad \forall (i, j) \in \boldsymbol{A}^\uparrow \quad (8')$$

where

$$W_{ij} \geq \min\left\{\hat{R}, \hat{R} + r_i\right\}$$

Finally, constraints on time are shown by inequalities (10)-(12). (10) and (11) are about the visiting time of each user's node. (10) indicates that, if an SAV moves on link (i, j), visiting time to node j is the summation of that to node i, service duration at node i, and travel time between nodes i and j. To linearize this constraint, Ropke et al. [16] proposed the following inequality:

$$B_j \geq B_i + d_i + t_{\phi(i)\phi(j)} - M_{ij} \cdot \left(1 - x_{ij}\right) \qquad \forall (i, j) \in \boldsymbol{A}^\uparrow \quad (10')$$

where

$$M_{ij} \geq \max\left\{0, l_i + d_i + t_{\phi(i)\phi(j)} - e_j\right\}$$

(11) guarantees every visiting time of a user is within his/her designated time window. (12) is about in-vehicle time, and it ensures that the in-vehicle time of each user is not shorter than the travel time needed to move directly from one's boarding node to the corresponding alighting node by a vehicle and is not longer than the maximum in-vehicle time. The in-vehicle time of each user is defined as:

$$L_i = B_{i+n} - (B_i + d_i) \qquad \forall i \in \boldsymbol{U}^+ \quad (21)$$

UE constraints (13)-(19) are incorporated to the DARP. Here, y_{ij}^{rs} is defined by introducing the Address Book Mapping as:

$$y_{ij}^{rs} = \begin{cases} 1 & \text{if } \phi(i) = r, \phi(j) = s \\ 0 & \text{otherwise} \end{cases} \qquad \forall (i, j) \in \boldsymbol{A}^\uparrow, \forall rs \in \boldsymbol{\Omega} \quad (22)$$

Equation (13) represents the relationship between OD demands in the road network and SAV operations, as described in Section 2.3. When $x_{ij} \cdot y_{ij}^{rs} = 1$, there exists an SAV that moves on link (i, j) in the inter-personal network. The move corresponds to SAV traffic between nodes r and s in the road network. Therefore, we can obtain a traffic demand between OD pair rs by summing up $x_{ij} \cdot y_{ij}^{rs}$ for all $(i, j) \in \boldsymbol{A}^\uparrow$. (14)-(19) are the complementarity conditions and constraints which represent the Wardrop's first principle.

Finally, note that the number of possible combinations of the decision variable x_{ij} is 2^{4n^2}, which is the same as in the conventional DARP. Due to the nonlinearity and nonconvexity of [OP], it is a more difficult problem to solve than the DARP which is formulated in a linear form.

3 Solution algorithm

As previously noted, [OP] is more difficult to solve than the conventional DARP. Originally, the conventional DARP is known as an NP hard problem. Its exact solution can be obtained by an algorithm for a mixed integer linear programming only when at most tens of users are considered.

In this section, we propose an algorithm to solve [OP] by defining an approximation problem for [OP]. The proposed algorithm for [OP] is inferred from those for the DARP which has a similar problem structure to [OP].

Firstly, we approximate [OP] as an easier problem, i.e., a mixed-integer quadratic approximation problem, which can be solved exactly by a solver. As the approximation technique, we use the sensitivity analysis for UE and convex approximation. Inputting an initial feasible solution, the sensitivity analysis for UE enables us to approximate travel times in the form of linear equations. Then, the objective function becomes a quadratic function. However, it is not always convex because the coefficient matrix of the quadratic term might be indefinite. To address this issue, we approximate this matrix using the closest positive semidefinite matrix. Here, we call this convex approximation. Hence, our model can be reformulated as a mixed-integer quadratic problem.

By solving this problem, we obtain a better solution than the initial solution. However, we might find an even better solution by using the obtained solution as the new initial solution. By repeating this procedure, we eventually obtain an unimprovable solution. This procedure can be formulated as a fixed-point problem. Therefore, a solution algorithm for this fixed-point problem yields an unimprovable solution to [OP].

In this section, we firstly list newly introduced notations. Then, we show the way to obtain travel times by the sensitivity analysis for UE. After that, we present the method of convex approximation and formulate the quadratic problem. Finally, we formulate a fixed-point problem to find an unimprovable solution, and we present the algorithm to solve it.

3.1 Notations

Matrices	
$\boldsymbol{Q}$	Positive semidefinite matrix of order N^2
$\boldsymbol{\Pi}$	OD/path incidence matrix where the entry in the $N \cdot (r-1) + s$th row and kth column is 1 if the kth path belongs to the OD

	pair rs, 0 otherwise. Only the used paths are enumerated.
$\boldsymbol{\Delta}$	Link/path incidence matrix where the entry in the ath row and the kth column is 1 if the link a is in the kth path. Only used paths are enumerated.
Vectors	
$\boldsymbol{x}$	$4n^2$-dimensional column vector where the $2n \cdot (i-1)+j$th element is x_{ij}
$\boldsymbol{x}^0$	An initial solution to $\boldsymbol{x}$
$\boldsymbol{t_\Omega}$	N^2-dimensional column vector where the $N \cdot (r-1)+s$th element is t_{rs}
$\boldsymbol{t}_{\boldsymbol{\Omega}}^0$	Exact OD travel times obtained from $\boldsymbol{\omega}^0$
$\boldsymbol{t_A}$	A-dimensional column vector where the ath element is t_a
$\boldsymbol{\omega}(\boldsymbol{x})$	N^2-dimensional column vector where the $N \cdot (r-1)+s$th element is ω_{rs}
$\boldsymbol{\omega}^0$	Initial OD demand values $\boldsymbol{\omega}(\boldsymbol{x}^0)$
$\overline{\boldsymbol{\omega}}$	N^2-dimensional column vector where the $N \cdot (r-1)+s$th element is $\overline{\omega}_{rs}$
$\boldsymbol{\varepsilon}$	N^2-dimensional column vector where the $N \cdot (r-1)+s$th element is a perturbation of ω_{rs}

3.2 Sensitivity analysis for UE

Firstly, we assume that an initial feasible solution $\boldsymbol{x} = \boldsymbol{x}^0$ has already been obtained. The perturbation vector $\boldsymbol{\varepsilon}$ is then given as follows:

$$\boldsymbol{\varepsilon}(\boldsymbol{x}) = \boldsymbol{\omega}(\boldsymbol{x}) - \boldsymbol{\omega}^0 \tag{23}$$

It is obvious that

$$\nabla_\varepsilon \boldsymbol{\omega}(\boldsymbol{x}^0) = \boldsymbol{I} \tag{24}$$

Moreover, we do not consider perturbations in the parameters of the link cost function, we have,

$$\nabla_\varepsilon \boldsymbol{t_A} = \boldsymbol{0} \tag{25}$$

The first derivatives of OD travel times with respect to $\boldsymbol{\varepsilon}$, evaluated at $\boldsymbol{\varepsilon} = \boldsymbol{0}$ under (24) and (25), are given by the following expression (see Tobin and Friesz [25] for more details).

$$\nabla_\varepsilon \boldsymbol{t_\Omega} = \left(\boldsymbol{\Pi}\left(\boldsymbol{\Delta}^T \nabla_\xi \boldsymbol{t_A} \boldsymbol{\Delta}\right)^{-1} \boldsymbol{\Pi}^T\right)^{-1} \tag{26}$$

where

$$\left(\nabla_\xi \boldsymbol{t_A}\right)_{aa'} = \begin{cases} \dfrac{\partial t_a(\xi_a)}{\partial \xi_a} & \text{if } a = a' \\ 0 & \text{otherwise} \end{cases} \quad \forall a, a' \in \boldsymbol{A}^{\downarrow} \tag{27}$$

Finally, the vector of link travel times at the UE, $\boldsymbol{t_\Omega}$, is approximated as follows:

$$\boldsymbol{t_\Omega} \approx \boldsymbol{t}_{\boldsymbol{\Omega}}^0 + \nabla_\varepsilon \boldsymbol{t_\Omega} \boldsymbol{\varepsilon}(\boldsymbol{x}) \tag{28}$$

3.3 Approximation as a quadratic problem

Firstly, we reformulate the objective function (1) of [OP] using the results in the previous section. The first term of (1) can be represented as the inner product of vectors as follows:

$$\begin{aligned} Z(\boldsymbol{x}) &= (\boldsymbol{\omega} + \overline{\boldsymbol{\omega}})^T \boldsymbol{t_\Omega} \\ &\approx (\boldsymbol{\varepsilon}(\boldsymbol{x}) + \boldsymbol{\omega}^0 + \overline{\boldsymbol{\omega}})^T \left(\boldsymbol{t}_{\boldsymbol{\Omega}}^0 + \nabla_\varepsilon \boldsymbol{t_\Omega} \boldsymbol{\varepsilon}(\boldsymbol{x})\right) \\ &= \boldsymbol{\varepsilon}(\boldsymbol{x})^T \boldsymbol{M} \boldsymbol{\varepsilon}(\boldsymbol{x}) + \boldsymbol{l}^T \boldsymbol{\varepsilon}(\boldsymbol{x}) + Z(\boldsymbol{x}^0) \end{aligned} \tag{29}$$

where

$$\boldsymbol{M} = \nabla_\varepsilon \boldsymbol{t_\Omega} \tag{30}$$

$$\boldsymbol{l} = \boldsymbol{t}_{\boldsymbol{\Omega}}^0 + \nabla_\varepsilon \boldsymbol{t}_{\boldsymbol{\Omega}}^T (\boldsymbol{\omega}^0 + \overline{\boldsymbol{\omega}}) \tag{31}$$

$$Z(\boldsymbol{x}^0) = (\boldsymbol{\omega}^0 + \overline{\boldsymbol{\omega}})^T \boldsymbol{t}_{\boldsymbol{\Omega}}^0 \tag{32}$$

It is obvious that $\boldsymbol{M} + \boldsymbol{M}^T$ is a symmetric matrix, and the following equation always holds for a square matrix $\boldsymbol{M}$.

$$\boldsymbol{\varepsilon}(\boldsymbol{x})^T \boldsymbol{M} \boldsymbol{\varepsilon}(\boldsymbol{x}) = \frac{1}{2} \boldsymbol{\varepsilon}(\boldsymbol{x})^T (\boldsymbol{M} + \boldsymbol{M}^T) \boldsymbol{\varepsilon}(\boldsymbol{x}) \tag{33}$$

However, $\boldsymbol{M} + \boldsymbol{M}^T$ may not be a positive semidefinite matrix. If we directly substitute (33) into (29), the objective function is not necessarily convex. For this reason, we approximate $\boldsymbol{M} + \boldsymbol{M}^T$ by the closest positive semidefinite matrix $\boldsymbol{Q}$, and we call this the convex approximation. To determine the optimal $\boldsymbol{Q}$, we solve the following optimization problem:

$$\min \|\boldsymbol{Q} - (\boldsymbol{M} + \boldsymbol{M}^T)\| \tag{34}$$

subject to

$$\boldsymbol{Q} \in \mathcal{S}_+^{N^2} \tag{35}$$

Here, $\mathcal{S}_+^{N^2}$ denotes the set of cones shaped by a N^2th-order positive semidefinite matrix. According to Tanaka and Nakata [26], the optimal $\boldsymbol{Q}$ is obtained as follows when we use Frobenius norm as the matrix norm.

$$\boldsymbol{Q} = \sum_{i:\lambda_i > 0} \lambda_i \boldsymbol{p}_i \boldsymbol{p}_i^T \tag{36}$$

where

$$\boldsymbol{M} + \boldsymbol{M}^T = \sum_{i=1}^{N^2} \lambda_i \boldsymbol{p}_i \boldsymbol{p}_i^T \tag{37}$$

Equation (37) represents the eigendecomposition of $\boldsymbol{M} + \boldsymbol{M}^T$, which is a symmetric matrix. λ_i is an eigenvalue and $\boldsymbol{p}_i$ is the corresponding eigenvector. By replacing $\boldsymbol{M} + \boldsymbol{M}^T$ in (33) with the matrix $\boldsymbol{Q}$ obtained in (36) and substituting it into (29), we can define a new objective function:

$$\tilde{Z}(\boldsymbol{x}|\boldsymbol{x}^0) = \frac{1}{2} \boldsymbol{\varepsilon}(\boldsymbol{x})^T \boldsymbol{Q} \boldsymbol{\varepsilon}(\boldsymbol{x}) + \boldsymbol{l}^T \boldsymbol{\varepsilon}(\boldsymbol{x}) + Z(\boldsymbol{x}^0) \tag{38}$$

This objective function is convex because $\boldsymbol{Q}$ is a positive semidefinite matrix. Combining this with DARP constraints, we can formulate the following optimization problem [MIQAP: mixed-integer quadratic approximation problem]

[MIQAP]

$$\min\ (38) \text{ subject to } (2)-(7), (8'), (9), (10'), (11)-(13)$$

However, a solution to [MIQAP] may not be feasible for [OP] or may not improve upon the initial solution when the traffic assignment is performed exactly with the UE conditions (13)-(19). Therefore, such a solution must be eliminated by adding a valid inequality to [MIQAP]. When a solution to [MIQAP] $\boldsymbol{x}^R = \left\{x_{ij}^R \middle| \forall (i,j) \in \boldsymbol{A}^{\uparrow}\right\}$ is obtained, it always satisfies (2)-(9) because they are not related with travel times. Therefore, only what we have to confirm is whether $B_i\ \forall i \in \boldsymbol{N}^{\uparrow}$ which satisfy (10)-(19) exist or not when $\boldsymbol{x} = \boldsymbol{x}^R$. Moreover, we consider the following constraint which ensures that an obtained solution is better than the original one.

$$Z(\boldsymbol{x}) \leq Z(\boldsymbol{x}^0) \quad (39)$$

When $\boldsymbol{x}$ does not satisfy (10)-(19) or (39) does not hold when $\boldsymbol{x} = \boldsymbol{x}^R$, we continue to solve [MIQAP] again after adding the following constraint, referred to as "congestion infeasible constraint" that excludes $\boldsymbol{x}^R$ from the feasible region:

$$\boldsymbol{s}^T\boldsymbol{x} + 1 \leq \sum_{(i,j)\in \boldsymbol{A}^{\uparrow}} x_{ij}^R \quad (40)$$

where

$$s_{ij} = \begin{cases} 1 & \text{if } x_{ij}^R = 1 \\ -1 & \text{if } x_{ij}^R = 0 \end{cases} \quad \forall i,j \in \boldsymbol{N}^{\uparrow} \quad (41)$$

For example, when $n = 2$, $\boldsymbol{x}^R = [1 \quad 0 \quad 0 \quad 1]^T$, $\boldsymbol{s} = [1 \quad -1 \quad -1 \quad 1]^T$, constraint (40) is

$$[1 \quad -1 \quad -1 \quad 1] \cdot \boldsymbol{x} + 1 \leq 2$$
$$x_{11} - x_{12} - x_{21} + x_{22} + 1 \leq 2$$
$$(1 - x_{11}) + x_{12} + x_{21} + (1 - x_{22}) \geq 1$$

The above constraint is violated if and only if $\boldsymbol{x} = \boldsymbol{x}^R$. By solving [MIQAP] with these additional inequalities, we can obtain a better solution than the initial one. Moreover, since it is feasible for [OP], we can solve [MIQAP] again inputting that solution as the new initial solution.

3.4 Reformulation as a fixed-point problem

As mentioned earlier, an optimal solution to [MIQAP] may be further improved by using it as an initial solution for [MIQAP] and solving it again. Therefore, we formulate the following [FPP: fixed-point problem] to develop an algorithm for [OP]:

[FPP]

$$\text{Find } \boldsymbol{x}^* \text{ such that } \boldsymbol{x}^* = \arg\min_{\boldsymbol{x}} \tilde{Z}(\boldsymbol{x}|\boldsymbol{x}^*) \quad (42)$$

subject to

$$(2) - (7), (8') - (10'), (11) - (19), (39)$$

Because of (39), the optimal solution to [OP] is a fixed point. However, it is not necessarily the unique fixed point. Assuming that a solution $\boldsymbol{x}^R$ is obtained, there may exist a better solution $\boldsymbol{x}$ than $\boldsymbol{x}^R$ such that

$$\tilde{Z}(\boldsymbol{x}|\boldsymbol{x}^0) > \tilde{Z}(\boldsymbol{x}^R|\boldsymbol{x}^0) \quad (44)$$
$$Z(\boldsymbol{x}) \leq Z(\boldsymbol{x}^R) \quad (45)$$

In this case, $\boldsymbol{x}^R$ remains a fixed point even though it is not the optimal solution to [OP]. If there is a large gap between $Z(\boldsymbol{x}^R)$ and the optimal value $Z(\boldsymbol{x}^*)$ where $\boldsymbol{x}^*$ is the optimal solution, however, the likelihood of $\boldsymbol{x}^R$ being a fixed point is low. Hence, we can consider a fixed-point solution to be at least close to the optimal one, and further improvement is difficult.

Accordingly, the algorithm for finding a fixed point of [FPP] is presented in Table 1. As an initial feasible solution in Step 1, we can set an obviously feasible solution. For example, a solution in which the number of SAVs equals the number of SAV users, and each SAV carries only one SAV user during its travel from one depot to another depot is obviously feasible. In general, however, it is expected that the computation time will be shorter if an appropriate initial feasible solution in Step 1, which is close to the exact solution, is found. A solution to the DARP with constant travel times seems an appropriate initial solution in many cases.

To solve [MIQAP] in Step 3, firstly, the relaxation problem for [MIQAP] is defined by fixing the variables x_{ij} as 0 or 1 for some $(i,j) \in \boldsymbol{A}^{\uparrow}$ and changing the domain of the variables x_{ij} from discrete variables, $x_{ij} \in \{0,1\}$, to continuous variables, $0 \leq x_{ij} \leq 1$, for other $(i,j) \in \boldsymbol{A}^{\uparrow}$. By solving the relaxation problem of [MIQAP] iteratively following the Branch-and-Cut algorithm, the original [MIQAP] is solved. The termination criteria of the solution process for [MIQAP] in Step 3 are set by three conditions: the maximum number of total iterations to solve the relaxation problem for [MIQAP], the maximum number of iterations to solve the relaxation problem for [MIQAP] after finding the solution that gives the minimum objective value compared to the solutions obtained so far, and the optimality gap. The optimality gap is defined as $1 - (\text{lower bound})/(\text{upper bound})$ where the lower bound and upper bound are from [MIQAP].

Next, we discuss the applicability of this algorithm. In Step 3, we solve [MIQAP], which has similar constraints to the conventional DARP and the quadratic objective function. Solving [MIQAP] takes longer time than solving the conventional DARP which has a linear objective function. Ropke et al. [16] reported that the maximum number of users addressed in the DARP, which can be solved was 36, and that 37.57 minutes were needed for the computation. In Step 3, a longer computation time seems to be needed for solving [MIQAP] addressing over 50 users, and addressing over 100 users may be difficult to solve.

Table 1. Solution algorithm of [FPP]

Step 1: Provide a feasible solution $\boldsymbol{x}^0$ for [OP].
Step 2: Calculate $\boldsymbol{Q}$ and $\boldsymbol{l}$ from $\boldsymbol{x}^0$.
Step 3: Solve [MIQAP].
If a feasible solution $\boldsymbol{x}^R$ for [MIQAP] is found,
Check whether $\boldsymbol{x}^R$ is also feasible for [OP] by substituting it into (10)-(19) and (39).
If $\boldsymbol{x}^R$ is infeasible for [OP]
Add a constraint (40) to [MIQAP] and continue to solve.
If one of termination criteria is satisfied, proceed to Step 4.
Step 4: Update the initial solution.
If the solution to [MIQAP] does not equal $\boldsymbol{x}^0$,
Update $\boldsymbol{x}^0$ to the solution to [MIQAP] and proceed to Step 2.
Otherwise, output $\boldsymbol{x}^0$ as a fixed point $\boldsymbol{x}^*$, and terminate the algorithm.

4 Numerical experiments

4.1 Settings

We conducted a numerical experiment using the Sioux-Falls network to verify the validity of the proposed algorithm. Figure 3 illustrates the network structure and link properties (free flow travel times and capacities). The parameters of the BPR function are $\alpha = 0.15, \beta = 4$. Here, depot nodes are 4, 6, 20 and 24. Figure 4 presents OD demands for users and nonusers. There are 40 users and 120 nonusers in this experiment. The service duration of each user is $d_i = 0\ \forall i \in \boldsymbol{N}^{\uparrow}$. The operation starts at time 0 and ends at 300 (unit time). Each user has a loose time window $[e_i, l_i] = [0{,}300]$ at either the boarding or alighting node and has a tight time window at the other nodes. e_i of a tight time window is set by randomly sampling from the uniform distribution with the support $30 \leq e_i \leq 300 - \hat{L}$ if i is a boarding node, and from the one with the support $30 + \hat{L} \leq e_i \leq 255$ if i is an alighting node. l_i is set by adding a randomly sampled value from the uniform distribution defined between 15 and 45 to e_i. Moreover, let $\hat{R} = 4, \hat{L} = 60$. We use the obvious feasible solution mentioned above as the initial solution for [FPP]. As the termination criteria of [MIQAP] in Table 1, we set the maximum number of total iterations to solve relaxation problems for [MIQAP] at Step 3 to 25000, the maximum number of iterations to solve relaxation problems for [MIQAP] after finding a solution that gives the minimum objective value to 10000, and the optimality gap to 17%. Thus, we terminate the solution process of [MIQAP] if one of the above three criteria is satisfied.

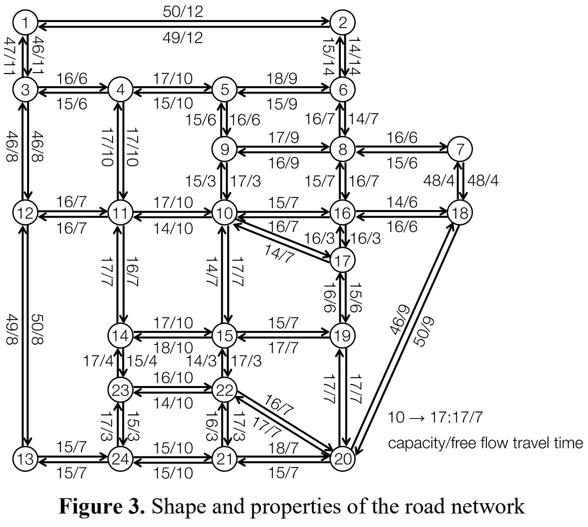


Figure 3. Shape and properties of the road network

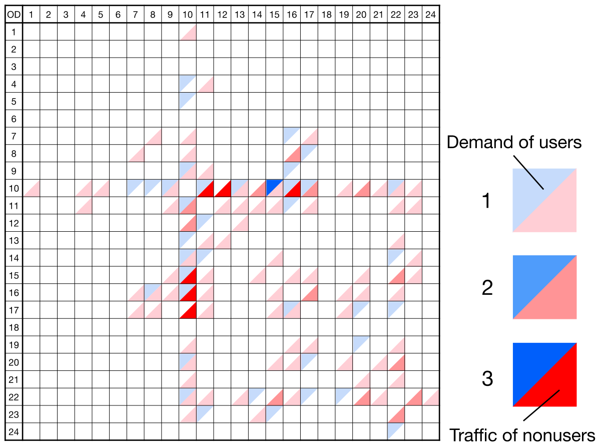


Figure 4. OD demands of users and nonusers

To assess the validity of the solution obtained from [FPP], we compare the solution of [FPP] with that from a comparison algorithm. The comparison algorithm in this study is comprised of the Branch-and-Cut algorithm and the UE assignment. We need to enumerate the candidate solutions to [OP]. However, due to the nonlinear constraints or equilibrium constraints, we cannot apply the Branch-and-Cut algorithm directly to [OP]. Thereby, we solve first the relaxation problem of [OP] referred to as "[OP] without UE constraints" for calculating OD demands. We can define [OP] without UE constraints by assuming the travel times as fixed free flow travel times, which corresponds to the conventional DARP as follows:
[OP] without UE constraints

$$\min Z_{\mathrm{wo}} = \sum_{rs \in \boldsymbol{\Omega}} t_{rs}^{0} \cdot (\omega_{rs} + \bar{\omega}_{rs}) \tag{46}$$

subject to $(2) - (13)$, where

$$t_{rs}^{0} = \sum_{a \in \boldsymbol{A}^{\downarrow}} \delta'^{rs}_{a} \cdot t_{a}^{0} \qquad \forall rs \in \boldsymbol{\Omega} \tag{47}$$

Here, δ'^{rs}_{a} is a variable which equals 1 if link a is contained in the shortest path between nodes of OD pair rs, and 0 otherwise. Thus, t_{rs}^{0} is the shortest OD travel time between nodes of OD pair rs without congestion. Then, by applying the Branch-and-Cut algorithm to [OP] without UE constraints, we can enumerate the candidate solutions in which the objective function values Z_{wo} are sorted in ascending order. For each candidate solution $\boldsymbol{x}$, OD demands $\omega_{rs}\ \forall rs \in \boldsymbol{\Omega}$ are calculated by substituting it into (13). We solve next the UE assignment problem for calculating OD travel times $t_{rs}\ \forall rs \in \boldsymbol{\Omega}$. Then, the total travel time Z is obtained by substituting both ω_{rs} and $t_{rs}\ \forall rs \in \boldsymbol{\Omega}$ into (1). Finally, a candidate solution with the shortest total travel time becomes the output of this algorithm.

If all candidate solutions are enumerated, it is the most ideal for obtaining the exact solution to [OP]. However, the feasible region of discrete variables in [OP] is enormous. It is practically impossible to enumerate all possible solutions. Instead, in this study, we obtain a solution that minimizes the objective variable as much as possible by repeating the generation of candidate solutions by the

Branch-and-Cut algorithm and the evaluation of the total travel time Z for a sufficiently long time, i.e., 24 hours. By comparing the values of the total travel time corresponding to all candidate solutions enumerated by the Branch-and-Cut algorithm, a solution that minimizes the objective variable is found. We define this solution as a seemingly exact solution and compare it with the solution obtained by the proposed algorithm.

The above series of processes is defined as the comparison algorithm. Table 2 illustrates the comparison algorithm, Branch-and-Cut and UE based algorithm, to obtain a seemingly exact solution based on the Branch-and-Cut algorithm. As far as the authors know, this algorithm is one of those that can provide a seemingly exact solution for problems such as [OP], which are difficult to find an exact solution. In the following section, we compare the solutions and the computation times by the proposed algorithm and the comparison algorithm to evaluate the validity of the proposed algorithm.

Table 2. Branch-and-Cut and UE based algorithm

Step 1: Set the total travel time of the currently optimal solution $Z^* = \infty$

Step 2: By Branch-and-Cut algorithm, solve [OP] without UE constraints where the travel times of all links are fixed as its constant free flow travel times.

If a feasible solution $\boldsymbol{x}^R$ for [OP] without UE constraints is found,

Calculate OD demands ω_{rs} $\forall rs \in \boldsymbol{\Omega}$ by substituting $\boldsymbol{x}^R$ into (13).

Carry out the UE assignment to obtain OD travel time t_{rs} $\forall rs \in \boldsymbol{\Omega}$ in [OP].

Calculate the total travel time Z by substituting ω_{rs} and t_{rs} $\forall rs \in \boldsymbol{\Omega}$ into (1).

Check if $\boldsymbol{x}^R$ is also feasible for [OP] by substituting it and t_{rs} $\forall rs \in \boldsymbol{\Omega}$ into (10)-(12).

If $\boldsymbol{x}^R$ is feasible for [OP] and $Z < Z^*$

Record $\boldsymbol{x}^R$ as the currently optimal solution, and update $Z^* = Z$

Step 3: Output the currently optimal solution as the seemingly exact solution of [OP] and terminate the algorithm

4.2 Result

By solving [FPP], we obtained a solution with a total travel time of 1869.9, and its computation time was 12 hours. Additionally, the Branch-and-Cut algorithm without considering congestion produced a solution with the same total travel time, and its computation time was 24 hours, as mentioned earlier. Figure 5 illustrates the time-series changes in total travel times obtained by the proposed algorithm and the comparison algorithm during the calculations. We confirm that the proposed algorithm converged to the seemingly exact solution faster than the comparison algorithm.

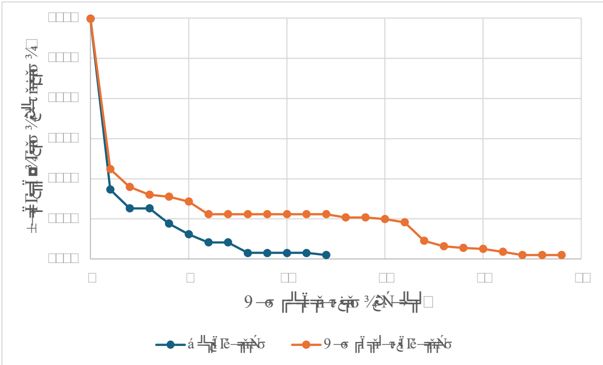


Figure 5. Changes in total travel times

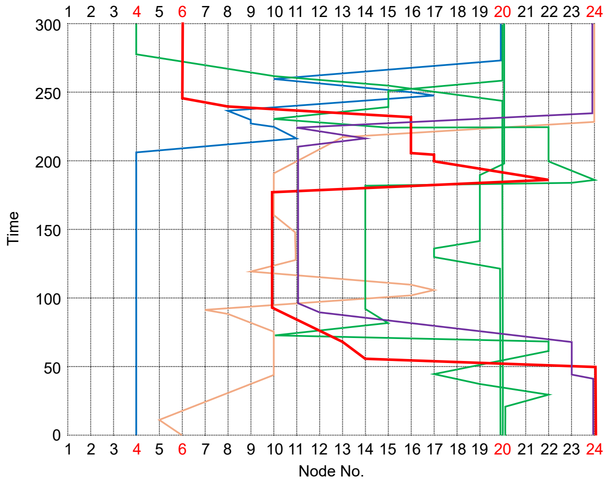


Figure 6. Obtained operation plan of all SAVs

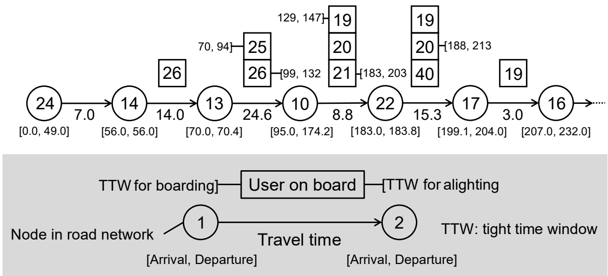


Figure 7. Detailed operation plan of an SAV

Figure 6 illustrates the movements of SAVs obtained from our model. Additionally, the numbers of SAVs allocated to depot nodes 4, 6, 20 and 24 before the operation are 1, 1, 3, and 2, respectively. Moreover, these numbers remain the same at the end of the operation. Therefore, the next operation can be restarted without any relocations.

Figure 7 presents a portion of the detailed operation of the SAV whose movements are represented by the red bold line segments in Figure 6, i.e., the operation of the SAV that departs from depot node 24 and eventually returns to depot node 6. Here, we confirm the SAV's arrival and departure times do not violate the time windows or the maximum in-vehicle time.

4.3 Discussion

Here, we discuss the relative accuracy of the convex approximation. Figure 8 presents histograms of the eigenvalues of $\boldsymbol{M}+\boldsymbol{M}^T$ calculated in the first iteration. The horizontal axis represents the orders of magnitude of the eigenvalues calculated by $\log_{10}\lambda_i\ \forall i\in\{1,2,\dots,N^2\}$. The upper and lower histograms correspond to positive and negative eigenvalues, respectively. Additionally, there are three eigenvalues equal to zero though they are not shown in Figure 8. The frequency distribution of positive eigenvalues, whose absolute values are smaller than 10^{-10}, is similar to the distribution of negative eigenvalues. In contrast, eigenvalues whose absolute values are greater than 10^{-10} are all positive. The matrix $\boldsymbol{Q}$ is obtained by replacing negative eigenvalues of $\boldsymbol{M}+\boldsymbol{M}^T$ with zeros. Figure 8 illustrates that the negative eigenvalues of $\boldsymbol{M}+\boldsymbol{M}^T$ are close to zero. Thus, we can obtain an accurate total travel time by substituting a solution into (38), compared to the value obtained using (29) without approximation. In fact, the difference between the objective values of (29) and (38) at the solution is smaller than 10^{-5}.

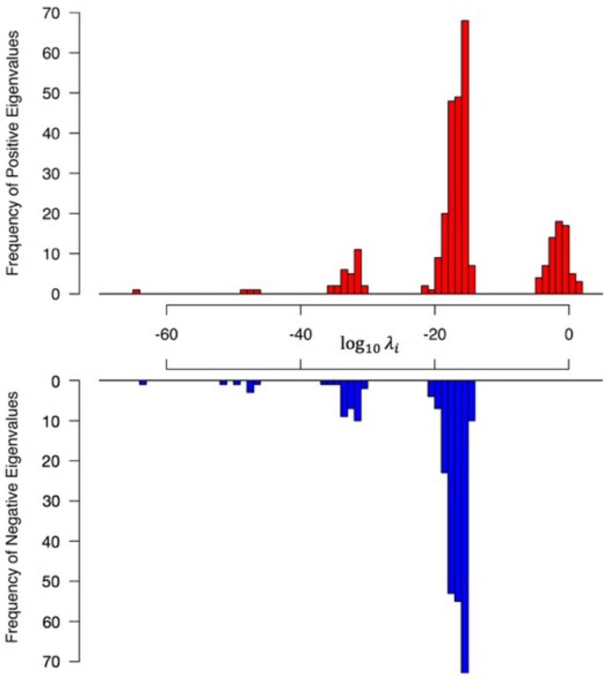


Figure 8. Eigenvalues of $\boldsymbol{M}+\boldsymbol{M}^T$

5 Conclusion

In this research, we construct a model for operating SAVs that accounts for congestion in the road network by combining a static UE traffic assignment with the conventional DARP. As far as the authors know, this is the first model that considers congestion theoretically and provides detailed input and output corresponding to each vehicle's operation. Moreover, we propose a computationally efficient solution algorithm to obtain a seemingly exact solution. The results of numerical calculations demonstrate the validity of the model and the solution algorithm.

Here, we list future challenges. Firstly, traffic assignment can be more realistic by applying stochastic UE or semi-dynamic UE (e.g. [28]). Semi-dynamic UE would help us to make a plan for a full-day operation. Secondly, the validity of the convex approximation needs to be verified by mathematically analyzing the properties of $\boldsymbol{M}+\boldsymbol{M}^T$. Thirdly, the applicability of our model and its solution algorithm to the real-size road network problem is currently limited, as the algorithm can only handle problems with a relatively small number of users. Therefore, a more efficient algorithm that can solve the problem with a significantly larger number of users is needed. Finally, a method for obtaining the exact optimal solution remains undeveloped. While approaches such as those by Wang et al. [29] and Wang and Lo [30] could be considered, their models assume OD traffic to be constant, which differs from our approach.

List of Abbreviations: SAV: Shared autonomous vehicle; DARP: Dial-a-Ride problem; OD: Origin and destination; UE: User equilibrium; GA: Genetic algorithm, OP: Original problem, FPP: Fixed-point problem, MIQAP: Mixed-integer quadratic approximation problem

Statements and Declarations

Competing Interests: The authors declare that they have no conflict of interest.

References

1. Duncan, M.: The cost saving potential of carsharing in a US context, Transportation **38**, 363-382, (2011).
2. Krueger, R., Rashidi, T. H., and Rose, J. M.: Preferences for shared autonomous vehicles, Transp. Res. Part C **69**, 343-355, (2016).
3. de Almeida Correia, G. H. and van Arem, B.: Solving the User Optimum Privately Owned Automated Vehicles Assignment Problem (UO-POAVAP): A model to explore the impacts of self-driving vehicles on urban mobility, Transp. Res. Part B **87**, 64-88, (2016).
4. Fagnant, D. J. and Kockelman, K.: Preparing a nation for autonomous vehicles: opportunities, barriers and policy recommendations, Transp. Res. Part A **77**, 167-181, (2015).
5. Levin, M. W., Kockelman, K. M., Boyles, S. D. and Li, T.: A general framework for modeling shared autonomous vehicles with dynamic network-loading and dynamic ride-sharing application, Comput., Environ. and Urban Syst. **64**, 373-383, (2017).
6. Gurumurthy, K. M., de Souza, F., Enam, A. and Auld, J.: Integrating Supply and Demand Perspectives for a Large-Scale Simulation of Shared Autonomous Vehicles, Transp. Res. Rec. **2674**(7), 181-192, (2020).
7. Hörl, S., Becker, F. and Axhausen, K. W.: Simulation of price, customer behaviour and system impact for a cost-covering automated taxi system in Zurich, Transp. Res. Part C **123**, 102974, (2021).

8. Oh, S., Seshadri, R., Azevedo, C. L., Kumar, N., Basak, K. and Ben-Akiva, M.: Assessing the impacts of automated mobility-on-demand through agent-based simulation: A study of Singapore, Transp. Res. Part A **138**, 367-388, (2020).

9. Pinto, H. K., Hyland, M. F., Mahmassani, H. S. and Verbas, O.: Joint design of multimodal transit networks and shared autonomous mobility fleets, Transp. Res. Part C **113**, 2-20, (2020).

10. Jaw, J.-J., Odoni, A. R., Psaraftis, H. N. and Wilson, N. H.: A heuristic algorithm for the multi-vehicle advance request dial-a-ride problem with time windows, Transp. Res. Part B **20**(3), 243-257, (1986).

11. Psaraftis, H. N.: An Exact Algorithm for the Single Vehicle Many-to-Many Dial-A-Ride Problem with Time Windows, Transp. Sci. **17**, 351-357, (1983).

12. Dumas, Y., Desrosiers, J. and Soumis, F.: Large scale multi-vehicle dial-a-ride problems, Tech. Rep. G-89-30, GERAD, HEC Montreal, Montreal, Quebec, Canada, (1989).

13. Cordeau, J.-F. and Laporte, G.: A tabu search heuristic for the static multi-vehicle dial-a-ride problem, Transp. Res. Part B **37**, 579-594, (2003).

14. Cubillos, C., Rodriguez, N. and Crawford, B.: A Study on Genetic Algorithms for the DARP Problem, Bio-inspired Modeling of Cognitive Tasks. IWINAC 2007. Lecture Notes in Computer Science 4527. Springer, Heidelberg, Berlin, Germany, 498-507, (2007).

15. Cordeau, J.-F.: A Branch-and-Cut Algorithm for the Dial-a-Ride Problem, Oper. Res. **54**(3), 573-586, (2006).

16. Ropke, S., Cordeau, J.-F. and Laporte, G.: Models and branch-and-cut algorithms for pickup and delivery problems with time windows, Networks **49**(4), 258-272, (2007).

17. Levin, M. W.: Congestion-aware system optimal route choice for shared autonomous vehicles, Transp. Res. Part C **82**, 229-247, (2017).

18. Tani, R, Nakauchi, T. and Uchida, K.: Ridesharing Traffic Flow Model based on Link-based Kinematic Wave Theory, J. of the East. Asia Soc. for Transp. Studies **14**, 634-646, (2021).

19. Liang, X., de Almeida Correia, G. H., An, K. and van Arem, B.: Automated taxis' dial-a-ride problem with ride-sharing considering congestion-based dynamic travel times, Transp. Res. Part C **112**, 260-281, (2020).

20. Fan, Q., van Essen, T. J. and de Almeida Correia, G. H.: Optimising fleet sizing and management of shared automated vehicle (SAV) services: A mixed-integer programming approach integrating endogenous demand, congestion effects, and accept/reject mechanism impacts, Transp. Res. Part C **157**, 104398, (2023).

21. Zhang, Z., Liu, W. and Zhang, F.: On the joint network equilibrium of parking and travel choices under mixed traffic of shared and private autonomous vehicles, Transp. Res. Part C **153**, 104226, (2023).

22. Wardrop, J. G.: Some Theoretical Aspects of Road Traffic Research, Proc. of the Inst. of Civ. Eng. Part II **1**(3), 325-378, (1952).

23. Bureau of Public Roads, Traffic Assignment Manual, U.S. Dept. of Commerce, Urban Planning Division, Washington D.C., USA., (1964).

24. Ruland, K. S. and Rodin, E. Y.: The pickup and delivery problem: Faces and branch-and-cut algorithm, Comput. and Math. with Appl. **33**(12), 1-13, (1997).

25. Tobin, R. L. and Friesz, T. L.: Sensitivity Analysis for Equilibrium Network Flow, Transp. Sci. **22**(4), 242-250, (1988).

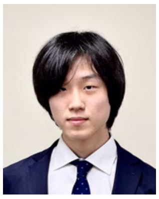
26. Tanaka, M. and Nakata, K.: Positive definite matrix approximation with condition number constraint, Optim. Lett. **8**, 939-947, (2014).

27. Farahani, R. Z., Miandoabchi, E., Szeto, W. Y. and Rashidi, H.: A review of urban transportation network design problems, Eur. J. of Oper. Res. **229**(2), 281-302, (2013).

28. Fujita, M., Matsui, H. and Mizokami, S.: Modelling of the time-of-day traffic assignment over a traffic network, Proc. of Jpn. Soc. of Civ. Eng. **389**, 111-119, (1988). (in Japanese)

29. Wang, S., Meng, Q. and Yang, H.: Global optimization methods for the discrete network design problem, Transp. Res. Part B **50**, 42-60, (2013).

30. Wang, D. Z. and Lo, H. K.: Global optimum of the linearized network design problem with equilibrium flows, Transp. Res. Part B **44**(4), 482-492, (2010).



Sorachi Matsumoto is a master student at Hokkaido University, Japan. His research interest includes transportation planning and policy for autonomous vehicles.

Ryuichi Tani is an Assistant Professor of Hokkaido University, Japan. He received a PhD from Hokkaido University in 2020. His research interest includes transportation planning, and travel time reliability.

Kenetsu Uchida is a Professor of Hokkaido University, Japan. He received a Doctor of Engineering from Hokkaido University in 1999. His research interest includes transportation planning, and travel time reliability.